Irish Mathematical Society
Cumann Matamaitice na hÉireann

Bulletin

Number 75 Summer 2015

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, Irish thesis abstracts, 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
NUI, Maynooth
Co. Kildare

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

`http://www.maths.tcd.ie/pub/ims/`

Printed in the University of Limerick

EDITORIAL

At the 150th birthday celebration of the LMS, those present were given a copy of the speech of Augustus De Morgan, President, at the first meeting of the Society, January 16th, 1865. He pointed out that the new society would, inevitably, be influenced mainly by a few energetic individuals, but would come to nothing if it did not *please* the general body of members. Having reviewed previous British societies and journals and linked their failure to capture by special interests and a "preponderance of subjects of one particular kind", he said that his desire was to see the society *"having the great bulk of its business adapted to the great bulk of its members"*[1]. With an emphasis on the future of mathematics, and granting that one cannot actually foresee how mathematical science will develop, he tentatively suggested some directions that deserved more attention in his day, to wit history, logic[2], and (is anything new?) his desire to see elementary students think more:

> "Mathematics is becoming too much of a machinery, and this is more especially the case with reference to the elementary students. They put the data of the problems into a mill and expect the result to come out ready ground at the other end. An operation which bears a close resebmblance to putting in hemp seeds at one end of a machine and taking out ruffled shirts ready for use at the other end. This mode is undoubtedly exceedingly effective in producing results, but it is certainly not so the teaching the mind and in exercising thought."

The Bulletin of the IMS has the same aim as that expressed for the LMS journals by De Morgan: to publish material of general interest to members, without limiting the areas. Members (and others) are encouraged to contribute.

In the present issue we publish a couple of surveys related to stochastic integrals and their application, a classroom note about inverse functions, as well as some reviews of books on varied pure and applied topics, and the problem page. The surveys are by Danny

[1]— and clearly he was not referring to the BMI of the members!

[2]He referred warmly to Boole's work.

Duffy, who gives an account from the perspective of an experienced practitioner of the kind of numerical methods used in finance, and by Pat Muldowney, who gives a gentle introduction to the basic issues around stochastic and other integrals, giving due credit to his mentor Ralph Henstock, whose fundamental work on integration deserves to be better-known in Ireland. Henstock was for many years Professor of Pure Mathematics at the University of Ulster in Coleraine. I can testify from personal experience that he was a tolerant, kindly and hospitable fellow. Both Duffy and Muldowney have published books which may be consulted by readers who wish to learn more about the material outlined.

The next main scientific meeting of the Society will take place in UCC, on 27-28 August, and will form a part of the Boole Centenary celebrations there.

The IMS Committee has adopted revised guidelines for conference organisers who wish to apply for support. These may be found at the IMS website. Organisers are reminded that reports should be submitted to the Bulletin by December, in good time for the Winter issue.

The Treasurer asked me to draw the attention of members aged over 65 to the fact that a reduced subscription rate applies. See page 2, and contact him to take advantage of this.

AOF. Department of Mathematics and Statistics, Maynooth University, Co. Kildare

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

LINKS FOR POSTGRADUATE STUDY

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

DCU: (Olaf Menkens)
http://www.dcu.ie/info/staff_member.php?id_no=2659
DIT: (Chris Hills)
mailto://chris.hills@dit.ie
NUIG:
mailto://james.cruickshank@nuigalway.ie
NUIM:
http://www.maths.nuim.ie/pghowtoapply
QUB:
http://www.qub.ac.uk/puremaths/Funded_PG_2012.html
TCD:
http://www.maths.tcd.ie/postgraduate/
UCD:
mailto://nuria.garcia@ucd.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].

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

[1] http://www.maths.tcd.ie/pub/ims/bulletin/

Irish Mathematical Society Bulletin 75, Summer 2015

CONTENTS

NOTICES FROM THE SOCIETY

Officers and Committee Members 2014

President	Dr M. Mackey	University College Dublin
Vice-President	Prof S. Buckley	Maynooth University
Secretary	Dr R. Quinlan	NUI Galway
Treasurer	Prof G. Pfeiffer	NUI Galway

Dr P. Barry, Prof J. Gleeson, Dr B. Kreussler, Dr R. Levene, Dr M. Mac an Airchinnigh, 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 C. Hills
	DCU	Dr M. Clancy
	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

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

DECEASED MEMBERS

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

Dr Derek O'Connor of Donard, Co. Wicklow, died on 27 March 2013. He was a member of the Society for 7 years.

Prof. Allan Solomon, late of the Open University, died on 3 April 2013. He had been a member of the Society from its beginning.

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 5–6
ISSN 0791-5578

FLAT SURFACES OF FINITE TYPE IN THE 3-SPHERE

ALAN MCCARTHY

This is an abstract of the PhD thesis *Flat surfaces of Finite Type in the 3-Sphere* written by Alan McCarthy under the supervision of Martin Kilian at the School of Mathematical Sciences, University College Cork, and submitted in May 2014.

The local theory of flat surfaces in $\mathbb{S}^3$ through the use of asymptotic curves in $\mathbb{S}^3$ was already known by Bianchi [1], however the problem of classifying flat tori was first posed by Yau [4] in 1974. Kitagawa [2] provided a classification by using asymptotic lifts of 'admissible pairs' of closed curves on $\mathbb{S}^2$. Flat tori were also classified by Weiner [3] in terms of their Gauss maps.

My thesis is concerned with finite gap flat surfaces. These are surfaces whose generating curves on $\mathbb{S}^2$ have finite gap geodesic curvatures, which means that eventually all flows of the mKdV hierarchy are finite linear combinations of preceeding ones.

We provide a summary of finite gap curves in terms of Lax pairs, Killing fields, their spectral curves and provide conditions that ensure that the curve remains closed and spherical. We also provide a discussion of the isoperiodic deformations and monodromy associated to the frame of the curves.

As an application we show that given an admissible pair of curves $\gamma_1, \ \gamma_2$ with geodesic curvatures $k_1, \ k_2 \in L^2(\mathbb{S}^1, \mathbb{R})$, there exists a pair of finite gap curvature functions that generate curves on $\mathbb{S}^2$ that are admissible and that these finite gap curvatures are also dense in the Sobolev norm.

2010 *Mathematics Subject Classification.* 51H25, 35P30,58A02,53C02.

Key words and phrases. differential geometry, integrable systems, flat, tori, finite type, finite gap.

Received on 23-1-2015.

Support from University College Cork through a part time lecturship is gratefully acknowledged.

References

[1] L. Bianchi. Sulle superficie a curvatura nulla in geometria ellittica. *Ann. Mat. Pura Appl.*, 24(1):9–19, 1896.

[2] Y. Kitagawa. Periodicity of the asymptotic curves on flat tori in S^3. *J. Math. Soc. Japan*, 40(3):457–476, 1988.

[3] J. L. Weiner. Flat tori in S^3 and their Gauss maps. *Proc. London Math. Soc. (3)*, 62(1):54–76, 1991.

[4] S. T. Yau. Submanifolds with constant mean curvature. I, II. *Amer. J. Math.*, 96:346–366; ibid. 97 (1975), 76–100, 1974.

(from August 2014) School of Mathematics and Statistics, University of New South Wales, Sydney

E-mail address, A. McCarthy: a.mccarthy@unsw.edu.au

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 7–19
ISSN 0791-5578

FROM NAVIER-STOKES TO BLACK-SCHOLES: NUMERICAL METHODS IN COMPUTATIONAL FINANCE

DANIEL J. DUFFY

ABSTRACT. In this article we give a general overview of the numerical methods (in particular the finite difference method) to approximate the partial differential equations that describe the behaviour of financial products (such as stocks, options, commodities and interest rate products). These products are traded in the marketplace and it is important to price them using accurate and efficient algorithms. Furthermore, financial institutions need to compute and monitor the risks associated with these financial instruments and portfolios of these instruments.

The focus in this article is to trace the emergence of advanced numerical techniques and their applications to computational finance during the last twenty-five years. It is aimed at a mathematical audience with a passing acquaintance of partial differential equations (PDEs) and finite difference methods. In particular, time-dependent convection-diffusion-reaction PDEs will take centre-stage because they model a wide range of financial products.

1. A short History of Computational Finance

Computational Finance can be defined as a set of mathematical and engineering techniques to solve complex problems in finance. It has grown steadily during the last thirty years as financial services became global and computing power increased exponentially. When the Cold War ended the market had acquired access to a large pool of physicists, mathematicians and computer scientists, or quants as they became known on Wall Street. These quants applied their knowledge to solve complex derivatives pricing problems. Growth was explosive until the financial crash of 2007/2008. After the crash

2010 *Mathematics Subject Classification.* 91G60, 65C30.
Key words and phrases. Financial Mathematics, Options, Derivatives, Computation, PDE.
Received on 20-2-2015; revised 8-3-2015.

many of the exotic structured products that these quants had invented were abandoned in favour of simpler ones. The events of the last seven years have certainly proved that finance is not physics and that models are, after all, just models of reality and not reality itself.

2. Computational Finance 101: Plain Call Options

Before we jump into the mathematics and numerical analysis of partial differential equations we try to sketch the financial context in which they are used. It is impossible to discuss the context in any great detail and we refer the reader to Wilmott [11]. It is written in a style that should appeal to mathematicians.

In order to reduce the scope we focus exclusively on the most fundamental of all financial instruments, namely equity (also known as stock or shares). Holding equity means that you own part of a company. If the company goes bankrupt the value of your shares is effectively zero or thereabouts. In short, you have lost your investment! The investor paid up front and she was probably expecting the share price to increase in the future. Most people are optimists and hence they buy shares in the hope that they will rise in price. But this is risky because if the share price drops they will make a loss. There is however, a less risky approach. Let us assume that you expect the share price of the ABC company to rise from \$100 to \$140 in the next three months (for example, you consulted your crystal ball on this and that is what it told you). So, instead of buying the share for \$100 now you might like to have the option to wait for three months and then buy the share. You can then buy a *call option* that gives you the right but not the obligation to buy the share three months into the future for a certain *strike price*. Of course, having the right but not the obligation to buy a share at some time in the future comes at a price and this must be paid by the investor up-front. For example, you can buy a call option with strike price \$120 that expires in three months time. If the price is greater than \$120 then you have made a profit. If the price is less than \$120 at expiration the option is worthless and you have lost your initial investment.

We have not addressed the issue of how to compute the option price. This is precisely the famous Black Scholes formula [1] that allows us to compute the option price analytically. In more complicated cases we need to resort to numerical methods as discussed in this article.

3. The Mathematics of PDEs in Computational Finance: Helicopter View

In general, the PDEs of relevance are of the *convection-diffusion-reaction* type in n space variables and one time variable. The space variables correspond to underlying financial quantities such as an asset or interest rate while the non-negative time variable t is bounded above by the *expiration* T. The space variables take values in their respective positive half-planes.

We model derivatives that are described by so-called initial boundary value problems of parabolic type [10]. To this end, consider the general parabolic equation:

$$Lu \equiv \sum_{i,j=1}^{n} a_{ij}(x,t)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{j=1}^{n} b_j(x,t)\frac{\partial u}{\partial x_j} + c(x,t)u - \frac{\partial u}{\partial t} = f(x,t) \tag{1}$$

where the functions a_{ij}, b_j, c and f are real-valued $a_{ij} = a_{ji}$, and

$$\sum_{i,j=1}^{n} a_{ij}(x,t)\alpha_i\alpha_j > 0 \qquad \text{if} \qquad \sum_{j=1}^{n} \alpha_j^2 > 0. \tag{2}$$

In equation (2) the variable x is a point in n-dimensional space and t is considered to be a positive time variable. Equation (1) is the general equation that describes the behaviour of many derivative types. For example, in the one-dimensional case ($n = 1$) it reduces to the famous Black-Scholes equation (Here $t^* = T - t$):

$$\frac{\partial V}{\partial t^\star} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - D)S\frac{\partial V}{\partial S} - rV = 0 \tag{3}$$

where V is the derivative type (for example a call or put option), S is the underlying asset (or stock), σ is the constant volatility, r is the interest rate and D is a dividend. Equation (3) is a special case and it can be generalised to include more general kinds of options.

Equation (3) can be generalised to the multivariate case:

$$\frac{\partial V}{\partial t^\star} + \sum_{j=1}^{n}(r - D_j)S_j\frac{\partial V}{\partial S_j} + \frac{1}{2}\sum_{i,j=1}^{n}\rho_{ij}\sigma_i\sigma_j S_i S_j \frac{\partial^2 V}{\partial S_i \partial S_j} = rV. \tag{4}$$

This equation models a multi-asset environment. In this case σ_i is the volatility of the i^{th} asset and ρ_{ij} is the correlation $(-1 \leq \rho_{ij} \leq 1)$ between assets i and j. In this case we see that equation (4) is written as the sum of three terms:

- Interest earned on cash position

$$r\Big(V - \sum_{j=1}^{n} S_j \frac{\partial V}{\partial S_j}\Big). \tag{5}$$

- Gain from dividend yield

$$\sum_{j=1}^{n} D_j S_j \frac{\partial V}{\partial S_j}. \tag{6}$$

- Hedging costs or slippage

$$-\frac{1}{2}\sum_{i,j=1}^{n} \rho_{ij}\sigma_i\sigma_j S_i S_j \frac{\partial^2 V}{\partial S_i \partial S_j}. \tag{7}$$

Our interest is in discovering robust numerical schemes that produce reliable and accurate results irrespective of the size of the parameter values in equation (4).

Equation (1) has an infinite number of solutions in general. In order to reduce this number to one, we need to define some constraints. To this end, we define so-called *initial condition* and *boundary conditions* for (1). We achieve this by defining the space in which equation (1) is assumed to be valid. In general, we note that there are three types of boundary conditions associated with equation (1) (see [10]). These are:

- First boundary value problem (Dirichlet problem).
- Second boundary value problem (Neumann, Robin problems).
- Cauchy problem.

The first boundary value problem is concerned with the solution of (1) in a domain $D = \Omega \times (0, T)$ where Ω is a bounded subset of $\mathbb{R}^{\mathrm{n}}$ and T is a positive number. In this case we seek a solution of (1) satisfying the conditions:

$$\begin{aligned} u|_{t=0} &= \varphi(x) \qquad \text{(initial condition)} \\ u|_{\Gamma} &= \psi(x,t) \qquad \text{(boundary condition)} \end{aligned} \tag{8}$$

where Γ is the boundary of Ω. The boundary conditions in (8) are called Dirichlet boundary conditions. These conditions arise when we model single and double barrier options in the one-factor case (see [5]). They also occur when we model plain options.

The second boundary value problem is similar to (8) except that instead of giving the value of u on the boundary Γ the directional derivatives are included, as seen in the following specification:

$$\Big(\frac{\partial u}{\partial \eta} + a(x,t)u\Big)|_\Gamma = \psi(x,t). \tag{9}$$

In this case $a(x,t)$ and $\psi(x,t)$ are known functions of x and t, and $\frac{\partial}{\partial \eta}$ denotes the derivative of u with respect to the outward normal η at Γ . A special case of (9) is when $a(x,t) \equiv 0$; then (9) represents the *Neumann boundary conditions.* These occur when modelling certain kinds of put options. Finally, the solution of the Cauchy problem for (1) in the strip $\mathbb{R}^n \times (0,T)$ is given by the initial condition:

$$u|_{t=0} = \varphi(x) \tag{10}$$

where $\varphi(x)$ is a given continuous function and $u(x,t)$ is a function that satisfies (1) in $\mathbb{R}^n \times (0,T)$ and that satisfies the initial condition (10). This problem allows negative values of the components of the independent variable $x = (x_1, \ldots, x_n)$. A special case of the Cauchy problem can be seen in the modelling of one-factor European and American options (see [11]) where x plays the role of the underlying asset S. Boundary conditions are given by values at $S = 0$ and $S = \infty$. For European options these conditions are:

$$\begin{aligned} &C(0,t) = 0 \\ &C(S,t) \to S \qquad \text{as} \qquad S \to \infty. \end{aligned} \tag{11}$$

Here C (the role played by u in equation (1)) is the variable representing the price of the call option. For European put options the boundary conditions are:

$$\begin{aligned} &P(0,t) = Ke^{-r(T-t^\star)} \\ &P(S,t) \to 0 \qquad \text{as} \qquad S \to \infty. \end{aligned} \tag{12}$$

Here P (the role played by u in equation (1)) is the variable representing the price of the put option, K is the strike price, r is the risk-free interest rate, T is the expiration and t is the current time.

From this point on we assume the following 'canonical' form for the operator L in equation (1):

$$Lu \equiv -\frac{\partial u}{\partial t} + \sigma(x,t)\frac{\partial^2 u}{\partial x^2} + \mu(x,t)\frac{\partial u}{\partial x} + b(x,t)u = f(x,t) \tag{13}$$

where σ, μ, b and f are known functions of x and t.

We have given a global introduction to the kinds of linear partial differential equations that are used in computational finance. We are unable to discuss other topics such as nonlinear PDEs, free and moving-boundary value problems, qualitative properties of equation (1) (for example, criteria for existence and uniqueness of the solution of equation (1)) and applications to computational finance. For a discussion of these topics we refer the reader to [5].

For the rest of this article we restrict our attention to the linear one-factor PDE defined by equation (13) in conjunction with auxiliary conditions to ensure existence and uniqueness. We also assume that all the coefficients and inhomogeneous term in equation (13) are known.

4. The Finite Difference Method (FDM) in Computational Finance

For completeness, we formulate the initial boundary value problem whose solution we wish to approximate using the finite difference method.

Define the interval $\Omega = (A, B)$ where A and B are two real numbers. Further let $T > 0$ and $D = \Omega \times (0, T)$.
The formal statement of the idealised problem is:
With

$$Lu \equiv -\frac{\partial u}{\partial t} + \sigma(x,t)\frac{\partial^2 u}{\partial x^2} + \mu(x,t)\frac{\partial u}{\partial x} + b(x,t)u,$$

find a function $u : D \to \mathbb{R}^1$ *such that*

$$Lu = f(x,t) \text{ in } D \tag{14}$$

$$u(x,0) = \varphi(x),\ x \in \Omega \tag{15}$$

$$u(A,t) = g_0(t),\ u(B,t) = g_1(t),\ t \in (0,T). \tag{16}$$

The initial-boundary value problem (14)-(16) is general and it subsumes specific cases from the option pricing literature (in particular it is a generalisation of the original Black-Scholes equation).

In general, the coefficients $\sigma(x,t)$ and $\mu(x,t)$ represent *volatility* (diffusivity) and *drift* (convection), respectively. Equation (14) is called the *convection-diffusion-reaction equation.* It serves as a model for many kinds of physical and economic phenomena. Much research has been carried out in this area, both on the continuous problem and its discrete formulations (for example, using finite difference and finite element methods). In particular, research has shown that standard centred-difference schemes fail to approximate (14)-(16) properly in certain cases (see [4]) .

The essence of the finite difference method is to discretise equation (14) by defining so-called discrete mesh points and approximating the derivatives of the unknown solution of system (14) - (16) in some way at these mesh points. The eventual goal is to find accurate schemes that will be implemented in a programming language such C++ or C# for the benefit of traders and risk management. Some typical attention points are:

- The PDE being approximated may need to be pre-processed in some way, for example transforming it from one on a semi-infinite domain to one on a bounded domain.
- Determining which specific finite difference scheme(s) to use based on quality requirements such as accuracy, efficiency and maintainability.
- Essential difficulties to resolve: convection dominance, avoiding oscillations and how to handle discontinuous initial conditions, for example.
- Developing the algorithms and assembling the discrete system of equations prior to implementation.

Our goal is to approximate (14)-(16) by finite difference schemes. To this end, we divide the interval $[A, B]$ into the sub-intervals:

$$A = x_0 < x_1 < \ldots < x_J = B$$

and we assume for convenience that the mesh-points $\{x_j\}_{j=0}^{J}$ are equidistant, that is:

$$x_j = x_{j-1} + h, \quad j = 1, \ldots, \quad J. \ \left(h = \frac{B-A}{J}.\right)$$

Furthermore, we divide the interval $[0, T]$ into N equal sub-intervals $0 = t_0 < t_1 < \ldots < t_N = T$ where $t_n = t_{n-1} + k, \quad n = 1, \ldots, N \quad (k = T/N)$.

(It is possible to define non-equidistant mesh-points in the x and t directions but doing so would complicate the mathematics and we would be in danger of losing focus).

The essence of the finite difference method lies in replacing the derivatives in (14) by divided differences at the mesh-points (x_j, t_n). We define the difference operators in the x-direction as follows:

$$
\begin{aligned}
&D_+ u_j = (u_{j+1} - u_j)/h, \quad D_- u_j = (u_j - u_{j-1})/h \\
&D_0 u_j = (u_{j+1} - u_{j-1})/2h, \quad D_+ D_- u_j = (u_{j+1} - 2u_j + u_{j-1})/h^2.
\end{aligned}
$$

It can be shown by Taylor expansions that D_+ and D_- are first-order approximations to $\frac{\partial}{\partial x}$, respectivily while D_0 is a second-order approximation to $\frac{\partial}{\partial x}$. Finally, $D_+ D_-$ is a second-order approximation to $\frac{\partial^2}{\partial x^2}$.

We also need to discretise the time dimension and to this end we consider the scalar initial value problem:

$$
\begin{cases}
Lu \equiv u^{'}(t) + a(t)u(t) = f(t), \forall t \ \in \ [0, T] \\
\ \text{with} \ \ a(t) \geq \alpha > 0, \forall t \ \in \ [0, T]. \\
u(0) = A.
\end{cases}
\tag{17}
$$

The interval where the solution of (17) is defined is $[0, T]$. When approximating the solution using finite difference equations we use a discrete set of points in $[0, T]$ where the discrete solution will be calculated. To this end, we divide $[0, T]$ into N equal intervals of length k where k is a positive number called the *step size*. In general all coefficients and discrete functions will be defined at these *mesh points*. We draw a distinction between those functions that are known at the mesh points and the solution of the corresponding difference scheme. We adopt the following notation:

$$
\begin{aligned}
&a^n = a(t_n), f^n = f(t_n) \\
&a^{n,\theta} = a(\theta t_n + (1 - \theta) t_{n+1}), 0 \leq \theta \leq 1, 0 \leq n \leq N - 1 \\
&u^{n,\theta} = \theta u^n + (1 - \theta) u^{n+1}, 0 \leq n \leq N - 1.
\end{aligned}
\tag{18}
$$

Not only do we have to approximate functions at mesh point but we also have to come up with a scheme to approximate the derivative appearing in (17). There are several possibilities and they are based on *divided differences*. For example, the following divided differences approximate the first derivative of u at the mesh point $t_n = n * k$;

$$\left.\begin{aligned} D_+u^n &\equiv \frac{u^{n+1}-u^n}{k} \\ D_-u^n &\equiv \frac{u^n-u^{n-1}}{k} \\ D_0u^n &\equiv \frac{u^{n+1}-u^{n-1}}{2k}. \end{aligned}\right\} \tag{19}$$

We now introduce a number of important and useful difference schemes that approximate the solution of (17). The main schemes are:

- Explicit Euler.
- Implicit Euler.
- Crank Nicolson (or box scheme).

The explicit Euler method is given by:

$$\begin{aligned} &\frac{u^{n+1}-u^n}{k} + a^n u^n = f^n,\ n = 0, \ldots, N-1 \\ &u^0 = A \end{aligned} \tag{20}$$

whereas the implicit Euler method is given by:

$$\begin{aligned} &\frac{u^{n+1}-u^n}{k} + a^{n+1}u^{n+1} = f^{n+1}, n = 0, \ldots, N-1 \\ &u^0 = A. \end{aligned} \tag{21}$$

Notice the difference: in (20) the solution at level $n+1$ can be directly calculated in terms of the solution at level n while in (21) we must rearrange terms in order to calculate the solution at level $n+1$. The next scheme is called the Crank-Nicolson or box scheme and it can be seen as an average of the explicit and implicit Euler schemes. It is given as:

$$\begin{aligned} &\frac{u^{n+1}-u^n}{k} + a^{n,\frac{1}{2}}\, u^{n,\frac{1}{2}} = f^{n,\frac{1}{2}},\ n = 0, \ldots, N-1 \\ &u^o = A \text{ where } u^{n,\frac{1}{2}} \equiv \tfrac{1}{2}(u^n + u^{n+1}). \end{aligned} \tag{22}$$

The discussion in this section has prepared us for a discussion of the Black-Scholes partial differential equation.

5. Example: The Black-Scholes PDE and its Approximation

Probably one of the most famous formulae in computational finance is due to Fischer Black, Myron Scholes and Robert Merton [1]. It has become popular with traders to price and hedge (a hedge is a trade to reduce risk) options.

We introduce the generalised Black Scholes formula to calculate the price of a call option on some underlying asset. In general the call price is a function of six parameters:

$$C = C(S, K, T, r, \sigma, t) \tag{23}$$

where the parameters have the following meaning [8]:

- S = asset price.
- K = strike (exercise) price.
- T = exercise (maturity) date.
- r = risk-free interest rate.
- σ = constant volatility.
- b = cost of carry.

We can view the call option price C as a function that maps a vector of parameters into a real value. The exact formula for C is given by:

$$C = Se^{(b-r)T}N(d_1) - Ke^{-rT}N(d_2) \tag{24}$$

where $N(x)$ is the standard cumulative normal (Gaussian) distribution function defined by

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy \tag{25}$$

and where

$$\begin{cases} d_1 = \dfrac{ln(S/K) \ + \ (b+\sigma^2/2)T}{\sigma\sqrt{T}} \\ d_2 = \dfrac{ln(S/K) \ + \ (b-\sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}. \end{cases} \tag{26}$$

The cost-of-carry parameter b has specific values depending on the kind of derivative security [8]:

- $b = r$, we have the Black and Scholes stock option model.
- $b = r - q$, the Morton model with continuous dividend yield q.
- $b = 0$, the Black futures option model.

- $b = r - R$, the Garman and Kohlhagen currency option model, where R is the foreign risk-free interest rate.

Thus, we can find the price of a plain call option by using formula (24).

The formula needs six input parameters, one of which (namely, the volatility) cannot be found from the market and then special methods must be employed to estimate it. A discussion of this problem is outside the scope of this article. Even though the assumptions upon which formula (24) are based do not hold in all practical cases (see Hull 2006 for a discussion) it is nonetheless the motivator for more general cases for which an analytical solution is not available. In these cases we must resort to numerical methods, for example using the finite difference method that approximates the so-called Black-Scholes PDE:

$$
\begin{aligned}
&LV \equiv -\frac{\partial V}{\partial t} + \sigma(S,t)\frac{\partial^2 V}{\partial S^2} + \mu(S,t)\frac{\partial V}{\partial S} + b(S,t)V \\
&\text{where} \\
&\sigma(S,t) = \tfrac{1}{2}\sigma^2 S^2 \\
&\mu(S,t) = rS \\
&b(S,t) = -r.
\end{aligned}
\tag{27}
$$

The corresponding fitted scheme is now defined as:

$$
\begin{aligned}
L_k^h V_j^n = &-\frac{V_j^{n+1} - V_j^n}{k} + \rho_j^{n+1} D_+ D_- V_j^{n+1} + \mu_j^{n+1} D_0 V_j^{n+1} \\
&+ b_j^{n+1} V_j^{n+1}, \\
\text{for} \quad & 1 \le j \le J-1, \text{ where} \\
\rho_j^n \equiv &\frac{\mu_j^n h}{2} \coth \frac{\mu_j^n h}{2\sigma_j^n}.
\end{aligned}
\tag{28}
$$

We define the discrete variants of the initial condition (15) and boundary conditions (16) and we realise them as follows:

$$
V_j^0 = \max(S_j - K, 0), \quad 1 \le j \le J-1 \tag{29}
$$

and

$$
\left.\begin{aligned} V_0^n &= g_0(t_n) \\ V_J^n &= g_1(t_n) \end{aligned}\right\} \quad 0 \le n \le N. \tag{30}
$$

The system (28), (29), (30) can be cast as a linear matrix system:

$$
A^n U^{n+1} = F^n, \quad n \ge 0 \text{ with } U^0 \text{ given} \tag{31}
$$

and we solve this system using LU decomposition, for example. A discussion of this topic with algorithms and implementation in C++ can be found in [3]. Summarising, the scheme (28) uses constant meshes in both space and time, centred differencing in space and backwards in time (fully implicit) marching. Furthermore, we use exponential fitting (see [2]) to ensure that the method remains stable and accurate for problems with small diffusion parameter or large convection parameter (This is the case of *convection dominance*). We note that equation (29) is the discrete payoff function for a call option. It plays the role of the discrete initial condition for the finite difference scheme (28), (29), (30).
Finally, we remark that scheme (28), (29), (30) is first-order accurate in space and time. For higher-order methods for one-factor and multi-factor Black Scholes PDEs, see [5] and [7].

6. Software Design and Implementation Issues

What happens when we have set up the system of equations (28), (29), (30)? In general, we implement the schemes in some modern object-oriented programming language, for example C++ or C# for use in production environments although languages such as Matlab and Mathematica are used for building and testing prototypes. Many pricing libraries have been developed during the last twenty-five years in C++ and its popularity can be attributed to the fact that it is an ISO standard and it is very efficient. It is a big language and the learning curve is steep.

A discussion of the software activities involved when designing software systems in computational finance is outside the scope of this article. See [3] for some applications to PDEs and to the finite difference method.

7. Conclusions and Future Scenarios: Computational Finance and Research Mathematics

We have written this article to show some of the mathematical, numerical and computational techniques that are used to price and hedge financial derivatives. We have focused on a small subset but important subset, namely the Black Scholes PDE and its numerical approximation using the finite difference method. There are many challenges and opportunities in this field in my opinion for applied and numerical mathematicians, computer scientists and engineers in the coming years as we enter an era of distributed and parallel computing.

References

[1] Black, F. and M. Scholes 1973 The pricing of options and corporate liabilities. Journal of Political Economy 81, 637-659.
[2] Duffy, D. J. 1980 Uniformly Convergent Difference Schemes for Problems with a Small Parameter in the Leading Derivative. PhD thesis. Trinity College Dublin.
[3] Duffy, D. J. 2004 Financial Instrument Pricing using C++. John Wiley and Sons. Chichester.
[4] Duffy, D. J. 2004A A critique of the Crank-Nicolson scheme, strengths and weaknesses for financial engineering. Wilmott Magazine. July 2004. pp. 68-76.
[5] Duffy, D. J. 2006 Finite difference methods in financial engineering. John Wiley and Sons. Chichester.
[6] Duffy, D.J. and Kienitz, J. 2009 Monte Carlo Frameworks, Building Customisable High Performance C++ Applications. John Wiley and Sons. Chichester.
[7] Duffy, D.J. and Germani, A. 2013 C# in Financial Markets. John Wiley and Sons. Chichester.
[8] Haug, E. 2007 The Complete Guide to Option Pricing Formulas. McGraw-Hill. New York.
[9] Hull J. 2006 Options, Futures and other Derivatives. Sixth Edition. Pearson. Upper Saddle River, New Jersey.
[10] Ilin, A.M, Kalashnikov, A.S. and Oleinik, O.A. 1962 Linear Equations of the Second Order of Parabolic Type, (translation) Russian Mathematical Surveys 17 (no. 3) 1-143.
[11] Wilmott, P. 2006 Paul Wilmott on Quantitative Finance. John Wiley and Sons. Chichester.

Daniel J. Duffy is founder of Datasim Financial, Amsterdam. He has been using C++ since 1989 and has a PhD in Numerical Analysis from Trinity College. His interests are in computation, mathematics and their applications.

Datasim Education BV, Schipluidenlaan 4, 1062 HE Amsterdam, The Netherlands
E-mail address: dduffy@datasim.nl

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 21–44
ISSN 0791-5578

INTEGRATION ISSUES IN PROBABILITY

PAT MULDOWNEY

Abstract. This essay explores the meaning of stochastic differential equations and stochastic integrals. It sets these subjects in a context of Riemann-Stieltjes integration. It is intended as a comment or supplement to [13].

1. Introduction

Famously, England and America are said to be divided by their common language. Similarly, mathematical analysts and probability theorists employ modes of expression which are superficially similar, but which may sometimes evoke different interpretations and connotations in each camp. This can be illustrated by the formula

$$\int_{-\infty}^{x} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\frac{dy}{\sigma\sqrt{2\pi}}.$$

To the analyst this expression may signify an improper indefinite integral, whereas the probabilist may see a cumulative normal distribution function. Aspects of the expression which are problematic or challenging to one may be trivially obvious to the other.

Another symptom is the probability/measure issue. In a kind of *coup d'etat* by mathematical analysis following the discoveries of A.N. Kolmogorov, the impression is sometimes given that the phenomenon of probability is now and forevermore to be understood in terms of the theory of measure.

But it is an overstatement to say that probability can be reduced to measure. Probability was a subject of interest long before modern measure theory existed, and there are aspects of random variation which are not amenable to explanation by the current methods of measure theory. On the other hand, an expert in probability is not,

2010 *Mathematics Subject Classification.* 65C30, 60H05.
Key words and phrases. Stochastic differential equation, Stochastic integral.
Received on 16-9-2014; revised 18-1-2015.

by virtue of that alone, an expert in the theory of measure. Neither subject encompasses the other.

This essay seeks to tease out some distinctive features of these two mathematical disciplines in respect of topics such as Itô's formula and stochastic calculus. In particular, it aims to demonstrate how the modern theory of integration can shed light on some challenging aspects of random variation.

Mathematical analysis does not "own" probability theory. But, after all these years, it can still contribute to it!

2. Itô's formula

Itô's formula is an example of a stochastic differential equation:

$$dY_s = \frac{\partial f(X_s)}{\partial s} ds + \frac{1}{2}\frac{\partial^2 f(X_s)}{\partial X_s^2} ds + \frac{\partial f(X_s)}{\partial X_s} dX_s. \tag{1}$$

For $Y_s = f(X_s)$, this formula is an evocative, shorthand way of writing

$$Y_t - Y_0 = \int_T dY_s = \int_T \frac{\partial f(X_s)}{\partial s} ds + \frac{1}{2}\int_T \frac{\partial^2 f(X_s)}{\partial X_s^2} ds + \int_T \frac{\partial f(X_s)}{\partial X_s} dX_s. \tag{2}$$

If the various expressions in this equation represented ordinary numbers and functions, then the presence in the equation of various integration processes might incline us to call (2) an integral equation.

But, while the symbol "f" in both equations is actually an ordinary deterministic function (such as the operation of taking the square of some operand), the symbols X and Y do not represent "ordinary" functions or definite numbers. Instead, they are "random variables", that is, quantities which are indefinite or unknown, to the extent that they can be predicted only within some margin of error.

The presence of "=" in the equation indicates that it is an exact statement about actual quantities. Itô's formula can be best regarded as an exact statement about margins of error in estimates of uncertain quantities or measurements.

In other words, it deals with probability distributions of unpredictable quantities which are obtained by means of various operations in the formula, such as the integration operation. So Itô's formula can be regarded as a kind of integral equation in which the integrals are the type known as *stochastic integrals*.

What is a stochastic integral? What is the meaning of Itô's formula? These questions are not trivial. They can be answered in a loose or intuitive manner, but deeper and more exact understanding can be challenging for non-specialists. And since they are a fundamental part of many important practical subjects, such as finance and communications, an understanding of them which is merely loose or hazy can be a serious barrier to competent practice in such subjects.

This essay seeks to outline an introduction to stochastic integrals which is less difficult than the standard textbook treatment of this subject. It uses Henstock's non-absolute integration instead of Lebesgue integration. It explores, compares, and contrasts these two methods of integration, with a view to assessing their role in stochastic integrals.

3. Random variables

Broadly speaking—at the risk of haziness and looseness!—a *random variable* is a mathematical representation of a measurement (an experiment, trial, or observation) of some uncertain or unpredictable occurrence or value. For instance, the random variable Z could represent a single throw of a die, so Z represents possible outcomes $\{z = 1, \dots, z = 6\}$ with probabilities $\{\frac{1}{6}, \dots, \frac{1}{6}\}$. Or it could represent measurement of a standard normal variable whose possible values are the real numbers $z \in \mathbf{R}$, with standard normal probability distribution $\mathbf{N}(0,1)$.

Suppose the throw of the die yields a payoff or outcome $y = f(z)$ obtained by the following deterministic calculation:

$$y = \begin{cases} -1 & \text{if } z = 1, \\ +1 & \text{if } z = 6, \\ 0 & \text{otherwise.} \end{cases}$$

This particular experiment or game depends on (is *contingent* on) the outcome of the experiment Z, and can be denoted by $Y = f(Z)$. Where Z has six possible outcomes, with a uniform probability distribution, Y has three possible outcomes whose probability distribution can easily be deduced by means of the deterministic calculation

f. The probability distribution[1] of Y is $y = -1$ with probability $\frac{1}{6}$, $y = +1$ with probability $\frac{1}{6}$, and $y = 0$ with probability $\frac{2}{3}$.

We can easily invent such *contingent* random variables or gambling games using more than one throw of the die, and with payoff Y dependent on some calculation based on the joint outcome of the successive throws.

This intuitive formulation is compatible with the formal and rigorous conception of a random variable as a P-measurable function whose domain is a P-measurable sample space Ω. This twentieth century injection of mathematical rigor by A.N. Kolmogorov and others brought about a great extension of the depth and scope of the theory of probability and random variation, including the development of many new spheres of application of the theory.

These applications often involve *stochastic processes*. Suppose T is some set of indexing elements $\{s\}$. For instance, T could be an interval of real numbers $[a, b]$. A stochastic process $Y = Y_T$ is a family $Y = (Y(s))_{s\in T}$, for which each element $Y(s) = Y_s$ is a random variable. A sample path $(y(s))_{s\in T}$ of the process $Y = Y_T$ can be thought of as a function $y : T \mapsto \mathbf{R}$ in which, for each s, $y(s)$ (or y_s) is a possible outcome of the random variable (measurement, experiment, trial) $Y(s)$.

4. Stochastic integrals

Take $T = [0, t]$. Equation (2) above appears to be the result of applying an integration operation $\int_T$ to the equation (1). If this is the case, and if this step is justified, then comparison of (1) and (2) implies (without delving into their actual meaning) that

$$\int_T dY_s = \int_0^t dY_s = Y_t - Y_0; \text{ or } \int_T dY(s) = \int_0^t dY(s) = Y(t) - Y(0). \tag{3}$$

[1]The probability distribution ("margin of error") carries the essential information specifying the character of the random variable or experiment. It is often convenient to include other "potential" values or outcomes which are **not** actually possible or "potential". For instance, in the die-throwing experiment we can declare that every real number is a potential outcome. In that case we assign probability zero to the impossible outcomes. This does not change the random variable or its probability distribution in any essential way that affects its mathematical meaning

On the face of it, a clear and precise understanding of this simplest of all possible stochastic integrals would seem to be the *sine qua non* of this subject. Expressed as a stochastic differential equation, it is the tautology $dY_s = dY_s$. Whatever (3) actually means, it seems consistent enough with more familiar forms of integration of the Stieltjes kind, in the somewhat loose and uncritical sense that the integral (or sum) of increments dY gives an overall increment.

Advancing a little bit further, take a deterministic function f, and consider $\int_T f(Y_s)dY_s$ (or $\int_T f(Y(s))dY(s)$), which is a more general version of $\int_T dY_s$. If y is a sample path of the process Y, the expression

$$\int_T f(y(s))dy(s) \quad \text{or} \quad \int_T f(y_s)dy_s \tag{4}$$

is a Stieltjes-type integral, which, if it exists, may be thought of as some limit of Riemann sums

$$\sum f(y(s))\Delta y(s) \text{ or } \sum f(y(s_j))\left(y(t_j) - y(t_{j-1})\right),$$

where the finite set of points t_j form a partition of the interval $T = [0, t]$, with $t_{j-1} \leq s_j \leq t_j$ for each j.

From the point of view of basic mathematical analysis, unlike (3) which is about "margins of error" in probabilistic measurement, there is nothing problematic about (4)—this Riemann-Stieltjes-type integral may or may not exist for particular functions y and f, but it is a fairly familiar subject for anyone who has studied basic Riemann-type integration.

In the Riemann sums for (4), some applications require that $s_j = t_{j-1}$ for each j. Cauchy's approach to the theory of integration used approximating sums with $s_j = t_{j-1}$ or $s_j = t_j$, so such sums can be called *Cauchy sums* rather than Riemann sums. In any event, there are various ways, including the Lebesgue method, in which we can seek to define an integral $\int_T f(y(s))dy(s)$ for sample paths $y_T = (y(s))_{s\in T}$ of a stochastic process $Y = Y_T$.

Suppose a Stieltjes-type integral of $f(y(s))$ is calculated with respect to the increments $y(I) := y(t_j) - y(t_{j-1})$ of the function y_T. For instance, if f is a function taking some fixed, real, constant value such as 1, then a "naive" Riemann sum calculation on the domain $T = [0, t]$, with $t_0 = 0$ and $t_n = 1$ gives

$$\sum f(y(s))y(I) \quad = \quad \sum_{j=1}^{n} y(I) \quad =$$

$$((y(t_1) - y(0)) + ((y(t_2) - y(t_1)) + \cdots + ((y(1) - y(t_{n-1})) = y(1) - y(0)$$

for **every** sample outcome y_T of the process Y_T. So it is reasonable—in some "naive" way—to claim that, for this particular function f, the Riemann-Stieltjes integral exists for all outcomes y_T:

$$\int_T f(y_s) dy_s = \int_0^t dy(s) = y(t) - y(0).$$

One might then be tempted[2] to apply such an argument to step functions f, and perhaps to try to extend it to some class of continuous functions f, especially if we are only concerned with sample paths y_T which are continuous.

But the key point here is that, given a stochastic process $Y = Y_T$, and given certain deterministic functions f, real values $\int_T f(y(s))dy(s)$ can be obtained for each sample path $y = y_T$ by means of a recognizable Stieltjes integration procedure.

Can this class of real numbers or outcomes be related somehow to some identifiable random variable Z which possesses some identifiable probability distribution (or "margin of error" estimates)?

If so, then Z might reasonably be considered to be the random variable obtained by integrating, in some Stieltjes fashion, the random variable $f(Y_s)$ with respect to the increments $Y(I) = Y(t_j) - Y(t_{j-1})$ of the stochastic process Y_T.

In other words, Z is the stochastic integral $\int_T f(Y_s)dY_s$.

To justify the latter step, a probability distribution (or "margin of error" data) for Z must be determined. But, in the case of the constant function f given above ($f(y_s) = 1$), this is straightforward. Because, with $f(y_s) = 1$ for all outcomes y_s in all sample paths (or joint outcomes) y_T, the distribution function obtained for the Riemann sum values $\sum f(y_s)y(I)$ is simply the known distribution function of the outcomes $y(t) - y(0)$ of the random variable $Y(t) - Y(0)$.

This distribution is the same for all partitions of $T = [0, t]$. So it is reasonable to take it to be the distribution function of the stochastic integral $Z = \int_T f(Y_s)dY_s$. For constant f this seems to provide meaning and rationale for (4).

[2]A warning against this temptation is provided in Example 8.3 below.

What this amounts to is a naive or intuitive interpretation of stochastic integration which seems to hold for some elementary functions f. This approach can be pursued further to give a straightforward interpretation—indeed, a "proof"—of Itô's formula, at least for the unchallenging functions f mentioned above.

But what of the standard or rigorous theory of stochastic integration?

5. Standard theory of stochastic integration

Unfortunately, this theory cannot accommodate the naive or intuitive construction of the simple stochastic integrals described in the preceding section. Broadly speaking, the elementary Riemann sum type of calculation is not adequate for the kinds of analysis needed in this subject. It is not possible, for instance, to apply a monotone convergence theorem, or a dominated convergence theorem, to simple Riemann and Riemann-Stieltjes integrals. Historically, these kinds of analysis and proof have been supplied by Lebesgue-type integrals which, while requiring a measure function as integrator, cannot be simply defined by means of the usual arrangement[3] of Riemann sums.

And this is where the difficulty is located. Suppose, for instance, that the stochastic process Y_T that we are dealing with is a standard Brownian motion. In that case any sample path y_T is, on the one hand, almost surely continuous—which is "nice"; but, on the other hand, it is almost surely *not* of bounded variation in every interval J of the domain $T = [0, 1]$. And the latter is "nasty".

This turns out to be very troublesome if we wish to construct a Lebesgue-Stieltjes integral using the increments $y(I) = y(t_j) - y(t_{j-1})$ of a sample path which is continuous but not of bounded variation in any interval.

The problem is that, in order to construct a Lebesgue-Stieltjes measure from the increments $y(I)$, we must separate the non-negative increments $y_+(I)$ from the negative-valued increments $y_-(I)$,

$$y(I) = y_+(I) - |y_-(I)|,$$

and try to construct a non-negative measure from each of the components. But, because y is not of bounded variation, the construction

[3]But Section 8 shows that Lebesgue integrals are essentially Riemann-Stieltjes integrals.

for each component diverges to infinity on every interval J. Thus the standard theory of stochastic integration encounters a significant difficulty at the very first step (4).

To summarize:

- In the standard Itô or Lebesgue integral approach, the most basic calculation of the integral of a constant function $f(Y_T)$, with respect to the increments dY of a Brownian process, fails because the Lebesgue-Stieltjes measure does not exist.
- On the other hand, if Riemann sums of the increments of the process Y_T are used, then, by cancellation, a finite result is obtained for each Riemann sum—a result which agrees with what is intuitively expected.

In the standard Lebesgue (or Itô) theory of stochastic integration—in [16] for instance—this problem is evaded by *postulating* a finite measure $\mu_y(J)$ for each sample path, and then constructing a weak form of integral which, in the case of Brownian motion, is based on certain helpful properties of this process.

The trouble with this approach is that it produces a quite difficult theory which does not lend itself to the natural, intuitive interpretation described above.

However, elementary Riemann-sum-based integration is not generally considered to have the analytical power possessed by Lebesgue-style integration. And a great deal of analytical power is required in the theory of stochastic processes. So at first sight it seems that we are stuck with the standard theory of stochastic integration, along with all its baggage of subtlety and complication.

But this is not really the case. The good news is that is actually possible to formulate the theory of stochastic integrals using Riemann sums instead of the measures of Lebesgue theory.

6. Integration of functions

To see this, it is first necessary to review the various kinds of integration which are available to us.

First consider the basic Riemann integral, $\int_a^b f(s)ds$, of a real-valued, bounded, continuous function $f(s)$ on an interval $[a, b]$. Let $\mathcal{P}$ be a partition of $[a, b]$;

$$\mathcal{P}: \quad a = t_0 < t_1 < t_2 < \cdots < t_n = b,$$

for any choice of positive integer n and any choice of t_j, $1 \leq j < n$. For any $u < v$ and any interval I with end-points u and v, write $|I| = v - u$. Denoting intervals $]t_{j-1}, t_j]$ by I_j let

$$U_{\mathcal{P}} = \sum_{j=1}^{n} P_j |I_j|, \qquad L_{\mathcal{P}} = \sum_{j=1}^{n} p_j |I_j|$$

where

$$P_j = \sup\{f(s) : s \in I_j\}, \qquad p_j = \inf\{f(s) : s \in I_j\}.$$

Definition 6.1. Define the *upper Riemann integral* of f by

$$U := \inf\{L_{\mathcal{P}} : \text{ all partitions } \mathcal{P} \text{ of } [a, b]\},$$

and the *lower Riemann integral* of f by

$$L := \sup\{l_{\mathcal{P}} : \text{ all partitions } \mathcal{P} \text{ of } [a, b]\}.$$

Then $U_{\mathcal{P}} \geq L_{\mathcal{P}}$ for all $\mathcal{P}$, and if $U = L$ we say that f is *Riemann integrable*, with

$$\int_a^b f(s) ds := U = L.$$

Write the partition $\mathcal{P}$ as $\{I\}$ where each I has the form $I_j =]t_{j-1}, t_j]$, with $|I_j| = t_j - t_{j-1}$, and Riemann sum

$$(\mathcal{P}) \sum f(s)|I| = \sum_{j=1}^{n} f(s_j)|I_j|.$$

Suppose $g(s)$ is a real-valued, monotone increasing function of $s \in [a, b]$, so $g(s) \geq g(s')$ for $s > s'$. For any interval I with end-points u and v $(u < v)$, define the increment or interval function $g(I)$ to be $g(v) - g(u)$.

Definition 6.2. If $|I|$ and $|I_j|$ are replaced by $g(I)$ and $g(I_j)$ in Definition 6.1 of the Riemann integral, then the resulting integral is called the *Riemann-Stieltjes integral* of f with respect to g, $\int_a^b f \, dg$ or $\int_a^b f(s) dg(s)$.

In fact if we start with the latter definition the Riemann integral is a special case of it, obtained by taking the point function g to be the identity function $g(s) = s$.

If $g(s)$ has bounded variation it can be expressed as the difference of two monotone increasing, non-negative point functions,

$$g(s) = g_+(s) - (-g_-(s)),$$

and the Riemann-Stieltjes integral of f with respect to g can then be defined as the difference of the Riemann-Stieltjes integrals of f with respect to g_+ and $-g_-$, respectively.

The following result is well known: if real-valued, bounded f is continuous and if real-valued g has bounded variation then $\int_a^b f\,dg$ exists.

As suggested earlier, the Lebesgue integral of a real-valued point function k with respect to a measure μ can be viewed, essentially, as a Riemann-Stieltjes integral in which the point-integrand $k(\omega)$ satisfies the condition of *measurability*. To explain this statement further, consider a measure space $(\Omega, \mathcal{A}, \mu)$ with non-negative measure μ on a sigma-algebra $\mathcal{A}$ of μ-measurable subsets of the arbitrary measurable space Ω. Thus, if $\mu(\Omega) = 1$, the measure space is a probability space. Suppose the point-integrand k is a bounded real-valued μ-measurable function on the domain Ω. Then there exist real numbers c and d for which

$$c \leq k(\omega) \leq d \quad \text{for all} \quad \omega \in \Omega.$$

Also, for each sub-interval J of $[c, d]$, measurability of k implies $\mu(k^{-1}(J))$ is defined. The basic definition of the Lebesgue integral of k with respect to μ on Ω is as follows.

Definition 6.3. Let $\mathcal{Q} = \{J_j\} = \{]v_{j-1}, v_j]\}$ be a partition of $[c, d]$,

$$\mathcal{Q}: \quad c = v_0 < v_1 < v_2 < \cdots < v_n = d,$$

and let

$$L_{\mathcal{Q}} = \sum_{j=1}^{n} v_{j-1}\mu(k^{-1}(J_j)), \qquad U_{\mathcal{Q}} = \sum_{j=1}^{n} v_j\mu(k^{-1}(J_j)).$$

Let $L := \sup\{L_{\mathcal{Q}} : \mathcal{Q}\}$, $U := \inf\{U_{\mathcal{Q}} : \mathcal{Q}\}$, the supremum and infimum being taken over all partitions $\mathcal{Q}$ of $[c, d]$. If $L = U$, then their common value is the *Lebesgue integral* $\int_\Omega k(\omega)d\mu$.

An advantage of Lebesgue integration over Riemann integration is that the former has theorems, such as the dominated and monotone convergence theorems which, under certain condition, make it possible for instance to change the order of integration and differentiation. Also, Fubini's and Tonelli's theorems allow exchange of order of multiple integrals.

What makes "good" properties such as these possible is *measurability* of the integrand k. But the Lebesgue integral itself is, by definition, a Riemann-Stieltjes-type integral. To see this, for each $u \in [c,d]$ define the monotone increasing function

$$g(u) = \mu\left(k^{-1}([c,u])\right), \tag{5}$$

and take the point function $h(u)$ to be the identity function $h(u) = u$. Then the construction[4] in Definition 6.3 shows that

$$\int_{\Omega} k(\omega) d\mu \quad = \quad \int_c^d h(u)\, dg(u), \quad = \quad \int_c^d u\, dg. \tag{6}$$

In other words, when combined with the measurability property of the point-integrand, this particular Riemann-Stieltjes construction gives the "good" properties required in the integration of functions.

7. Riemann definition

But in fact a Riemann construction can give these "good" properties **without** postulating measurability in the definition[5] of the integral. To see this, we start again by considering a more general and more flexible definition of basic Riemann and Riemann-Stieltjes integration which generalizes the construction of these integrals as given above in Definitions 6.1 and 6.2.

The proposed, more general, definition of the Riemann-Stieltjes integral is applicable to real- or complex-valued functions f (bounded or not); and to real- or complex-valued functions g, with or without bounded variation.

Definition 7.1. The function f is Riemann-Stieltjes integrable with respect to g, with integral α, if, given $\varepsilon > 0$, there exists a constant $\delta > 0$ such that, for every partition $\mathcal{P} = \{I\}$ of $[a,b]$ satisfying $|I| < \delta$ for each $I \in \mathcal{P}$, the corresponding Riemann sum satisfies

$$\left|\alpha - (\mathcal{P})\sum f(s)g(I)\right| < \varepsilon,$$

so $\alpha = \int_a^b f\, dg$.

[4]The integral of a point function $h(u)$ with respect to a point function $g(u)$ can be addressed either as a Riemann-Stieltjes construction or as a Lebesgue-Stieltjes construction. When $h(u) = u$ and $g(u) = \mu\left(k^{-1}([c,u])\right)$ the former approach gives the Lebesgue integral $\int_\Omega k(\omega)d\mu$. On the other hand, if the Lebesgue-Stieltjes construction is attempted with $h(u) = u$ and $g(u) = \mu\left(k^{-1}([c,u])\right)$, we simply replicate the Riemann-Stieltjes construction of the Lebesgue integral $\int_\Omega k(\omega)d\mu$, and nothing new emerges.

[5]And if measurability is redundant in the definition, then so is the measure space structure.

If g is the identity function $g(s) = s$ then Definition 7.1 reduces to the ordinary Riemann integral of f, $\int_a^b f(s)ds$.

Definition 7.1 does not embody conditions which ensure the existence of the integral. Such integrability conditions are not postulated but are deduced, in the form of theorems, from the definition of the integral.

Thus, if the function properties specified, respectively, in Definitions 6.1, 6.2, and 6.3 above are assumed, the integrability in each case follows from Definition 7.1; and Definitions 6.1, 6.2, and 6.3 become theorems of Riemann, Riemann-Stieltjes, and Lebesgue integration, respectively.

Definition 6.3 can now be expressed in terms of Definition 7.1, using the formulations (5) and (6), and assuming measurability of the integrand f with respect to measure space $(\Omega, \mathcal{A}, \mu)$.

Definition 7.2. The function f is Lebesgue integrable with respect to measure μ, with integral $\int_\Omega f(\omega)d\mu = \alpha$, if, given $\varepsilon > 0$, there exists a constant $\delta > 0$ such that, for every partition $\mathcal{Q} = \{J\}$ of $[c, d]$ satisfying $|J| < \delta$ for each $J \in \mathcal{Q}$, the corresponding Riemann sum satisfies

$$\left|\alpha - (\mathcal{Q})\sum h(u)g(J)\right| < \varepsilon,$$

where $h(u) = u$ is the identity function on $[c, d]$; so $\alpha = \int_c^d h(u)dg(u)$ $= \int_c^d u\, dg$.

Thus, by definition, the Lebesgue integral $\int_\Omega f(\omega)d\mu$, with domain Ω, is the Riemann-Stieltjes integral $\int_c^d u\, dg$, with domain $[c, d]$.

The following result is an obvious consequence of Definition 7.1. If f has constant value β and if g is an arbitrary real- or complex-valued function, then $\int_a^b f\, dg$ exists and equals $\beta(g(b) - g(a))$. This follows directly from Definition 7.1 since, for every partition $\mathcal{P}$ of $[a, b]$, cancellation of terms gives

$$(\mathcal{P})\sum f(s)g(I) = \beta\sum_{j=1}^{n}(g(t_j) - g(t_{j-1})) = \beta\left(g(b) - g(a)\right).$$

This result does not in general hold for Lebesgue-Stieltjes integration, as the latter requires that $g(s)$ be resolved into its negative and non-negative components, $g(s) = g_+(s) - (-g_-(s))$, and convergence may fail when the integral is calculated with respect to each of these components separately.

Example 8.3 below shows that, though constant functions f are Riemann-Stieltjes integrable with respect to any integrator function g, this does not necessarily extend to step functions f.

Definition 7.1 of the Riemann or Riemann-Stieltjes integral does not postulate any boundedness, continuity, measurability or other conditions for the integrand f. But, as already stated, in the absence of integrand measurability and the construction in Definition 6.3, this method of integration does not deliver good versions of monotone and dominated convergence theorems, or Fubini's theorem.

8. -Complete integration

Developments in the subject since the 1950's—developments which were originated independently by R. Henstock and J. Kurzweil—have made good this deficit in the basic Riemann and Riemann-Stieltjes construction. In this new development of the subject, Definition 7.1 of the Riemann-Stieltjes integral is amended as follows.

Definition 8.1. A function f is *Stieltjes-complete* integrable with respect to a function g, with integral α if, given $\varepsilon > 0$, there exists a function $\delta(s) > 0$ such that

$$\left|\alpha - (\mathcal{P})\sum f(s)g(I)\right| < \varepsilon$$

for every partition $\mathcal{P}$ such that, in each term $f(s)g(I)$ of the Riemann sum, we have $s - \delta(s) < u \leq s \leq v < s + \delta(s)$, where u and v are the end-points of the partitioning interval I.

In other words, where $|I|$ is less than a constant δ in the basic Riemann-Stieltjes definition, we have $|I| < \delta(s)$ in the new definition. Write $\alpha = \int_{[a,b]} f(s)g(I)$, or $\int_{[a,b]} f \;\, dg$, for the Stieltjes-complete integral whenever it exists.

Again, if the integrator function g is the identity function $g(s) = s$, the resulting integral (corresponding to the basic Riemann integral), is the *Riemann-complete* integral of f, written $\alpha = \int_{[a,b]} f(s)|I|$, or $\int_{[a,b]} f(s)ds$. The latter is also known as the Henstock integral, the Kurzweil integral, the Henstock-Kurzweil, the generalized Riemann integral, or the *gauge* integral since in this context the function $\delta(s) > 0$ is called a gauge.

It is obvious that every Riemann (Riemann-Stieltjes) integrable integrand is also Riemann-complete (Stieltjes-complete) integrable, as the gauge function $\delta(s) > 0$ of Definition 7.1 can be taken to be the constant $\delta > 0$ of Definition 6.1 and Definition 6.2.

This argument indicates a *Lebesgue-complete* extension of the Lebesgue integral, by replacing the constant $\delta > 0$ of Definition 7.2 with a variable gauge $\delta(u) > 0$:

Definition 8.2. Let $h(u) = u$ be the identity function on $[c, d]$. The function f is **Lebesgue-complete** integrable with respect to measure μ, with integral $\int_\Omega f \, d\mu = \alpha$, if, given $\varepsilon > 0$, there exists a gauge $\delta(u) > 0$ for $c \leq u \leq d$, such that

$$\left| \alpha - (\mathcal{Q}) \sum h(u) g(J) \right| < \varepsilon,$$

for every partition $\mathcal{Q} = \{J\}$ of $[c, d]$ satisfying

$$u - \delta(u) < v_{j-1} \leq u \leq v_j < u + \delta(u)$$

for each $J =]v_{j-1}, v_j] \in \mathcal{Q}$.

In that case $\alpha = \int_{[c,d]} h(u)g(J) = \int_{[c,d]} u \; g(J)$, and the Lebesgue-complete integral is a special case of the Stieltjes-complete integral—a special case in which a measure space structure exists and for which the integrand is measurable. So it is again clear that every Lebesgue integrable integrand is Lebesgue-complete integrable; since the former is, in effect, a Riemann-Stieltjes integral, the latter is a Stieltjes-complete integral, and every Riemann-Stieltjes integrable function is also Stieltjes-complete integrable. (No special notation has been introduced here to distinguish the Lebesgue integral $\int_\Omega f \; d\mu$ from its Lebesgue-complete counterpart.)

If the measurable domain Ω is a real interval such as $[a, b]$, then some ambiguity arises in the interpretation of the Lebesgue integral as an integral of the gauge, or generalized Riemann, kind. The reason for the ambiguity is as follows. Assuming the existence of the Lebesgue integral $\int_\Omega f(\omega) d\mu, = \int_{[a,b]} f(\omega) d\mu$, where ω now represents real numbers in the domain $[a, b]$, then we are assured of the existence of the Stieltjes and Stieltjes-complete (or Lebesgue-complete) integrals $\int_c^d u \; dg$ and $\int_{[c,d]} u \; g(J)$, respectively, with

$$\int_{[a,b]} f(\omega) d\mu = \int_c^d u \; dg = \int_{[c,d]} u \; g(J),$$

where the values $u = h(u)$ are elements of $[c, d]$ and h is the identity function on $[c, d]$.

But in this case, letting $\omega = s$ denote points of the domain $[a,b]$ and with I denoting subintervals of $[a,b]$, the function $\mu(I)$ is defined on intervals I, and two different Stieltjes-type constructions are possible.

First, there is the Riemann-Stieltjes integral $\int_c^d u\, dg$ which defines the Lebesgue integral $\int_\Omega f(\omega)d\mu, = \int_{[a,b]} f(\omega)d\mu$. Secondly, there is the gauge integral $\int_{[a,b]} f(s)\mu(I)$ which has a Stieltjes-complete construction.

It is then meaningful to consider whether, with f measurable, existence of the Lebesgue integral $\int_{[a,b]} f(\omega)d\mu$ implies existence of the Stieltjes-complete integral $\int_{[a,b]} f(s)\mu(I)$, and whether

$$\int_c^d u\, dg = \int_{[a,b]} f(s)\mu(I)$$

holds,[6] the first of these integrals being the Lebesgue integral $\int_{[a,b]} f(\omega)d\mu$, which, by Definition 7.2, is interpreted as the Riemann-Stieltjes integral $\int_c^d u\, dg$.

To see that these two integrals coincide, take f to be a bounded, measurable function on $[a,b]$. This can be expressed as the difference of two non-negative, bounded, measurable functions f_+ and f_-. Accordingly, and without loss of generality, take f to be non-negative, bounded, measurable. Then the Lebesgue integrable function f is the μ-almost everywhere point-wise limit of a monotone increasing sequence of step functions f_j. With $\omega = s$, each step function f_j is Lebesgue integrable, with Lebesgue integral $\int_{[a,b]} f_j(\omega)\ d\mu$; and each step function f_j is Stieltjes-complete integrable, with Stieltjes-complete integral $\int_{[a,b]} f_j(s)\mu(I)$, and

$$\int_{[a,b]} f_j(\omega)\ d\mu = \int_{[a,b]} f_j(s)\mu(I)$$

for each j. (This statement is also true if "Lebesgue integral" and "Lebesgue integrability" are replaced by "Lebesgue-complete integral" and "Lebesgue-complete integrability".)

[6]There is a considerable literature on this question, which is usually answered as: *"Every Lebesgue integrable function on an interval of the real numbers* $\mathbf{R}$ *is also Henstock-Kurzweil integrable."* If the domain of the integrand is a measurable space Ω which is **not** a subset of $\mathbf{R}$ or $\mathbf{R}^n$, then the appropriate way to formulate the corresponding Henstock-Kurzweil (or -complete) integral is in the form $\int_{[c,d]} u\, g(J)$ described in Definition 8.2.

By the monotone convergence theorem of Lebesgue integration (or, respectively, by the monotone convergence theorem of Lebesgue-complete integration),

$$\int_{[a,b]} f_j(\omega)d\mu \to \int_{[a,b]} f(\omega)d\mu$$

as $j \to \infty$. By the monotone convergence theorem of Stieltjes-complete integration, $f(s)\mu(I)$ is Stieltjes-complete integrable and

$$\int_{[a,b]} f_j(s)\mu(I) \to \int_{[a,b]} f(s)\mu(I)$$

as $j \to \infty$. Since corresponding integrals of the pair of sequences are equal, their limits are equal:

$$\int_{[a,b]} f(\omega)d\mu = \int_{[a,b]} f(s)\mu(I).$$

This is the gist of a proof that existence of a Lebesgue integral (or of a Lebesgue integral) on a real domain implies existence of the corresponding Stieltjes-complete integral on the same domain, and equality of the two.

Thus the above argument can be applied to either the Lebesgue or the Lebesgue-complete integral on $\Omega = [a, b]$ in conjunction, respectively, with the corresponding Stieltjes-complete integral on the same domain. In effect, if the domain Ω is a subset of $\mathbf{R}$, and if f is Lebesgue integrable or Lebesgue-complete integrable with respect to μ, then $f(s)\mu(I)$ is also Stieltjes-complete integrable and the two integrals are equal.

The specific properties of the Lebesgue-complete integral have not been investigated.

As mentioned earlier, constant functions f are Riemann-Stieltjes integrable, and hence Stieltjes-complete integrable, with respect to any integrator function g. But as the following counter-example shows, this does not necessarily extend to step functions f, or any other functions which are not constant.

Example 8.3. Dirichlet function: For $0 \leq s \leq 1$ let $D(s)$ be 1 if s is rational, and 0 otherwise. For $I =]u, v]$ let $D(I) = D(v) - D(u)$. Let $D([0, v]) = D(v) - D(0)$. The point function $D(s)$ is discontinuous everywhere, and has infinite variation on every interval $J \subseteq [0, 1]$. If $f(s)$ is constant for $0 \leq s \leq 1$, then the Riemann-Stieltjes integral $\int_0^1 f(s)\, dD$ exists and equals $D(1) - D(0)$; that is,

$\int_0^1 f(s)\,dD = 0$. **But if f is not constant on $[0,1]$, then the Riemann-Stieltjes integral of f with respect to D does not exist.** What about Stieltjes-complete integrability of $f(s)D(I)$? In fact, if f is not constant on $[0,1]$, then the Stieltjes-complete integral of f with respect to D does not exist. This is proved in Theorem 1 of [6], and the proof is reproduced in Theorem 67 of [13]. **Thus $f(s)D(I)$ is Riemann-Stieltjes integrable and Stieltjes-complete integrable on $[0,1]$ if and only if $f(s)$ is constant for $0 \leq s \leq 1$.**

Historically this is the first published result (Theorem 1 of [6]) in the theory of -complete integration.

9. -Complete approach to stochastic integrals

Returning to stochastic integrals, the -complete method of integration allows us to construct Stieltjes-type Riemann sums for highly oscillatory expressions which include both positive and negative terms. Cancellation of terms can occur in the Riemann sum approximations, so the possibility of convergence is preserved by this construction.

The Lebesgue construction, on the other hand, requires integral convergence, separately and independently, of the positive and negative components of the integrand. The difficulty this presents is illustrated in the alternating or oscillating series $\sum_{j=1}^{\infty}(-1)^j j^{-1}$. If the positive and negative terms of the series are considered as two separate series then each of them diverges. But the series itself is conditionally (or non-absolutely) convergent. Similarly, for sample paths $y(s)$ of a stochastic process Y_T the integral $\int_{[0,t]} dy(s)$ does not generally exist when considered as a Lebesgue-Stieltjes integral. But it exists for all sample paths y_T, with value $y(t) - y(0)$, when considered as a Stieltjes-complete integral.

There is no analytical cost or disadvantage in relinquishing the Lebesgue construction in favor of the -complete method. This is because the important theorems of Lebesgue integration, such as monotone and dominated convergence, are also valid for the -complete approach. Furthermore, there are other convergence theorems of a similar kind, specifically designed to deal with highly oscillatory functions such as those which occur in the theory of stochastic processes but which are beyond the scope of the Lebesgue method. See [13] for details of these.

However, stochastic integration includes novelties and challenges which have not yet been addressed in this essay.

For Brownian motion processes X_T, one of the most important stochastic integrals is $\int_0^t dX_s^2 = t$. The corresponding integral for a sample path $x(s)$ $(0 \leq s \leq t)$ is "$\int_0^t (dx(s))^2$". But this expression does not have the familiar form of a Stieltjes-type integral: $\int_a^b f(s)dg$, which, when g is the identity function, reduces to the even more familiar $\int_a^b f(s)ds$.

In Riemann sum approximation we are dealing with expressions $\sum(x(I))^2$, where, for $I =]u, v]$, $x(I) = x(v) - x(u)$. But traditionally, while a Riemann sum for a Stieltjes integral involves terms $f(s)x(I)$ with integrator function $x(I)$ (in which $f(s)$ can be identically 1), we do not usually expect to see integrators such as $(x(I))^2$ or dX_s^2.

Another important stochastic integral in Brownian motion theory is

$$\int_0^t X_s dX_s = \frac{1}{2}X_t^2 - \frac{1}{2}t.$$

For a sample path $x(s)$ of Brownian motion, this involves $\int_0^t x(s)dx(s)$, or, in Riemann sum terms, $\sum x(s)x(I)$. The latter, as it stands, is a finite sum of terms $x(s)(x(v)-x(u))$ where $I =]u, v]$ and $u \leq s \leq v$. And if we are using the Stieltjes-complete approach as described above, then we might suppose that each s in the Riemann sum is the special point used in partitions which are constrained by a gauge $\delta(s)$,

$$s - \delta(s) < u \leq s \leq v < s + \delta(s).$$

But in fact this is not what is required in the stochastic integral $\int_0^t X_s dX_s$. In Riemann sum format, what is required is

$$\sum x(u)x(I), \quad \text{or} \quad \sum x(u)\left(x(v) - x(u)\right),$$

where the first factor $x(u)$ in the integrand is a point function evaluated at the left hand end-point u of the interval $I =]u, v]$.

Sometimes the form $\sum x(w)(x(v) - x(u))$ is used, with $w = u + \frac{1}{2}(v - u)$.

In a way, integrands of form $x(I)^2$, $x(u)x(I)$, or $x(w)x(I)$, are an unexpected innovation. Their value is calculated from the numbers

u and v which specify the interval I. So they can be thought of as functions $h(I)$ of intervals I.

But these functions are **not additive**[7] on intervals. In that regard they are unlike the integrators $|I|$ and $x(I)$ which are themselves functions of I but are finitely additive on intervals, in the sense that, if $J = I_1 \cup \ldots \cup I_n$ is an interval, then

$$|J| = \sum_{j=1}^{n} |I_j|, \qquad x(J) = \sum_{j=1}^{n} x(I_j).$$

Broadly speaking, integration is a summation process in which the summed terms involve functions of intervals. Up to this point in this essay, the only integrands to be considered included a factor which was an additive function of intervals I, such as the length function $|I|$ or the Stieltjes-type functions $g(I)$ or $x(I)$. But there is nothing inherent in the definition of -complete integrals that requires any I-dependent factor in the integrand to be additive.

With this in mind, consider again the definition of the -complete integral on an interval $[a, b]$.

Firstly, a *gauge* is a function $\delta(s) > 0$, $a \leq s \leq b$. Given s, an interval $I =]u, v]$ for which s is either an end-point or an interior point, is $\delta(s)$-*fine* if $s - u < \delta(s)$ and $v - s < \delta(s)$. A finite collection $\mathcal{D} = \{(s_1, I_1), \ldots, (s_n, I_n)\}$ is a *division* of $[a, b]$ if each s_j is either an interior point or end-point of I_j and the intervals I_j form a partition of $[a, b]$. Given a gauge δ, a division $\mathcal{D}$ is δ-fine if each $(s_j, I_j) \in \mathcal{D}$ is δ-fine.

Now suppose h is a function of elements (s, I). Examples include: $h(s, I) = h_1(I) = |I|$, $h(s, I) = h_2(s) = s$, $h_3(s, I) = s^2|I|$, $h_4(I) = |I|^2$. Given a division $\mathcal{D} = \{(s, I)\}$ of $[a, b]$ whose intervals I form a partition $\mathcal{P}$, the corresponding Riemann sum is

$$(\mathcal{D}) \sum h(s, I), = \sum \{h(s, I) : I \in \mathcal{P}\}.$$

Definition 9.1. A function $h(s, I)$ is integrable on $[a, b]$, with integral $\int_{[a,b]} h(s, I) = \alpha$, if, given $\varepsilon > 0$. there exists a gauge $\delta(s) > 0$ so that, for each δ-fine division $\mathcal{D}$ of $[a, b]$,

$$\left| \alpha - (\mathcal{D}) \sum h(s, I) \right| < \varepsilon.$$

[7]If $h(I)$ were finitely additive on intervals I it could be used to define a point function $h(s) := h([0, s])$, and vice versa. Integrals with respect to finitely additive integrators are therefore representable as Stieltjes-type integrals, and vice versa.

Applying this definition to the examples, h_1 is integrable with integral $b-a$, h_2 is not integrable, h_3 is integrable with integral $\frac{1}{3}(b^3-a^3)$, and h_4 is integrable with integral 0. If $h(s,I)=h_5(I)=u^2|I|$ where, for each I, u is the left hand end-point of I, then it is not too hard to show that h_5 is integrable with integral $\frac{1}{3}(b^3-a^3)$.

Actually, it is the traditional custom and practice in this subject to only consider integrands $h(s,I)=f(s)p(I)$ where the integrator function $p(I)$ is a measure function or, at least, finitely additive on intervals I; and where the evaluation point s of the point function integrand $f(s)$ is the point s of (s,I) for each $(s,I)\in\mathcal{D}$. When $p(I)=|I|$, this convention is needed in order to prove the Fundamental Theorem of Calculus.[8].

But, while the Fundamental Theorem of Calculus is important in subjects such as differential equations, it hardly figures at all in some other branches of mathematics such as probability theory or stochastic processes. And we have seen that stochastic integration often requires point integrands $f(s)$ to be evaluated, not at the points s of $(s,I)\in\mathcal{D}$, but at the left hand end-points of the partitioning intervals I.

So, with $I=]u,v]$, $f(u)$ is, in fact, an integrand function which depends, not on points s but on intervals $]u,v]$.

These are a few of the “unexpected innovations” to be encountered in stochastic integration, giving it a somewhat alien and counter-intuitive feel to anyone versed in the traditional methods of calculus. Indeed, these are further examples of probability and analysis losing contact with each other.

For instance, the stochastic integral $\int_0^t XdX$ is given the value $\frac{1}{2}X(t)^2-\frac{1}{2}t$ when the process $X(s)$ (with $X(0)=0$) is a Brownian motion. Introductory treatments of this problem sometimes contrast the expression $\int_0^t XdX$ with the elementary calculus integral $\int xdx$ whose indefinite integral is $\frac{1}{2}x^2$, in which the use of symbols X and x can, in the mind of an inexperienced reader, set up an inappropriate and misleading analogy.

In terms of sample paths, the stochastic integral $\int_0^t X(s)dX(s)$ has representative sample form $\int_0^t x(s)dx(s)$ which is a Stieltjes-type integral with integrator function $x(I)=x(v)-x(u)$, formed from a typically “zig-zag” Brownian path $x(s)$, $0<s\leq t$, with $x(0)=0$.

[8]The Fundamental Theorem of Calculus states that if $F'(s)=f(s)$ then $f(s)$ is integrable on $[a,b]$ with definite integral equal to $F(b)-F(a)$

Then the notation for the contrasting elementary calculus integral is not $\int x\,dx$, but $\int s\,ds$, with value $\frac{1}{2}s^2$. Putting the latter in Stieltjes terms, $\int s\,ds$ is the Stieltjes integral $\int_0^t x(s)dx(s)$ where the sample path path or function x is the identity function $x(s) = s$, $0 \leq s \leq t$.

Clearly a Stieltjes integral involving a "typical" Brownian path $x(s)$ (which though continuous is, typically, nowhere differentiable) is a very different beast from a Stieltjes integral involving the straight line path $x(s) = s$. So in reality it is not surprising that there is a very big difference between the two integrals

$$\int X(s)dX(s) = \frac{1}{2}X(t)^2 - \frac{1}{2}t, \quad \text{and} \quad \int s\,ds = \frac{1}{2}s^2. \tag{7}$$

The first integral typically involves Stieltjes integrals using very complicated and difficult Brownian paths $x(s)$. It should be distinguished sharply from the more familiar and simpler Stieltjes integrals in which, for instance, the point function component of the integrand is a continuous function, and the integrator or interval function is formed from increments of a monotone increasing or bounded variation function.

It is easy to overlook this distinction. Example 60 of [13] illustrates the potential pitfall. In this Example, X_T is an arbitrary stochastic process and, with a fixed partition of $T =]0, t]$, $0 = \tau_0 < \tau_1 < \cdots < \tau_m = t$, the function $\sigma(s)$ is constant for $\tau_{j-1} < s \leq \tau_j$. Example 60 claims, in effect, that the stochastic integral $\int_{\tau_{j-1}}^{\tau_j} \sigma(s)dX_s$ exists for each j in the same way that, for constant β, $\int_{\tau_{j-1}}^{\tau_j} \beta\,dX_s$ exists and equals $\beta(X(\tau_j) - X(\tau_{j-1}))$.

But Example 8.3 above shows that this claim is false. As a step function, $\sigma(\tau_{j-1})$ is not generally equal to the constant $\beta = \sigma(s)$ when $s > \tau_{j-1}$. So if the sample path $x(s)$ is the Dirichlet function $D(s)$, the Stieltjes integral $\int_{\tau_{j-1}}^{\tau_j} \sigma(s)dx(s)$ does not exist, and the claim in Example 60 is invalid.

However, if X_T is a Brownian motion process, then each of the significant sample paths $x(s)$ satisfies a condition of uniform continuity. In that case Example 60 is valid. But it requires some proof, similar to the proof of Theorem 229 on the succeeding page.

So what is truly surprising in (7) is, not that the two integrals give very different results, but that any convergence at all can be found for the first integral.

Why is this so? This essay has avoided giving any precise meaning to expressions such as $\int_0^t X dX$—or even to a random variable X_s. But the meaning of the random variable $\int_0^t X_s dX_s$ is somehow representative of a Stieltjes-type integral which can be formulated for **every** sample path $\{x(s) : 0 < s \leq t\}$. These sample paths may consist of joined-up straight line segments (as in the archetypical jagged-line Brownian motion diagram), or smooth paths, or everywhere discontinuous paths (like the Dirichlet function). Thus any claim that all of the separate and individual Stieltjes integrals $\int_0^t x(s)dx(s)$ of the class of such sample paths x—a very large class indeed—have integral values $\frac{1}{2}x(t)^2 - \frac{1}{2}t$ must be somehow challenging and dubious.

The integrals $\int_0^t dX(s) = X(t)$, $\int_0^t dx(s) = x(t)$, show that each member of a large class of Stieltjes integrals **can** indeed yield a common, single, simple result. Our discussion of the Riemann sum calculation of these integrals illustrates how this happens: regardless of the values of $x(s)$ for $s < t$, adding up increments ensures that **all** values $x(s)$ cancel out, except the terminal value $x(t)$.

Thus, if $f(s)$ takes constant value β for $0 \leq s \leq t$, then, for every sample path $x(s)$, the Riemann-Stieltjes (and Stieltjes-complete) integral $\int_0^t f(s)dx(s)$ exists, and $\int_0^t f(s)dx(s) = \beta x(t)$ (or $\beta(x(t) - x(0))$ if $x(0) \neq 0$. This is the basis of the claim that the stochastic integral $\int_0^t f(s)dX(s)$ exists, and is the random variable $\beta X(t)$.

However, Example 8.3 demonstrates that caution must be exercised in pursuing further the logic of Riemann sum cancellation. Because if the sample path $x(s)$ is the function $D(s)$ of Example 8.3, the expression $f(s)D(I)$ is not integrable on $[0, t]$, in either the Riemann-Stieltjes sense or the Stieltjes-complete sense, even when $f(s)$ is a step function (non-constant).

It is indeed possible to take the Riemann sum cancellation idea further. Theorem 229 of [13] shows how this can be done.

But many important stochastic integrands are not actually integrable in the basic sense of the Definition 9.1. If various sample paths $x(s)$ are experimented with in the integral $\int_0^t dX_s^2$, many different results will be found. So what is the meaning of the result $\int_0^t dX_s^2 = t$?

While, for different sample paths x, $\int_0^t dx_s^2$ is not generally convergent to any definite value, there is a weak sense of convergence of the integral which makes "$\int_0^t dX_s^2 = t$" meaningful. Most importantly in this case, the weak limit t is a fixed quantity rather than a random or unpredictable quantity such as $x(t)$. But this question goes beyond the scope of the present essay, whose aim is to explore some of the basic concepts of this subject, and hopefully to illuminate them a little. A more extensive exploration is presented in [13].

References

[1] Bartle, R.G., *Return to the Riemann integral*, American Mathematical Monthly 103(8) (1980), 625–632.
`http://mathdl.maa.org/images/upload_library/22/Ford/Bartle625-632.pdf`

[2] Bullen, P.S., *Nonabsolute integration in the twentieth century*, American Mathematical Society Special Session on Nonabsolute Integration, Toronto, 23–24 September, 2000,
`www.emis.de/proceedings/Toronto2000/papers/bullen.pdf`

[3] Chung, K.L., and Williams, R.J., *Introduction to Stochastic Integration*, Birkhäuser, Boston, 1990.

[4] Itô, K., and McKean, H.P., *Diffusion Processes and their Sample Paths*, Academic Press, New York, 1965.

[5] Jarrow, R., and Protter, P., *A short history of stochastic integration and mathematical finance: the early years, 1880–1970*,
`http://people.orie.cornell.edu/ protter/WebPapers/historypaper7.pdf`

[6] Henstock, R., *The efficiency of convergence factors for functions of a continuous real variable*, Journal of the London Mathematical Society 30 (1955), 273–286.

[7] Karatzas, I., and Shreve, S. E., *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1991.

[8] Kolmogorov, A.N., *Grundbegriffe der Wahrscheinlichkeitreichnung*, Ergebnisse der Mathematik, Springer, Berlin, 1933 (*Foundations of the Theory of Probability*, Chelsea Publishing Company, New York, 1950).

[9] McKean, H.P., *Stochastic Integrals*, Academic Press, New York, 1969.

[10] McShane, E.J., *A Riemann Type Integral that Includes Lebesgue–Stieltjes, Bochner and Stochastic Integrals*, Memoirs of the American Mathematical Society No. 88, Providence, 1969.

[11] McShane, E.J., *A unified theory of integration*, American Mathematical Monthly 80 (1973), 349–359.

[12] McShane, E.J., *Stochastic Calculus and Stochastic Models*, Academic Press, New York, 1974.

[13] Muldowney, P., *A Modern Theory of Random Variation, with Applications in Stochastic Calculus, Financial Mathematics, and Feynman Integration*, Wiley, New York, 2012.

[14] Muldowney, P., *A Riemann approach to random variation*, Mathematica Bohemica 131(2) (2006), 167–188.

[15] Muldowney, P., *Henstock on random variation*, Scientiae Mathematicae Japonicae, No. 247 67(1) (2008), 51–69.
[16] Øksendal, B., *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1985.

Pat Muldowney served as lecturer at the Magee Business School of the University of Ulster for over twenty years. Dr. Muldowney's books include A General Theory of Integration in Function Spaces, Pitman, 1987 and A Modern Theory of Random Variation, with Applications in Stochastic Calculus, Financial Mathematics, and Feynman Integration, Wiley, 2012.

Lisselty Dunmore East Co. Waterford.
E-mail address: pat.muldowney@googlemail.com

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 45–57
ISSN 0791-5578

ON THE CONTINUITY OF THE INVERSES OF STRICTLY MONOTONIC FUNCTIONS

HEIKO HOFFMANN

ABSTRACT. In this short note we present an elementary, but seemingly not well known result on the continuity of the inverse of a strictly monotonic function and we discuss the relation of this result to the question when order and subspace topology are identical, both on the real line as well as in the abstract framework of connected linearly ordered spaces.

1. INTRODUCTION.

It is a fundamental question in analysis under which conditions the inverse of a continuous bijection, say between two topological spaces, is itself continuous. There are well-known results like the invariance of domain theorem or the classical (and easy to prove) result that the inverse of a continuous bijection from a compact space onto a Hausdorff space is also continuous; see also [5] for a complete characterisation of all subsets of $\mathbb{R}$ such that every continuous injection defined on a set of this kind is a homeomorphism onto its range.

It seems that results like the ones just mentioned have influenced the presentation of similar results at the level of undergraduate courses. So it seems that the following statement is most widespread in such courses.

If $\emptyset \neq I \subseteq \mathbb{R}$ *is an interval and if* $f : I \to \mathbb{R}$ *is continuous and injective, then* $f^{-1} : f(I) \to \mathbb{R}$ *is continuous, too.*

Usually, the proofs given for this result make use of the continuity of f in such a way that the continuity assumption appears to be indispensable at a first cursory glance. However, there is a more general result (see, e.g., [4, 37.1]), which, unfortunately, seems to be seldom taught in undergraduate courses.

2010 *Mathematics Subject Classification.* 26A48.
Key words and phrases. continuous inverse, strictly monotonic function.
Received on 19-3-2015; revised 6-6-2015.

If $\emptyset \neq I \subseteq \mathbb{R}$ *is an interval and if* $f : I \to \mathbb{R}$ *is strictly monotonic, then* $f^{-1} : f(I) \to \mathbb{R}$ *is continuous, too.*

This statement demonstrates that the premise of the continuity of f is entirely superfluous (of course, injectivity must be replaced by strict monotonicity) and proofs based on this premise might disguise the deeper reason for this phenomenon. In fact, from the point of view of topology, the true reason lies in the following observation (readers not very well familiar with abstract topology may skip the subsequent explanation at their first reading): a strictly monotonic function $f : I \to f(I)$ is a homeomorphism if I and $f(I)$ both carry the order topology induced by the order inherited from $\mathbb{R}$ instead of the usual subspace topology. Since the subspace topology is finer than the order topology the mapping $f^{-1} : f(I) \to I$ is continuous if $f(I)$ is endowed with the subspace topology and I carries the order topology. But since for intervals the order and subspace topology coincide, we conclude that $f^{-1} : f(I) \to I$ is continous where I and $f(I)$ now both carry the usual subspace topology.

Clearly, the same argument works for every strictly monotonic function $f : A \to \mathbb{R}$ ($\emptyset \neq A \subseteq \mathbb{R}$) whenever the order and subspace topology of A coincide. Unfortunately, the above proof (no matter how simple it is) is in general out of reach for an undergraduate course due to the topological conceptual framework. So at this point three questions arise:

(1) Is there a simple (i.e., ideally so simple that it is easily accessible to undergraduate students with no knowledge of abstract topology) description of those subsets of $\mathbb{R}$ for which the order and subspace topology of A coincide?
(2) Is there an elementary proof for the above statement about the continuity of the inverse of a strictly monotonic function defined on such a set?
(3) Does there exist a subset of $\mathbb{R}$ such that each strictly monotonic function defined on this set has a continuous inverse, but the subspace and order topology on this set are distinct?

In this note we answer the first two questions affirmatively and we present such an elementary proof, which might be easily incorporated into an undergraduate course. This proof is given in the next section, where we choose a formulation that completely avoids mention of the order topology and we get along only with notions easily accessible to undergraduate students. Furthermore, we shall show that this result is optimal in the sense that on each non-empty subset of $\mathbb{R}$ for which order and subspace topology differ there exists a strictly monotonic function whose inverse is not continuous, thus giving a negative answer to the third question.

In the last section we take up once again the abstract topologist's position in order to complete our picture and to relate Proposition 2.1 and Proposition 2.4 below to the topological point of view described above. This link is provided by Lemma 3.3, which in fact answers the first of the above questions (see Corollary 3.5).

2. Strictly monotonic functions on subsets of $\mathbb{R}$

In this section we do not want to presuppose that the reader is familiar with abstract topology in order to make sure that this part of the note is also readable, e.g., for undergraduate students. For this reason we first clarify some notions occuring in what follows.

The symbol $\mathbb{N}$ denotes the set of strictly positive integers, while $\mathbb{N}_0 := \mathbb{N}\dot{\cup}\{0\}$.
Let A be a subset of the reals $\mathbb{R}$. A set $C \subseteq A$ is called a *(connected) component of* A if C is an interval (where we include the degenerate cases of the empty set and singletons) and if each interval $I \subseteq A$ containing C already coincides with C. Each set A is the disjoint union of all its connected components. This is most easily seen by defining an equivalence relation on A by setting $a \sim a' :\Longleftrightarrow [\min\{a, a'\}, \max\{a, a'\}] \subseteq A$ for $a, a' \in A$. Then the equivalence classes of $\sim$ are precisely the connected components of A.
We call A an *open* set if for each $a \in A$ there exists $\varepsilon > 0$ such that $(a - \varepsilon, a + \varepsilon) \subseteq A$.
We denote by

$$\partial A := \{x \in \mathbb{R} : \forall \varepsilon > 0 : (x - \varepsilon, x + \varepsilon) \cap A \neq \emptyset \neq (x - \varepsilon, x + \varepsilon) \setminus A\}$$

the boundary of A. The set A is *closed* if and only if $\partial A \subseteq A$. Notice that A is closed if and only if $\mathbb{R} \setminus A$ is open.
A function $f : A \to \mathbb{R}$ is *continuous at a point* $a \in A$ if for every sequence $(a_n)_n$ in A converging to A the sequence $(f(a_n))_n$ converges to $f(a)$. The function $f : A \to \mathbb{R}$ is *continuous (on* A*)* if it is continuous at each point of A.

Now we can state and prove the announced result on the continuity of the inverse of strictly monotonic functions.

Proposition 2.1. *Let $\emptyset \neq A \subseteq \mathbb{R}$ be a set such that every bounded component of $\mathbb{R} \setminus A$ is either closed or open. Furthermore, let $f : A \to \mathbb{R}$ be a strictly monotonic function on A. Then the function $f^{-1} : f(A) \to \mathbb{R}$ is continuous.*

Proof. We suppose that f is strictly increasing (the case that f is strictly decreasing can be treated in a similar way).

Let $y_0 \in f(A)$ be arbitrary. We want to show that $f^{-1} : f(A) \to \mathbb{R}$ is continuous at y_0. For this purpose, let $(y_n)_n$ be an arbitrary convergent sequence in $f(A)$ with limit y_0. We then have to show that $(x_n)_n := (f^{-1}(y_n))_n \in A^{\mathbb{N}}$ converges to $x_0 := f^{-1}(y_0) \in A$.

It is easy to verify that there are $u, v \in f(A)$ with $u \leq v$ such that $y_n \in [u, v]$ for all $n \in \mathbb{N}_0$. We put $a := f^{-1}(u)$ and $b := f^{-1}(v)$. Then we have $x_n \in [a, b] \cap A$ for all $n \in \mathbb{N}_0$. In particular, the sequence $(x_n)_n$ is bounded and it thus suffices to verify that x_0 is its only possible limit point in order to conclude that $(x_n)_n$ converges to x_0, which completes the proof. Indeed, suppose to the contrary that x_0 is the only possible limit point of the sequence $(x_n)_n$, but this sequence does not converge to x_0. Then we may pass to a subsequence $(x_{n_k})_k$ such that $|x_{n_k} - x_0| \geq r$ for all $k \in \mathbb{N}$ and some $r > 0$. Since $(x_{n_k})_k$ is bounded as well, it has an accumulation point thanks to the Bolzano-Weierstraß theorem, say x_0', and we deduce $|x_0' - x_0| \geq r$, i.e., $x_0' \neq x_0$ on the one hand. But on the other hand x_0' is also an accumulation point of the sequence $(x_n)_n$ itself and therefore $x_0 = x_0'$ by hypothesis and so we end up with a contradiction.

Suppose now that $(x_n)_n$ possesses a limit point ξ different from x_0 and let $(x_{n_k})_k$ be a subsequence converging to ξ. We then either have $\xi > x_0$ or $\xi < x_0$. We only treat the first case (the second one is analogous) and we shall show that we obtain a contradiction.

First, assume additionally that ξ does not belong to A and denote by I that component of $\mathbb{R} \setminus A$ that contains ξ. Observe that we have $\xi \in \partial I$ because of $\xi \in \partial A$.

If ξ is the left endpoint of I and if I is not a singleton, then there exists a $k_0 \in \mathbb{N}$ with $x_{n_{k_0}} \in (x_0, \xi)$ and an index $k_1 \in \mathbb{N}$ with $x_{n_k} \in (x_{n_{k_0}}, \xi)$ for all $k \geq k_1$. This yields

$$y_{n_k} = f(x_{n_k}) \geq f(x_{n_{k_0}}) > f(x_0) = y_0$$

for all $k \geq k_1$. As $k \to \infty$ we obtain the contradiction $y_0 \geq f(x_{n_{k_0}}) > y_0$.

If ξ is the right endpoint of I (which includes the case that I is a singleton), then I must be bounded due to $x_0 < \xi$. By assumption I is either closed or open, but due to $\xi \in \partial I \cap (\mathbb{R} \setminus A)$, the set I must be closed. Therefore we then have $I = [\alpha, \xi]$ with an $\alpha \leq \xi$ such that $\alpha \notin A$.

We may now choose an element $z \in (x_0, \alpha) \cap A$. (Note that this is indeed possible: If $\alpha < \xi$, this follows from $\alpha \in \partial A$ and $x_0 < \xi$, which yields $x_0 < \alpha$. If however $\alpha = \xi$, then $(x_0, \xi) \cap A$ is nonvoid since otherwise we would obtain $(x_0, \xi] \subseteq I = \{\xi\}$, which is impossible.) There exists a $k_0 \in \mathbb{N}$ such that $x_{n_k} > z$ for all $k \geq k_0$. This implies

$$y_{n_k} = f(x_{n_k}) \geq f(z) > f(x_0) = y_0$$

for all $k \geq k_0$ and we arrive at the contradition $y_0 \geq f(z) > y_0$.

Summarizing, we infer that ξ must be an element of A. Here we distinguish between two cases: $(x_0, \xi) \cap A \neq \emptyset$ or $(x_0, \xi) \cap A = \emptyset$. In the first case we choose $z \in (x_0, \xi) \cap A$ and proceed as in the above case where ξ was a right endpoint of the above I to arrive at a contradiction.

So let us assume that $(x_0, \xi) \cap A = \emptyset$. Then there exists a $k_0 \in \mathbb{N}$ such that $x_{n_k} \geq \xi$ for every $k \geq k_0$. This yields $y_{n_k} \in [f(\xi), \infty)$ for each $k \geq k_0$, which leads to the contradiction $y_0 \geq f(\xi) > f(x_0) = y_0$.

Altogether we arrive at the conclusion that $\xi > x_0$ is not possible. $\square$

Proposition 2.1 gives rise to the following characterisation of the continuity of a strictly monotonic function.

Corollary 2.2. *Let $\emptyset \neq A \subseteq \mathbb{R}$ such that every bounded component of $\mathbb{R} \setminus A$ is closed or open. Then for a strictly monotonic function $f : A \to \mathbb{R}$ the following assertions are equivalent.*

(a) *The function $f : A \to \mathbb{R}$ is continuous.*
(b) *Each bounded component of $\mathbb{R} \setminus f(A)$ is closed or open.*

If either assertion holds, then the sets A and $f(A)$ are homeomorphic. In particular, f is continuous if its range $f(A)$ is closed, open or an interval. Moreover, the implication "(b) $\Longrightarrow$ (a)" is still true if we drop the assumption imposed on A.

Proof. Applying Proposition 2.1 to the function $f^{-1} : f(A) \to \mathbb{R}$ gives us the implication "(b) $\Longrightarrow$ (a)"; even without the assumption imposed on A.

Now assume that f is continuous as well as, without loss of generality, that f strictly increases. Furthermore, suppose to the contrary that $\mathbb{R} \setminus f(A)$ possesses a bounded component that is neither closed nor open, thus having the form $(u, v]$ or $[u, v)$. We only treat the first case.

Then $u \in f(A)$, $v \notin f(A)$ and there is a strictly decreasing sequence $(y_n)_n$ in $f(A)$ with limit v. We set $x_n := f^{-1}(y_n)$ for $n \in \mathbb{N}$ and $x := f^{-1}(u)$. The sequence $(x_n)_n$ is strictly decreasing and bounded from below by x, thus it converges to $\xi := \inf_{n \in \mathbb{N}} x_n$ in $\mathbb{R}$. The number ξ does not belong to A since otherwise the continuity of f would imply

$$v = \lim_{n \to \infty} y_n = \lim_{n \to \infty} f(x_n) = f(\xi) \in f(A),$$

which is impossible because of $v \notin f(A)$. Now consider an arbitrary $z \in A$ with $z > x$. We then have $f(A) \ni f(z) > f(x) = u$ and thus $f(z) > v$. Consequently, there exists an index $n \in \mathbb{N}$ with $v < y_n = f(x_n) < f(z)$, which implies $\xi < x_n < z$. We conclude that $(x, \xi]$ is a component of $\mathbb{R} \setminus A$ (because $x \in A$ and $A \ni x_n \to \xi \notin A$ as $n \to \infty$), which contradicts the assumption on A.

The first part of addendum is clear by Proposition 2.1. $\square$

Remark 2.3. The characterisation of the continuity of a strictly monotonic function obtained in the preceding corollary fails if the adverb "strictly" is dropped. Indeed, just consider the function $f : \{\frac{1}{n};\ n \in \mathbb{N}\} \cup \{0\} \to \mathbb{R}$ given by $f(0) := 0$ and $f(\frac{1}{n}) = 1$ $(n \in \mathbb{N})$.

As announced we now demonstrate that Proposition 2.1 is in some sense optimal.

Proposition 2.4. *Let $\emptyset \neq A \subseteq \mathbb{R}$ be a set such that $\mathbb{R} \setminus A$ possesses a bounded component that is neither closed nor open. Then there exists a strictly monotonic, continuous function $f : A \to \mathbb{R}$ such that the function $f^{-1} : f(A) \to \mathbb{R}$ is discontinuous.*

Proof. By assumption $\mathbb{R} \setminus A$ possesses a bounded component having the form $(a, b]$ or $[a, b)$ (with $a < b$). We only consider the first case since the second case is analogous.

Clearly, b is a cluster point of $(b, \infty) \cap A$. Therefore we can choose a strictly decreasing sequence $(x_n)_n$ in A converging to b. Moreover, we choose a strictly decreasing sequence $(y_n)_n$ in $\mathbb{R}$ with limit a.

Now we put $g(x_n) := y_n$ for $n \in \mathbb{N}$ and $g(a) := a$ and we extend g on (x_{n+1}, x_n) linearly. This gives us a strictly increasing, continuous function $g : \{a\} \cup (b, x_1] \to \mathbb{R}$, which we extend to a strictly increasing, continuous function $g : (-\infty, a] \cup (b, \infty) \to \mathbb{R}$ in any way. Then the function $f := g|_A$ (note that $A \subseteq (-\infty, a] \cup (b, \infty)$) is strictly increasing and continuous, but its inverse is discontinuous at a. In fact, we calculate $\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} y_n = a = f(a)$, while $\lim_{n\to\infty} f^{-1}(y_n) = \lim_{n\to\infty} x_n = b \neq a = f^{-1}(a)$. □

Remark 2.5. (a) Proposition 2.1 and Proposition 2.4 together characterise those nonvoid subsets A of $\mathbb{R}$ such that each (continuous) strictly monotonic function $f : A \to \mathbb{R}$ possesses a continuous inverse. These are precisely those non-empty sets A such that all bounded connected components of $\mathbb{R} \setminus A$ are closed or open.

(b) By Proposition 2.1, the function $g|_{(-\infty,a)\cup(b,\infty)}$ (where g is as in the proof of Proposition 2.4) has a continuous inverse. Therefore the point a is the only discontinuity of the above f^{-1}.

(c) Combined with the order topological considerations in the introduction, Proposition 2.4 furnishes a proof that the order and subspace topology of A do not coincide whenever $\mathbb{R} \setminus A$ possesses a bounded component that is neither closed nor open. The converse is also true, see Corollary 3.5 below in the next section.

3. Strictly monotonic functions on subsets of connected linearly ordered spaces

In this section we want to go beyond the scope of real functions and embed the results of the preceding section into a more general framework in order to supply the topological background underlying these results. In particular, we want to explore whether there exists a reasonable generalisation of Proposition 2.1. As we shall see, it turns out that there is indeed a perfect analogue in the setting of connected linearly ordered spaces (see Proposition 3.7 below).

We presuppose from now on that the reader is acquainted with the most basic notions of abstract topology. Nevertheless we start by reviewing some important notions.

In what follows $(X, \leq_X)$ and $(Y, \leq_Y)$ are linearly ordered sets, both endowed with their respective order topology $\tau_{(X,\leq_X)}$ and $\tau_{(Y,\leq_Y)}$. If no confusion is to be expected, we drop the indices and simply write $\leq$.

Let $\emptyset \neq A \subseteq X$. We may endow A with two reasonable topologies: the subspace topology, denoted by τ_A, generated by all sets of the form $(-\infty, x) \cap A$ or $(x, \infty) \cap A$ where $x \in X$, and the topology induced by the order on A inherited from X, denoted by $\tau_{(A,\leq)}$, generated by all sets of the form $(-\infty, a) \cap A$ or $(a, \infty) \cap A$ where $a \in A$. We always have $\tau_{(A,\leq)} \subseteq \tau_A$, but this inclusion can be strict. Notice that $\tau_{(A,\leq)} = \tau_A$ if, for instance, A is compact with respect to the subspace topology or an interval (see, e.g., [2, 4A2R (m)]).

We adopt the usual convention to write $(-\infty, x)$ resp. (x, ∞) resp. $(-\infty, x]$ resp. $[x, \infty)$ instead of $\{x' \in X : \ x' < x\}$ resp. $\{x' \in X : x' > x\}$ resp. $\{x' \in X : \ x' \leq x\}$ resp. $\{x' \in X : \ x' \geq x\}$. We write $\sup A = \infty$ if A is not bounded from above and $\inf A = -\infty$ if A is not bounded from below and by convention $-\infty < x < \infty$ for all $x \in X$ (even if $(X, \leq)$ has a minimum or maximum; in particular $\pm\infty \notin X$).

Recall that $(X, \leq)$ is called *Dedekind complete* if every non-empty subset A of X with an upper bound has a least upper bound denoted by $\sup A$. If $(X, \leq)$ is Dedekind complete and A is a nonvoid subset of X with a lower bound, then A possesses a greatest lower bound denoted by $\inf A$ (see, e.g., [1, 314B (b)]).

One says that $(X, \leq)$ is *dense* provided that for any two elements $x, x' \in X$ with $x < x'$ there exists an $x'' \in X$ such that $x < x'' < x'$.

A subset A of X is called *(order-) convex* if for any two elements $a, a' \in A$ the interval $[\min(a, a'), \max(a, a')]$ is contained in A.

The next lemma collects some basic facts concerning order topologies, which are probably folklore. For this reason we omit the easy proofs (see also Exercise 26G in [7] for assertion (a))

Lemma 3.1. *Let $(X, \leq)$ be a linearly ordered set endowed with the order topology and $\emptyset \neq A \subseteq X$.*

(a) *The space X is connected if and only if $(X, \leq)$ is Dedekind complete and dense.*
(b) *If A is connected w.r.t. τ_A, then it is convex.*
(c) *If A is an interval, then A is convex. The converse is true provided that $(X, \leq)$ is Dedekind complete.*

(d) *If* $(X, \leq)$ *is connected and* A *is an interval, then* A *is connected w.r.t.* τ_A*.*
(e) *If* $(X, \leq)$ *is connected, then the set of subsets of* X *that are connected (w.r.t. the subspace topology) coincides with the set of intervals.*

Example 3.2. Subintervals of $\mathbb{R}$, the extended real line $\mathbb{R} \cup \{\pm\infty\}$, the long line (see, e.g., counterexample 46 in [6]), the extended long line (see, e.g., counterexample 46 in [6]), the unit square with the lexicographical order (see, e.g., counterexample 48 in [6]) or lexicographic cubes (see section 2 of [3]) are examples for connected linearly ordered spaces $(X, \leq)$.

For a connected linearly ordered space $(X, \leq)$ the subsequent lemma provides a catchy characterisation of all subsets of X for which the subspace and order topology coincide. Moreover, the following Lemma 3.3 (resp. Corollary 3.5) links Proposition 2.1 and Proposition 2.4 to the topological consideration from the introduction and completes our picture. In fact, using Lemma 3.3 we obtain a perfect generalisation of Proposition 2.1 as we shall see later on.

Lemma 3.3. *Let* $(X, \leq)$ *be connected and* $\emptyset \neq A \subseteq X$*. Then the order and subspace topology of* A *coincide if and only if every component of* $X \setminus A$ *w.r.t. the subspace topology* τ_A *is closed or open.*

Proof. We first suppose that $X \setminus A$ possesses a component which is neither closed nor open. Thanks to part (e) of Lemma 3.1, there are $a, b \in X$ with $a < b$ such that either $[a, b)$ or $(a, b]$ is a component of $X \setminus A$. (Notice that all other kinds of intervals are surely closed or open.) We only treat the first case since the second one can be handled analogously.

We first observe that a cannot be the least element of $(X, \leq)$ (provided there exist any at all) because otherwise $[a, b)$ would be open. As a consequence, the set $(-\infty, a)$ is nonvoid. In addition, $(t, a) \cap A \neq \emptyset$ for each $t \in (-\infty, a)$. The latter assertion results from the fact that for a point $t \in (-\infty, a)$ with $(t, a) \cap A = \emptyset$ one would obtain $(t, b) \subseteq X \setminus A$. Since (t, b) is connected by Lemma 3.1 (d) and $[a, b)$ is a connected component of $X \setminus A$ with $[a, b) \cap (t, b) \neq \emptyset$, we infer $(t, b) \subseteq [a, b)$. But as $(X, \leq)$ is dense, the set (t, a) is nonempty. This yields $(-\infty, a) \cap [a, b) \neq \emptyset$, which is absurd. We now put $\mathcal{A} := \{t \in X : \ t < a\}$, we let $\preceq$ denote the partial order $\leq$ on $\mathcal{A}$ and we choose $x_t \in (t, a) \cap A$ for each $t \in \mathcal{A}$. Then $(\mathcal{A}, \preceq)$ is an upwards directed (nonvoid) set and $(x_t)_{t\in\mathcal{A}}$ is a net in $A\cap(-\infty,a)$ that converges in $(X,\leq)$ to a, as one easily verifies. In particular, for every $x\in A$ with $x<b$, which implies $x<a$, resp. for each $x'\in A$ with $x'>b$, there is a $t_0\in\mathcal{A}$ with $x_t\in(x,\infty)\cap A$, resp. with $x_t\in(-\infty,x')\cap A$ for all $t\succeq t_0$. Therefore $(x_t)_{t\in\mathcal{A}}$ converges to b with respect to the order topology on A.

If the order topology and the subspace topology of A coincided, then we could infer that $(x_t)_{t\in\mathcal{A}}$ converges to b with respect to the subspace topology on A, hence in $(X,\leq)$, which would yield $a=b$ (the order topology is always Hausdorff, see, e.g., [2, 4A2R (c)]) in contrast to $a<b$. As a result, the subspace topology of A is strictly finer than the order topology of A. This establishes the only-if-part.

Now we conversely assume that each component of $X\setminus A$ is either closed or open. In order to show that in this case the order and subspace topology of A are identical, it suffices to verify that each set of the form $(-\infty,\xi)\cap A$ or $(\xi,\infty)\cap A$, where $\xi\in X$, is open with respect to the order topology on A. We show this only for $(-\infty,\xi)\cap A$ because the remaining case can be treated similarly.

In the cases $\xi\in A$, $(-\infty,\xi)\cap A=\emptyset$ or $(-\infty,\xi)\cap A=A$ the assertion is clear. Therefore we may assume that $\xi\notin A$ and $(-\infty,\xi)\cap A\neq\emptyset$ and $(-\infty,\xi)\cap A\neq A$ or equivalently that $\xi\notin A$ and $(-\infty,\xi)\cap A\neq\emptyset$ and $(\xi,\infty)\cap A\neq\emptyset$. We denote by I that component of $X\setminus A$ that contains ξ. Due to $(-\infty,\xi)\cap A\neq\emptyset$, $(\xi,\infty)\cap A\neq\emptyset$ and part (b) of Lemma 3.1, the set I is bounded from above and from below. Thanks to Lemma 3.1 (e) the set I is an interval and we thus deduce that there are $a,b\in X$ such that $I\in\{(a,b],[a,b),[a,b],(a,b)\}$.

We next show that none of the cases $I=[a,b)$ or $I=(a,b]$ can occur. We establish this claim only for the first case because an analogous argument works in the second one. The same argument as utilised above in the proof of the only-if-part gives us a net $(x_t)_{t\in\mathcal{A}}$ in $A\cap(-\infty,a)$ that converges in $(X,\leq)$ to a. (For this notice that $(-\infty,a)$ is non-empty because $A\cap(-\infty,\xi)\neq\emptyset$ and $[a,\xi]\subseteq X\setminus A$.) By hypothesis, I is closed or open. If I were open, then there would exist a $t_0\in\mathcal{A}$ such that $x_t\in I$ for all $t\succeq t_0$, which is impossible because of $(-\infty,a)\cap I=\emptyset$. Hence, I is closed. Employing that $(X,\leq)$ is dense, one easily shows that b is a cluster point of I, which yields $b\in I\subseteq X\setminus A$. But as $[a,b]$ is connected by Lemma 3.1 (d) and a strict superset of I, which is a connected component of $X \setminus A$, the point b belongs to A. Contradiction!

Altogether we therefore either have $I = [a, b]$ with $a \leq \xi \leq b$ and $a, b \in X \setminus A$ or $I = (a, b)$ with $a < \xi < b$ and $a, b \in A$.

In the first case we can choose as before a net $(x_t)_{t\in\mathcal{A}}$ in $A \cap (-\infty, a)$ converging in $(X, \leq)$ to a. We then obtain

$$(-\infty, \xi) \cap A = (-\infty, a) \cap A = \bigcup_{t\in\mathcal{A}} ((-\infty, x_t) \cap A),$$

so that $(-\infty, \xi) \cap A$ is a union of sets open with respect to the order topology on A and consequently itself open with respect to the order topology on A.

In the second case we observe that $[b, \infty) \cap A$ is closed with respect to the order topology on A (because of $b \in A$). For this reason

$$(-\infty, \xi) \cap A = A \setminus ([b, \infty) \cap A)$$

is open with respect to the order topology on A. □

Example 3.4. (a) Lemma 3.3 applies to all closed subsets of a connected linearly ordered space $(X, \leq)$. Indeed, let A be a closed subset of X. Then $X \setminus A$ is open and can be expressed as a union of disjoint open intervals (see, e.g., [2, 4A2R (j)]). Hence, each connected component of $X \setminus A$ is open.

(b) Furthermore, we may apply Lemma 3.3 to open subsets U of a connected linearly ordered space $(X, \leq)$. To see this, note that each component of $X \setminus U$ is closed in $(X \setminus U, \tau_{X\setminus U})$ because components of a topological space are always closed in this space. Hence, each component of $X \setminus U$ is closed in $(X, \leq)$ since $X \setminus U$ is closed in $(X, \leq)$.

(c) Lemma 3.3 also applies to each subset D of a connected linearly ordered space $(X, \leq)$ which is dense in X. In fact, thanks to Lemma 3.1 every component of $X \setminus D$ is an interval. Since $(X, \leq)$ is dense, each non-empty interval that is not a singleton has nonvoid interior. Hence, every component of $X \setminus D$ is a singleton and consequently closed as $(X, \leq)$ is Hausdorff.

We record the following simple consequence of Lemma 3.3.

Corollary 3.5. *Let $\emptyset \neq A \subseteq \mathbb{R}$. Then the order and subspace topology of A coincide if and only if every bounded component of $\mathbb{R} \setminus A$ is either closed or open.*

Remark 3.6. If we combine Proposition 2.1, Proposition 2.4 and Corollary 3.5, we arrive at the following result:

Let $\emptyset \neq A \subseteq \mathbb{R}$. Then the order and subspace topology of A coincide if and only if every (continuous) strictly monotonic function $f : A \to \mathbb{R}$ possesses a continuous inverse $f^{-1} : f(A) \to A$, where A and $f(A)$ are endowed with their respective subspace topologies.

As previously promised we now arrive at the announced generalisation of Proposition 2.1.

Proposition 3.7. *Let $(X, \leq_X)$ and $(Y, \leq_Y)$ be two linearly ordered spaces, where $(X, \leq_X)$ is connected. Let A be a nonvoid subset of X such that each component of $X \setminus A$ (w.r.t. the subspace topology τ_A) is closed or open. Assume that $f : A \to Y$ is an injective order-preserving or injective order-reversing mapping. Then the inverse mapping $f^{-1} : (f(A), \tau_{f(A)}) \to (A, \tau_A)$ is continuous.*

Proof. If f is order-reversing, then we define another total order $\leq_Y^r$ on Y via

$$y \leq_Y^r y' :\Longleftrightarrow y \geq_Y y'$$

for $y, y' \in Y$. Then $f : (A, \leq_X) \to (Y, \leq_Y^r)$ is order-preserving and it is not hard to verify that $\tau_{(Y,\leq_Y)} = \tau_{(Y,\leq_Y^r)}$. For this reason we may and will assume w.l.o.g. that f is order-preserving. Then

$$f : (A, \tau_{(A,\leq_X)}) \to (f(A), \tau_{(f(A),\leq_Y)})$$

is a homeomorphism. Because of $\tau_{(f(A),\leq_Y)} \subsetneq \tau_{f(A)}$ and $\tau_{(A,\leq_X)} = \tau_A$ (the latter assertion results from Lemma 3.3 and the hypothesis), we conclude that

$$f^{-1} : (f(A), \tau_{f(A)}) \to (A, \tau_A)$$

is continuous as claimed. □

We close this note with the following Question:

Does also an analogue of Proposition 2.4 hold in all connected linearly ordered spaces? Or to put it another way: Is an analogon of the characterisation obtained in Remark 3.6 valid in every connected linearly ordered space?

References

[1] D. H. Fremlin, *Measure Theory – Volume 3*, Torres Fremlin, Colchester, 2004.
[2] D. H. Fremlin, *Measure Theory – Volume 4*, Torres Fremlin, Colchester, 2006.

[3] R. G. Haydon, J. E. Jayne, I. Namioka, C. A. Rogers, *Continuous functions on totally ordered spaces that are compact in their order topologies*, J. Funct. Anal. **178** (2000), 23–63.
[4] H. Heuser, *Lehrbuch der Analysis – Teil 1*, 13th ed., B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2000.
[5] M. J. Hoffman, *Continuity of inverse functions*, Math. Mag. **48** (1975), 66–73.
[6] L. A. Steen and J. A. Seebach Jr., *Counterexamples in Topology*, Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London, 1970.
[7] S. Willard, *General Topology*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1970.

Heiko Hoffmann received his Bachelor and Master degree from Saarland University in 2008 and 2010, respectively. He has recently completed his PhD thesis at the Karlsruhe Institute of Technology (KIT).

Department of Mathematics, Institute for Analysis, Karlsruhe Institute of Technology (KIT), 76128 Karlsruhe, Germany
E-mail address: heiko.hoffmann@kit.edu

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 59–62
ISSN 0791-5578

Michael J. Cloud, Byron C. Drachman and Leonid P. Lebedev: Inequalities With Applications to Engineering, Springer International Publishing, 2014. ISBN:978-3-319-05310-3, EUR 50.23, 239+xiii pp.

REVIEWED BY ANNA HEFFERNAN

This book doesn't have the best start; the very first section aims to convince the reader that inequalities are important for experimental work. However, this means that it jumps from topics like signal processing to induction coils and biomechanics at the drop of a hat. The point the authors are trying to make is clear but the jumps in topics come across as a little unsettling and leave the reader a little bewildered. Whereas one expects the opening section of a book to draw the reader in, this introduction almost has the opposite effect. Or maybe that is just the feeling of this reviewer. Maybe an engineer would have felt differently. However, having agreed to review the book, this reviewer marched on, and was really glad that they did.

The book is quite simply a very nice book. It reads at an undergraduate mathematics level but gives soft introductions and reviews on all topics leaving it possible for those without a formal mathematical or applied mathematical background to follow. The authors ease the reader through the basics of analysis with some linear algebra, selecting accordingly. They introduce, derive and apply several well known inequalties such as Bernoulli's, Cauchy-Schwarz, Minkowski's and show some neat applications in several areas ranging from bounding integrals to topology and electrostatics, finishing with a nice introduction to interval analysis. They have a gift in giving clear and concise descriptions with proofs for nearly all theorems used, backed up with worked examples and problems. In one or two areas, they drop the ball ever so slightly in their lucid descriptions but infrequently enough that you cannot hold it against them. This is an applied mathematics book and so would be enjoyed by mathematicians, both pure and applied. For those outside this area, it is still very readable but would require some patience and work but the authors will get you there.

Received on 22-5-2015.

This book is divided into seven chapters, six of which range from 16 to 32 pages while chapter five is a whopping 62 pages, leaving you feel this could have been better organised. In the first chapter, the authors bring the reader through a variety of basic inequality theorems. Each is well explained and the chapter proves to be very much self-contained, an undergraduate (or even advanced high school) student should follow the arguments and proofs quite easily. Chapter two runs through the required highlights of basic calculus that could be found in any first year undergraduate honours mathematics course. It leaves one or two theorems without proof, but none that are of a difficult level and/or not easily retrievable with a quick google, e.g., the fundamental theorem of calculus. By the end of chapter two, the reader is not only refreshed with the required calculus but the link to inequalities has been cemented in their mind.

Chapter three continues much in the same manner; with very clear descriptions of theorems, proofs and exercises. It follows on nicely from the previous chapters, bringing the reader through several well known inequalities, e.g., Young's, Cauchy-Schwarz, Chebyshev's, etc. Also, like previous chapters, there is a wealth of exercises (with hints) to assist you in becoming more comfortable with the subject area. The authors accomplish their goals with clarity and ease.

Unfortunately, things waiver slightly with the introduction of chapter four - Inequalities in abstract spaces. Understandably, the authors have to introduce the reader to a lot of mathematical concepts required for functional analysis that most likely would not cross the path of the reader unless they have studied pure or applied mathematics. Not only, are the authors introducing the basic concepts such as linear spaces, spanning bases and linear independence but slightly more complex notions such as inner product and metric spaces as well as operators. Saying that, the authors still present a very good review of the required definitions and theorems, following a 'need to know' basis which serves as a nice refresher for those previously familiar with some/most of the theories. Unfortunately with less examples and exercises than previous chapters, if viewing this material for the first time, the reader could easily get bogged down in definitions.

Chapter five is where the authors reveal many applications of the earlier mathematics. They bring you through a wide variety of topics and show some neat implications of inequalities. They start with more mathematical topics, like bounding integrals and looking at well known and used functions such as the Gamma or Bessel functions. But quickly move on to more applied areas like signal analysis and dynamical systems. Clearly at this point, the book can no longer be truly self contained; in several areas, formulae are introduced with no derivation, e.g., Euler's characteristic formula in topology or Poisson's equation in electrostatics. However, this is completely expected when touching on so many different topics and the authors draw a good line between what is required to illustrate the application and a full derivation of all equations.

Chapter six is dedicated to inequalities in differential equations but mainly builds towards an understanding of Ritz's method. Here the cleanliness and clarity seen in the description of previous problems is lost a little; the authors explain the requirement of material outside the scope of the book, e.g., Sobolev spaces, and cover that part of the material suitable, giving simpler examples. However, the explanations fall slightly short of the clear, concise descriptions of earlier chapters, and there are fewer examples and problems.

Chapter seven is the final chapter of the book and ends the book on a nice note. The authors give a brief introduction to interval analysis where they truly ease the reader into the topic. Clear and easy to follow descriptions and examples are given with references to earlier problems solved by other means. In particular, they pick up one or two of the initial problems put forth in the opening section of the book and show how the application of interval analysis eases their solving.

Any mathematician would enjoy this book and appreciate its clear, concise descriptions. Non-mathematicians who seek a better understanding behind some of the inequalities or even mathematics that arise in the different subject areas will also benefit greatly from working through the book. It simply is a very nice applied mathematics book.

Anna Heffernan is a research fellow in fundamental physics at the European Space Agency. Her PhD from University College Dublin was on the self-force problem, a method applied to the modelling of black hole binary systems. She

currently researches classical general relativity including the self-force as well as more experimental aspects such as relativistic positioning systems.

Advanced Concepts Team, European Space Agency, European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, Netherlands

E-mail address: anna.heffernan@esa.int

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 63–70
ISSN 0791-5578

David Bleecker and Bernhelm Booss-Bavnbek: Index Theory with Applications to Mathematics and Physics, International Press, October 2013. ISBN:978-1-57146-264-0, USD 95.00, 766+xxii pp.

REVIEWED BY MARIANNE LEITNER

1. Introduction

"In his [Bonn] Arbeitstagung lecture given 16 July 1962 Atiyah formulated the problem of expressing the index of elliptic operators in terms of topological invariants associated to their symbol and stated the fundamental conjecture for the Dirac operator ... A few months later, in Feburary 1963, Atiyah and Singer announced the general index formula for elliptic operators on closed manifolds and indicated the main steps of a proof ... K-theory which gave the essential framework for the statement of the index theorem had been introduced by Atiyah and Hirzebruch following Grothendieck's lead in their 1959 paper.... The *central and deep point* in this new cohomology theory was the Bott isomorphism", recalls Brieskorn (1936-2013) in [7] (see [1], [4], [3]). Both Brieskorn and Booß-Bavnbeck received their doctorates in Bonn under Friedrich Hirzebruch (1927-2012). In its draft version from 2012 [11], the book under review has been dedicated to Hirzebruch and Bleecker's PhD supervisor Chern.

The history of the book reflects this ancestry. It started as a German language textbook [6] from 1977, which was translated and somewhat extended by Bleecker in 1985. The book grew further to a 766 pages hardcover volume, more than twice that of the original textbook, or to a weight of 1.470kg (two pints, that is, and it may make you equally giddy).

2. Content of the book

The book is organised into

I: *Operators with Index and Homotopy Theory*, (Chapters 1-4, 132 pages),

Received on 21-5-2015.

II : *Analyis on Manifolds*, (Chapters 5-9, 118 pages),
III: *The Atiyah-Singer Index Formula*, (Chapters 10-13, 112 pages),
IV: *Index Theory in Physics and the Local Index Theorem* (Chapters 14-18, 342 pages).

The book provides two appendices, the first devoted to *Fourier Series and Integrals - Fundamental Principles*, the second to *Vector Bundles*.

A bounded linear operator T acting in a (complex) Hilbert space is *Fredholm*, written $T \in \mathcal{F}$, if $\ker T$ and coker$T$ are finite dimensional. The *index* of T is given by

$$\operatorname{index} T = \dim \ker T - \dim \operatorname{coker} T\ . \tag{1}$$

The index is invariant under small perturbations and in fact a homotopy invariant. Thus it generalises to continuous families of Fredholm operators over a compact parameter space X, $T : X \to \mathcal{F}$, mapping $x \mapsto T_x$, for $x \in X$. If the kernel of T_x and T_x^*, respectively, has constant dimension, it defines an isomorphism class of complex vector bundles in $\mathrm{Vec}(X)$. In order to make sense of the difference in the semi-group $\mathrm{Vec}(X)$, one introduces the Grothendieck group $K(X)$. Specifically, we have $\operatorname{index} T \in K(X)$ in eq. (1). The determinant line bundle generalises the index bundle with interesting links to recent developments in physics (zeta function regularisation, multiplicative anomaly), which are hardly discussed in this book though a reference to the work of Charles Nash is provided [10] (best wishes on the occasion of your retirement!).

Elliptic operators on sections of complex vector bundles provide a primary source of Fredholm operators (Part II, Chapters 5 and 6). Chapter 7 is a *Crash Course* on *Sobolev spaces*. In Chapter 8, elliptic *Pseudo-Differential Operators* are introduced. Let $P = \sum_{|\alpha|\leq m} A^\alpha(x)D^\alpha$ be a differential operator on $X = \mathbb{R}^n$, acting on smooth functions u with compact support. Then

$$(Pu)(x) = \int e^{i\langle k,x\rangle} p(x,k)\hat{u}(k)dk\ ,$$

where $\hat{u}$ is the Fourier transform of u and the polynomial $p(x,k) = \sum_{|\alpha|\leq m} A^\alpha(x)k^\alpha$ is the amplitude (or total symbol) of P. For more general C^∞ functions p of x and k, P defines a pseudo-differential operator, subject to a growth condition in the variable k. There is a notion of principal symbol for pseudo-differential operators. Unlike the amplitude p, the principal symbol turns out to have a geometric meaning, and for elliptic operators, its knowledge is sufficient for computing the index. Using charts, the definitions carry over to operators $P : C_0^\infty(X,E) \to C^\infty(X,F)$ between smooth sections of vector bundles E, F over a manifold X. Elliptic operators of this kind yield a space $\mathrm{Ell}(E,F)$. There is a way to construct a global amplitude [5] by considering the pull-back of E, F along the projection $\pi : T^*X \to X$. (Here T^*X is the cotangent bundle.) This gives rise to a one-to-one map $C^\infty(T^*X, \mathrm{Hom}(\pi^*E, \pi^*F)) \to \mathrm{Ell}(E,F)$, $p \mapsto P$, up to small perturbations of P that do not affect the index.

For the sake of "simplicity, accessibility and transparency", in Part III, the authors decide to "develop a larger portion of algebraic topology by means of a theorem of Raoul Bott concerning the topology of $\mathrm{GL}(N,\mathbb{C})$, i.e. on the basis of linear algebra, rather than on the basis of the theory of simplicial complexes and their homology and cohomology."(p. 252). *Winding numbers* (Chapter 10) play a central role in questions about stability of planetary orbits in celestial mechanics. In keeping visible the political sympathies of the authors, the book mentions challenging engineering tasks related to "the unmanned soft landing of the lunar module Luna 9 on February 3, 1966" (resp. Luna 1 in the previous versions). A very careful discussion of winding numbers follows, from a geometric, a combinatorial, a calculus, an algebraic and a functional analytic view point. The *index theorem* relates two of the possible generalisations to higher dimension: a local one (the topological index) and a global one (the analytic index). For example, the Euler characteristic $\chi(X)$ of a compact oriented differentiable surface X can be described as

- the number of isolated zeros of a tangent vector field on X, counted with proper multiplicity (the local index of the vector field at that point);
- the degree of the map $S^1 \to S^1$ (a winding number);
- the alternating sum of the number of vertices, edges and faces (for polyhedra).

The third approach generalises to higher dimension as an alternating sum of dimensions of cohomology groups of TX. Closely related is the description of $\chi(X)$ as

- the index of the elliptic differential operator $(d+d^*) : \Omega^{\text{even}}(X) \to \Omega^{\text{odd}}(X)$ based on the de Rahm operator d.

More generally, "one considers cycles where a given number of vector fields become dependent" (p. 262, citing Atiyah), linking topology to $GL(n, \mathbb{C})$. The mapping degree $\pi_{n-1}(\mathrm{GL}(N, \mathbb{C})) \to \mathbb{Z}$ (n even) is the first step towards Bott's Periodicity Theorem. If P is a pseudo-differential operator acting on $C_0^\infty(\mathbb{R}^n \times \mathbb{C}^N)$, for fixed $x \in \mathbb{R}^n$, its principal symbol defines a continuous map $\sigma(P)(x, .) : S^{n-1} \to \mathrm{GL}(N, \mathbb{C})$ (Chapter 11). Chapter 12 deals with Hermitian vector bundles E, F over a closed manifold X. Let $P \in \mathrm{Ell}(E, F)$ have principal symbol $\sigma(P)$. The restriction of $\sigma(P) : T^*X \to \mathrm{Hom}(\pi^*E, \pi^*F)$ to the sphere bundle $SX \subset T^*X$ defines isomorphisms, so there is a naturally associated element

$$[\sigma(P)] = [\pi^*E, \pi^*F; \sigma(P)|_{SX}] \in K(BX, SX) \cong K(T^*X) \,,$$

represented by the difference bundle obtained by gluing $\pi^*E|_{BX}$ and $\pi^*F|_{BX}$ (where $BX \subset T^*X$ is the ball bundle) on $BX \cup_{SX} BX$ along SX using $\sigma(P)|_{SX}$. It turns out that index P defined by eq. (1) depends only on the equivalence class $[\sigma(P)] \in K(TX)$. On the other hand, there is a notion of *topological index*, and the *Atiyah Singer Index Theorem* states that these two are equal,

$$\text{analytic index} = \text{topological index}$$

as group homomorphisms $K(TX) \to \mathbb{Z}$. While the analytic index is easy to define, it is hard to compute. In contrast, the topological index can be explicitely calculated in many cases, but its definition is too involved to be reproduced here.

A particular feature is the workout of the embedding proof of the Atiyah-Singer Index formula for non-trivial normal bundle of X. The crucial step is the *multiplicative property* of the index. The authors follow a suggestion made in ([9], p. 188) and apply the Bokobza-Haggiag formalism [5] to simplify this partial discussion. Eventually the cobordism proof is discussed shortly and compared to the embedding and the heat equation proof (see below).

Part IV gives a crash course in *Classical Field Theory* and in *Quantum Theory* (Chapter 14) and treats the *Geometric Preliminaries* like principal fiber bundles, connections and curvature, and characteristic classes (Chapter 15), which in part have been used already in Chapter 12. Chapter 16 on *Gauge Theoretic Instantons* investigates an application of the index theorem which is important for both mathematicians and physicists, namely the computation of the dimension of the moduli space of self-dual connections (instantons) on a principal G-bundle, where G is a compact semi-simple Lie group, on a compact oriented Riemannian 4-manifold. In the massive Chapter 17 (130 pages) it is shown that the classical geometric operators such as the signature operator, the de Rham operator, the Dolbeault operator and the Yang-Mills operator can be locally expressed in terms of twisted Dirac operators, so that *The Local Index Theorem for twisted Dirac operators* applies. In Section 4 of Chapter 17, an asymptotic expansion for the heat kernel is presented in great and useful detail over 27 pages, using the geometric concepts introduced previously in Chapter 15. The last Chapter in the book, Chapter 18, is devoted to the *Theory* of the *Seiberg-Witten* equation (1994). The authors don't try to keep up with recent developments but aim at a "digestible presentation of the main results" (p. 643). Sketches and some details of the proofs are given.

3. Comments and conclusion

The book tries to draw a complete and comprehensible picture of the field. In particular, it usually includes sketches of proofs that it cannot work out fully. The reader is encouraged to question the meaning of the formulae in a guided manner, and numerous exercises are included, mostly backed by helpful hints. The book is unusual in its willingness to go much into detail, which it does very carefully. In view of the amount of material it covers, structuring is a major issue and overall the book does a truly admirable job here. Inside the text, cohesion is established using many references to related discussions in other parts of the book. Luckily these come not only with the number of the relevant chapter and section, but also with a page number, so that a jump to other parts of the book is quick and easy and does not feel like a disruption. This also allows one to step in to the book at any place, and to get to know the book as a whole without reading it linearly from the beginning. A determined reader is suggested to follow a logical path through the material to approach the subject in one of the directions labelled as follows:

(1) Index Theorem and Topological K-Theory,
(2) Index Theorem via Heat Equation,
(3) Gauge Theoretic Physics,

(4) Spectral Geometry, and
(5) Global and Micro-Local Analysis.

The book is a rich source of citations and references for further reading, making it half an encyclopaedia, as a colleague would name it. Though these are included in the normal text, the conversational style keeps the reasoning running and the text does not appear overloaded. All in all the book is sagely written and pleasant to read.

This said, there are issues with the book. The definition of the topological index (Part III, Chapter 12) relies on the K-theoretical Thom isomorphism

$$K(X) \stackrel{\cong}{\to} K(V) \tag{2}$$

whenever $V \to X$ is a complex vector bundle over a compact manifold. For $V = \mathbb{R}^{2n} \times X$, (2) is just the Bott Periodicity isomorphism (given by the outer tensor product with a power of the Bott class $\mathbf{b} \in K(\mathbb{R}^2)$). The general case "can be considered as a reformulation or generalization of Bott Periodicity", where "the Bott class $\mathbf{b}$ ... corresponds to the canonical exterior class λ_V" (p. 289). Though efforts are made to define the class λ_V and thus the map, (2) is not actually explained or proved. Instead, the reader is referred to an "independent" proof in [6]. It seems that λ_V does not reappear in any later discussion. A survey in the literature indicates, however, that a complete proof of (2) is out of reach for the dedication and space in the book, which is not primarily devoted to K-theory.

In Part IV, Chapter 16, the introduction (p. 460f) of an invariant inner product on simple Lie algebras is cumbersome and should have been omitted. The matrix trace $\operatorname{tr} AB$ is sufficient for the purpose.

The same is true for the crash course in physics (Chapter 14) which is beyond the realm of the book. The presentation falls out of shape: Maxwell's equations for the field F are written in terms of components of the electric and magnetic fields, which are irrelevant to the book. One might have written

$$\begin{aligned} F &= dA\,, \\ d * F &= j\,, \end{aligned}$$

since the Hodge-star operator is introduced in Chapter 13. The book, however, avoids the use of the $*$ operator by introducing an extra letter δ which it takes two attempts to explain ("the codifferential (the formal adjoint of d)", p. 366). The use of the letters A and F for the one- and two-form, respectively, is standard convention but conflicts with the presentation a hundred pages later (Chapter 16, p. 463), where now F denotes an element in the gauge group (the role played by A before).

The comments about quantum electrodynamics (QED) are incomprehensible, since no quantised fermions are introduced. After a brief mention of Feynman integrals, we read "Contrary to popular misconceptions (even held by good physicists) a formal power series in α does not necessarily converge, even at $\alpha = 1/137$" (p. 383). This is a suspicious statement, even for the year 1977!

The main link between the mathematical content of the book and quantum field theory (QFT) is provided by *instantons*. These are absolute minima of the Yang-Mills (YM) functional with non-trivial winding number. The proper mathematical framework for quantum YM is lattice gauge theory. The notion of a winding number is not available on the lattice, however, and the reader would want to see at least an argument why instantons are relevant to this setting. Unfortunately, none of these issues is addressed throughout the more than 50 pages. The authors include the original construction of instantons by Atiyah, Drinfeld, Hitchin and Manin but state that a proof that this construction yields all instantons would take them too long (p. 482). They could have given a short proof by following Donaldson and Kronheimer [8] who use the simpler and more powerful approach by Nahm. The chapter culminates in the *Main Theorem* (Theorem 16.37 on page 511) which states that under rather strict conditions on the manifold, like being self-dual and having positive scalar curvature, the moduli space of self-dual connections has a manifold structure, and its dimension is specified. The authors are aware that work after 1982 (Kronheimer, Mrowka, Taubes, and Uhlenbeck) has removed the restrictions, but they only present the old argumentation.

We have already commented on Chapter 18 on *Seiberg Witten Theory*. The reader should note that the chapter is not about Seiberg-Witten QFT but about the classical equation. This is global analysis and only marginally involves index theory.

The authors cite Hilbert (p. 133): "Any true progress brings with it the discovery of more incisive tools and simpler methods which at the time facilitate the understanding of earlier theories and eliminate older more awkward developments." Unfortunately, the girth acquired by the book since 1977 does not pay any heed to this insight.

Though the book has considerable merits, the referee often felt a relief when she looked up the citations and read the short and clear expositions by Atiyah.

References

[1] Atiyah, M.F.: *Harmonic spinors and elliptic operators*, Arbeitstagung Lecture, Bonn, mimeographed, Notes taken by S. Lang, 16 July, 1962;

[2] Atiyah, M.F.: *K-Theory*, Notes by D.W. Anderson (Fall 1964), W.A. Benjamin, Inc. NY - Amsterdam 1967;

[3] Atiyah, M.F., and Hirzebruch, F.: *Riemann-Roch theorems for differentiable manifolds*, Bull. Amer. Math. Soc., Vol. **65**.4 (1959), 276–281;

[4] Atiyah, M.F., Singer, I.M.:*The index of elliptic operators on compact manifolds*, Bull. Amer. Math. Soc., Vol. **69** (1963), 422–433;

[5] Bokobza-Haggiag, J. :*Opérateurs pseudo-différentiels sur une variété diffé rentiable*, Ann. Inst. Fourier, Grenoble **19**,1 (1969), pp. 125–177;

[6] Booß, B.: *Topologie und Analysis: Einführung in die Atiyah-Singer-Indexformel*, Springer (1977);

[7] Brieskorn, E.: *Singularities in the work of Friedrich Hirzebruch* Surveys in Differential Geometry, Vol. **7** (2002), pp. 17–60;

[8] Donaldson, S.K., and Kronheimer, P.B.: *The geometry of four-manifolds*, Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, NY (1990), Oxford Science Publications;

[9] Lawson, B., and Michelsohn, M.: *Spin Geometry*, Princeton University Press (1989);

[10] Nash, C.: *Differential topology and quantum field theory*, Academic Press Ltd., London (1991);

[11] Bleecker, D., and Booß-Bavnbek, B.: *Index Theory with Applications to Mathematics and Physics*, `http://milne.ruc.dk/~Booss/A-S-Index-Book/BlckBss_2012_04_25.pdf`

Marianne Leitner is a postdoctoral Scholar at the School of Theoretical Physics, Dublin Institute for Advanced Studies. Her main research interest is the mathematical understanding of conformal field theories.

School of Theoretical Physics, Dublin Institute for Advanced Studies

E-mail address: leitner@stp.dias.ie

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 71–74
ISSN 0791-5578

Eduardo Cattani, Fouad El Zein, Phillip A. Griffiths, Lê Dũng Tráng (eds): Hodge Theory, Princeton University Press, 2014. ISBN:978-1-40085147-8, USD 90, 608 pp.

REVIEWED BY ANCA MUSTAŢA

This book provides a comprehensive survey of fundamental concepts and directions in Hodge Theory, leading into topics at the forefront of current research. The volume is based on lectures from the Summer School on Hodge Theory and Related Topics hosted by the ICTP in Trieste, Italy, and organized by E. Cattani, F. El Zein, P. Griffiths, Le D. T. and L. Goettsche. It comprises contributions of fourteen authors, written in a variety of styles and contexts from concrete, informal and local to abstract, general and highly structured. Despite these differences, all chapters benefit from a uniform strife for conciseness and efficiency, as befits the book aims. The emphasis is on providing the proper context for the development of each new idea. Some proof technicalities are omitted, as necessary in order to fit the rich material within the confines of one volume - but useful outlines of proofs and precise references to literature are provided most of the times. On the other hand, counter to the economical style of each chapter, there is some overlap of topics between different authors - including definitions of mixed Hodge stuctures, polarizations of Hodge structures, period maps, monodromy representations, local systems, the Gauss-Manin connection and, unsurprisingly, the Hodge conjecture - but each time within a somewhat different context or goal. As an overall result, the reader can focus on the connections between different concepts and form a general picture while also gaining enough familiarity with each topic.

While the homology and cohomology groups provide topological invariants for manifolds, Hodge theory encodes the structure of a complex projective manifold X into linear algebraic data on the cohomology of X. Chapter 1, by Eduardo Cattani, lays down the analytic background: starting with a brisk tour of complex, symplectic, Hermitian and Kahler structures on a manifold, followed

Received on 28-5-2015.

by constructions of the de Rham and Dolbeaut complexes of differential forms leading to the definition of de Rham and Dolbeaut cohomologies. The Hodge decomposition of the cohomology groups is introduced via harmonic forms. No proof is given, but the preliminary work is concrete and detailed enough, without over-reliance on long computations: these are relegated to exercises. The author pays due attention to the Kahler metric on a projective manifold, and the relation between the Kahler form and the hyperplane class. This induces an extra structure on the cohomology groups called polarization, or equivalently a Hermitian metric whose orthogonal decomposition is compatible with the Hodge decomposition, and leads to a further splitting of the Hodge groups (the Lefschetz decomposition).

While Chapter 1 starts with the definition of a holomorphic map, this inclusive beginning is rather deceptive. The following chapters will assume a background in algebraic geometry and homological algebra. Chapter 2 contains a new proof of *Grothendieck's Algebraic de Rham Theorem*, whereby the cohomology of an algebraic manifold (not necessarily compact) can be calculated based on sheaves of algebraic differential forms. Spectral sequences provide a unitary and elegant framework for the discussion in Chapters 2 and 3, by Fouad el Zein with co-authors Loring W. Tu and Lê Dũng Tráng: starting from sheaf cohomology, and continuing with mixed Hodge complexes and structures. Surprisingly though, spectral sequences are only defined on page 134, some 40 pages after their first use in Chapter 2, and are only truly fleshed out 20 pages later. The important concept of mixed Hodge structures (MHS) first occurs in a highly formal, abstract presentation, but the patient reader is fully rewarded by a nice geometric motivation at the end of Chapter 3: Given a non-compact quasi-projective variety X, we can embedd it in a projective manifold Y such that the complement D admits a nice structure, called a normal crossing divisor (NCD). Then the holomorphic differential forms on X can be related to differential forms on Y with poles along D, via so called residue maps. Thus the simplicial structure given by the components of D leads to a new filtration on the cohomology of X, *the weight filtration.* Together with the Hodge filtration, this forms a MHS, which will occur often in the rest of the book.

Chapter 4 by James Carlson is refreshingly concrete and serves as an analytical preamble for the presentation of Variations of Hodge Structures (VHS) by Eduardo Cattani in Chapter 7. A fixed topological space X can have many different complex structures. This leads to the construction of a classifying space of Hodge structures (with fixed polarization and Hodge numbers). Any family of complex structures $(X_b)_{b\in B}$ yields a map from the basis B to this moduli space, *the period map*, whose differential yields *the Gauss-Manin connection*. In the case when B is not compact, the remarcable properties of the Gauss-Manin connections lead to an extension of Hodge structures at the limit. Cattani presents the asymptotic behaviour of the period map, helpfully illustrating it by the concrete example of the mirror quintic.

Chapters 5, 6, and 8 by Luca Migliorini, Mark Andrea de Cataldo, Patrick Brosnan and Fouad El Zein respectively, all deal with variations of mixed Hodge structures (VMHS): for algebraic families of (possibly singular) varieties, they discuss the interplay of the corresponding MHS-s. Thinking of the members of such a family as fibres of an algebraic morphism, not necessarily smooth nor proper, they organize the study of the MHS-s in terms of a suitable stratification of the target. This leads to a decomposition theorem, which shows how the intersection cohomology groups of the domain split into a direct sum of intersection cohomology groups on the target. Again, this requires an intricate formalism, involving e.g. the category of perverse sheaves, and most arguments are only sketched in these chapters. The focus is on illustrating the theorem through a series of well chosen examples - de Cataldo's chapter is just a long sequence of exercises. One cannot help but feel that this chapter was written in some haste and the onus is left on the reader to slowly flesh it out - while references provided are sometimes imprecise. Patrick Brosnan and Fouad El Zein follow the evolution of the geometric ideas, focusing on the case when the fibres of the family are not compact, and building on from their discussion of the NCD case in Chapter 3. Linking in with the chapters on classifying spaces, they finish with recent results on admissible normal function (describing an admissible variation of graded-polarized mixed Hodge structures), and their algebraic zero-locus.

Chapter 9 by Jacob Murre is a beautiful survey on the various equivalence relations on algebraic cycles, the relations between them and between their respective quotient groups. The cycle map, Abel-Jacobi map and Albanese map provide links with Hodge structures. The exact nature of these connections has been long investigated, making the subject of the celebrated Hodge and Bloch-Beilison conjectures. Mark Green's Chapter 10 transposes the main themes of Hodge theory to the case of varieties generated over number fields, resulting in many conjectures. Recent results by Mark Green and Phillip Griffiths are illustrated by examples in the last sections.

Chapters 10 and 11 are due to a new generation of mathematicians who have already made important contributions to Hodge theory: Francois Charles, Christian Schnell and Matt Kerr. In these chapters they discuss arithmetic aspects of Hodge theory in well-thought out and largely self-contained presentations.

While the present volume cannot replace classics like [1], [2], [3], [4], [5], it can serve as a good reference, or road-map, for readers interested in Hodge theory. It outlines the development stages of main themes, and their interactions, and it can point the reader towards new exciting directions.

References

[1] Phillip Griffiths and Joseph Harris. *Principles of algebraic geometry.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1994. Reprint of the 1978 original.

[2] J. Carlson, S. Muller-Stach, C. Peters, *Period mappings and Period Domains*, Cambridge Studies in Advanced Mathematics 85, 2003.

[3] Claire Voisin. *Hodge theory and complex algebraic geometry.* I, volume 76 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, English edition, 2007. Translated from the French by Leila Schneps.

[4] Claire Voisin. *Hodge theory and complex algebraic geometry.* II, volume 77 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, english edition, 2007. Translated from the French by Leila Schneps.

[5] Jose Bertin, Jean-Pierre Demailly, Luc Illusie, and Chris Peters. *Introduction to Hodge theory*, volume 8 of SMF/*AMS Texts and Monographs.* American Mathematical Society, Providence, RI, 2002. Translated from the 1996 French original by James Lewis and Peters.

Anca Mustaţa is lecturer in mathematics at University College Cork since 2007. Her main research interests lie in the areas of algebraic and complex geometry.

School of Mathematical Sciences, University College Cork
E-mail address: a.mustata@ucc.ie

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 75–76
ISSN 0791-5578

Julian Havil: John Napier. Life, Logarithms, and Legacy., Princeton University Press, 2014. ISBN:978-0-691-15570-8, US 35.00, 123+xiv pp.

REVIEWED BY ERNESTO NUNGESSER

This book describes the life and work of John Napier with great care. It can be seen as a translation of the work of Napier into modern language and concepts. One of the main motivations of the book is to underline the importance of the work of this mathematician, in view of the fact that, for instance the 2010 Britannica publication, *The 100 Most Influential Scientists of All Time*, makes no mention of Napier or in an online poll conducted by the National Library of Scotland concerning the favourite Scottish scientist, he only made it to be the tenth.

The first chapter starts with the citation "May you live in interesting times" and is appropiate to describe the time in which John Napier lived. As is mentioned from the beginning there is not much known about his life, almost nothing about his youth, not even in which university he studied. Nevertheless some known facts and anecdotes give a good impression about his life and an epoch which was characterized by different political struggles which took the form of religous conflicts and the struggle between the so-called Kings' Men and the Queen's Men. In the second chapter we learn about his passion for and decipherment of the apocalypse. This topic is interesting and Napier thought he would be most remembered because of his contribution in the understanding of the revelation. However it seems that apart from being especially methodical, his conclusions, like for instance that the pope was the antichrist were rather standard during his time. In chapter three the tables are presented and the reader is introduced to Napier logarithms and how they relate to the definition of logarithms we know today. It is not until chapter four that we know how he actually constructed the tables and his genius becomes evident. The introduction of motion in his conception is the key that brings him to the relations we know and

Received on 13-5-2015.

love. In the remaining three chapters his work is put into context with modern computation and his legacy. That a new technology had been born is evident. As is described in the book in the beginning Napier was dealing with sines, until he later realized that his construction was not that artificial and really helped to treat relations in general. Logarithms are central in the conception of bases of numbers and is closely related to the decimal system. This is explained in detail in the last three chapters together with Napier's bones. Alltogether it seems difficult to put the work of this man into context with other mathematicians. What seems marvellous is the combination of mystical and practical thinking in his work. The year 2012 was the 400th anniversary of Napier's publication *Descriptio* and in view of the fact that the younger generations will not even know of the existence of the logarithmic tables, there is a good reason to write about the work of this man. It is clearly a very interesting book from a historical perspective.

Ernesto Nungesser is a postdoctoral researcher at Trinity College Dublin, currently funded by the Irish Research Council. His main research interests are related to relativistic kinetic theory.

School of Mathematics, Trinity College Dublin
E-mail address: ernesto@maths.tcd.ie

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 77–80
ISSN 0791-5578

Leiba Rodman: Topics in Quaternion Linear Algebra, Princeton University Press, 2014. ISBN:978-0-691-16185-3, USD 79.50, 384 pp.

REVIEWED BY RACHEL QUINLAN

This book consists of a comprehensive account of matrix theory over real quaternion division algebra $\mathbb{H}$. The first seven chapters present an adaptation of many of the essential principles of matrix theory (over the real and complex fields) to the quaternion context. Chapters 8 through 14 are mostly devoted to the theme of quaternion matrix pencils and are closely based on a series of recent research articles by the author. The exposition is detailed and careful, and readers familiar with concepts such as canonical forms and standard matrix factorizations and with basic knowledge of analysis and topology will find it accessible and largely self-contained.

Irish readers may be disappointed by the fact that an entire book on the algebra of quaternions mentions William Rowan Hamilton only once, in passing, as a warrant for the use of the symbol $\mathbb{H}$. Otherwise the book is extraordinarily comprehensive. The first chapter introduces some notational and other conventions, and the second discusses the basic arithmetic of quaternions, including such topics as automorphisms and involutions of the quaternion algebra. Particular matrix realizations of $\mathbb{H}$ via the real and complex regular representations are presented; these concrete interpretations of $\mathbb{H}$ are often used throughout the text for computational purposes. Readers who enjoy random mathematical challenges may like Exercise 2.7.21, which asks for a demonstration that every non-zero element of $\mathbb{H}$ can be written in infinitely many ways as the product of two pure quaternions.

Chapter 3 establishes some basic principles of linear algebra for vector spaces over the quaternions and for matrices with quaternion entries. Standard matrix factorizations over $\mathbb{R}$ and $\mathbb{C}$ that extend unproblematically to the quaternion setting are presented, such as

Received on 4-6-2015.

the Gaussian elimination algorithm, the singular value decomposition, and the QR factorization for a square quaternion matrix. A detailed analysis of topological and geometric properties of numerical ranges of quaternion matrices is provided, with respect to general involutions as well as the standard quaternion conjugation. Your (routine but informative) challenge from the exercises in this chapter is to show that every element of $\mathbb{H}^{n\times n}$ that has a right inverse also has a left inverse and that these coincide, and to calculate the inverse in $\mathbb{H}^{3\times 3}$ of the matrix

$$\begin{pmatrix} 0 & i & j \\ -i & 0 & k \\ -j & -k & 0 \end{pmatrix}.$$

The next two chapters consider canonical forms, for congruence classes of hermitian and skew-hermitian quaternion matrices (Chapter 4) and for similarity classes of square quaternion matrices (Chapter 5). A quaternion matrix A is said to be hermitian if it equal to its conjugate transpose A^*, and skew-hermitian if $A = -A^*$. If $A \in \mathbb{H}^{n\times n}$ is hermitian then $x^*Ax \in \mathbb{R}$ for all $x \in \mathbb{H}^{n\times 1}$, and so concepts such as positive definiteness for hermitian matrices extend unproblematically to the quaternion setting. Canonical forms are established in Chapter 4 for quaternion matrices that are hermitian or skew-hermitian, in the above sense or with respect to an involution ϕ other than the standard conjugation. The development and the results appear to be analagous to the theory over $\mathbb{C}$. Chapter 5 introduces the left and right spectra of a square quaternion matrix and the quaternion analogue of the Jordan canonical form. If $A \in \mathbb{H}^{n\times n}$, then $\lambda \in \mathbb{H}$ is a right eigenvalue of A if $Av = v\lambda$ for some nonzero $v \in \mathbb{H}^{n\times 1}$, and $\mu \in \mathbb{H}$ is a left eigenvalue of A if $Au = \mu u$ for some nonzero $u \in \mathbb{H}^{n\times 1}$. It is easily confirmed that if λ is a right (or left) eigenvalue of A then so also is every element $\alpha^{-1}\lambda\alpha$ of the conjugacy class of λ in the multiplicative group $\mathbb{H}^{\times}$ of the quaternion division algebra. So eigenvalues of quaternion matrices are not really single elements but conjugacy classes. There is no general connection between the left and right spectra of an element of $\mathbb{H}^{n\times n}$ and they may be disjoint. There is a Jordan canonical form theorem for quaternion matrices; it resembles the usual statement for algebraically closed fields except that the eigenvalues that appear in the Jordan blocks are determined only up to conjugacy in $\mathbb{H}^{\times}$. The exercises in Chapter 5 reveal some unsettling failures of reliable certainties of linear algebra to translate to the division ring setting. The harmless-looking example $\begin{pmatrix} 1 & i \\ j & k \end{pmatrix}$ shows that a quaternion matrix need not be similar to its transpose, and even that the transpose of an invertible quaternion matrix need not be invertible. One of the issues here seems to be that it is possible for a set of column vectors to be right linearly independent but not left linearly independent over $\mathbb{H}$.

Chapter 6 considers subspaces of $\mathbb{H}^{n\times 1}$ that are simultaneously invariant for one matrix in $\mathbb{H}^{n\times n}$ and have special properties (such as being totally isotropic) with respect to a form defined by another. Chapter 7 discusses the Smith normal form of a matrix written over the ring of polynomials in one (central) variable over $\mathbb{H}$, and the Kronecker canonical form for a quaternion matrix pencil. This theme serves as an introduction to the second half of the book, which is concerned with canonical forms of pencils of quaternion matrices of special forms (for example hermitian or skew-hermitian, or ϕ-(skew)-hermitian for a nonstandard involution ϕ). This detailed and extensive analysis is mostly drawn or adapted from a series of recent research articles by the author (and collaborators in some cases). Many open problems are included.

The book covers a huge range of material, from basic information about the algebra of quaternions to discussions of much more specialized interest in the later chapters. It is very well organized and the quality of exposition is high; care has been taken to provide an accessible and readable account with plenty of interesting problems. Each chapter concludes with some notes explaining aspects of the context and provenance of the content, and advising the reader of relevant literature. The early chapters would certainly be useful to lecturers and students of graduate courses on such themes as linear algebra or ring theory, and the entire volume will certainly be useful as a reference text for researchers in linear algebra. The author passed away in March this year, however his comprehensive and engaging book will be appreciated long into the future.

Rachel Quinlan is a Senior Lecturer in Mathematics at NUI Galway. She completed her PhD in 2000 at the University of Alberta and has research interests in linear algebra and representation theory.

School of Mathematics, Statistics and Applied Mathematics, NUI Galway
E-mail address: rachel.quinlan@nuigalway.ie

Irish Math. Soc. Bulletin
Number 75, Summer 2015, 81–84
ISSN 0791-5578

PROBLEMS

IAN SHORT

Problems

Let us begin with a classic.

Problem 75.1. What is the least positive integer n for which a square can be tessellated by n acute-angled triangles?

The second problem was proposed by Finbarr Holland of University College Cork. The inequality involving the exponential function that is considered in the problem is a generalisation of the useful inequalities

$$e^x \leqslant \frac{1}{1-x} \quad \text{and} \quad e^{2x} \leqslant \frac{1+x}{1-x} \quad (0 \leqslant x < 1),$$

which are strict inequalities unless $x = 0$.

Problem 75.2. Let

$$s_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}, \quad n = 0, 1, 2, \dots .$$

Suppose $0 < \alpha < 1$. Prove that when $n \geqslant 1$,

$$e^x \leqslant \frac{s_n(x) - \alpha x s_{n-1}(x)}{1 - \alpha x} \quad \text{for all } x \in [0, 1/\alpha)$$

if and only if $\alpha \geqslant 1/(n+1)$.

We finish with another inequality: the sort that might crop up in a mathematics olympiad.

Problem 75.3. Given positive real numbers a, b, and c, prove that

$$a + b + c \leqslant \sqrt[3]{abc}\left(\frac{a}{b} + \frac{b}{c} + \frac{c}{a}\right).$$

Received on 11-6-2015.

Solutions

Here are solutions to the problems from *Bulletin* Number 73.

The first problem was solved by Angel Plaza (Universidad de Las Palmas de Gran Canaria, Spain), the North Kildare Mathematics Problem Club, and the proposer, Finbarr Holland. We present the solution of the North Kildare Mathematics Problem Club.

Problem 73.1 **.** Let U_n denote the Chebyshev polynomial of the second kind of degree n, which is the unique polynomial that satisfies the equation $U_n(\cos\theta) = \sin((n+1)\theta)/\sin\theta$. The polynomial U_{2n} satisfies $U_{2n}(t) = p_n(4t^2)$, where

$$p_n(z) = \sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} z^{n-k}.$$

Prove that p_n is irreducible over the integers when $2n+1$ is a prime number.

Solution 73.1. Define $q_n(t) = p_n(2t+2)$, so that $q_n(2t^2-1) = U_{2n}(t)$. Since $q_n(\cos 2\theta) = U_{2n}(\cos\theta)$, the n roots of q_n are the numbers $\cos(2k\pi/(2n+1))$ for $k = 1, \dots, n$. We prove that if p_n is reducible, then $2n+1$ is not prime.

Suppose that p_n is reducible over the integers. Then so is q_n, and one of the proper factors of q_n has $a = \cos(2\pi/(2n+1))$ as a root. It follows that the degree of the extension $\mathbb{Q}(a)$ over $\mathbb{Q}$ is less than n. Now let $b = i\sin(2\pi/(2n+1))$. Since $b^2 = a^2 - 1$, the degree of the extension $\mathbb{Q}(a, b)$ over $\mathbb{Q}$ is less than $2n$. Notice that $\mathbb{Q}(a, b)$ contains $a+b$, a primitive root of unity. Therefore the cyclotomic polynomial $x^{2n} + \cdots + x^2 + x + 1$ of degree $2n$ splits in $\mathbb{Q}(a, b)$. However, this polynomial is irreducible when $2n+1$ is prime, as is well-known, so $2n+1$ cannot be prime. $\square$

The second problem was solved by Henry Ricardo (New York Math Circle, New York, USA), the North Kildare Mathematics Problem Club, and the proposer (the Editor, who learned the problem from Tony Barnard of King's College London). The solution we present is an amalgamation of the submitted solutions. Henry Ricardo pointed out that the problem (and solution) appear elsewhere; for example, see Problem 1339 in Math. Mag. 64 (1991), no. 1.

Problem 73.2. Find all positive integers a, b, and c such that

$$\begin{aligned} bc &\equiv 1 \pmod{a} \\ ca &\equiv 1 \pmod{b} \\ ab &\equiv 1 \pmod{c}. \end{aligned}$$

Solution 73.2. Without loss of generality, suppose that $a \leqslant b \leqslant c$. Since $bc-1$, $ca-1$, and $ab-1$ are divisible by a, b, and c, respectively, we see that

$$(bc-1)(ca-1)(ab-1) = (abc)^2 - (abc)(a+b+c) + (ab+bc+ca) - 1$$

is divisible by abc. Hence $ab + bc + ca - 1$ is divisible by abc. But $0 < ab + bc + ca - 1 < 3bc$, so $a < 3$.

Next, we know that

$$(ca-1)(ab-1) = a^2(bc) - (ab+ca) + 1$$

is divisible by bc, so $(ab + ca) - 1$ is divisible by bc. But $0 < (ab + ca) - 1 < 2ac$, so $b < 2a$.

From the inequalities $a < 3$ and $b < 2a$ we see that either $a = 1$ and $b = 1$ or $a = 2$ and $b < 4$. In the former case we obtain the solution $(1, 1, m)$, where m is any positive integer. In the latter case, the congruence $bc \equiv 1 \pmod{a}$ tells us that b is odd, so $b = 3$. From the congruence $ab \equiv 1 \pmod{c}$ we deduce that $c = 5$, which gives the only other solution $(2, 3, 5)$. □

The third problem was solved by Adnan Al (Mumbai, India), Angel Plaza (Universidad de Las Palmas de Gran Canaria, Spain), Henry Ricardo (New York Math Circle, New York, USA), the North Kildare Mathematics Problem Club, and the proposer (the Editor, who learned the problem from Tony Barnard). It was also solved by Finbarr Holland, and it is his short solution that we present here. Several contributors noted that there is literature on this kind of problem; see, for example, S. Koumandos, *Remarks on a paper by Chao-Ping Chen and Feng Qi*, Proc. Amer. Math. Soc. 134 (2006), no. 5, 1365–1367.

Problem 73.3. Prove that

$$\frac{1}{10\sqrt{2}} < \frac{1}{2} \times \frac{3}{4} \times \frac{5}{6} \times \cdots \times \frac{99}{100} < \frac{1}{10}.$$

Solution 73.3. Let

$$v_n = \frac{1}{2} \times \frac{3}{4} \times \frac{5}{6} \times \cdots \times \frac{2n-1}{2n}.$$

Then a quick check shows that the sequence $\sqrt{n}v_n$ is strictly increasing and the sequence $\sqrt{2n+1}v_n$ is strictly decreasing. Since $v_1 = 1/2$, we obtain the more general collection of inequalities

$$\frac{1}{2\sqrt{n}} < v_n < \frac{1}{\sqrt{2n+1}}, \quad n = 2, 3, \dots . \qquad \square$$

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.

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

Instructions to Authors

The *Bulletin* is typeset using LaTeX; therefore, authors should submit articles to the *Bulletin* as LaTeX input files together with any macro files or input files for the production of graphics. Typescripts are **not** acceptable. Source files should be suitable for processing using `pdflatex`. All files should be accompanied by any non-standard style or input files which have been used. Private `METAFONT` and TeX macros, reference input files and files for diagrams must also accompany submissions. The preferred method of submission is by e-mail to

`mailto://ims.bulletin@gmail.com`

Only if this is not possible, authors may send a CD with their files to the Editor together with a printout of the typescript. If articles are submitted in electronic form, no hard copies need to be sent unless requested by the Editor. 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. To this end, authors will be supplied with style files. Since normally no proofs are sent out before publication, it is the author's responsibility to check carefully for any misprints or typos. 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.

ISSN 0791-5579
9 770791 557007