McGill Undergraduate Mathematics Magazine

the δelta ε psilon

Spring 2009

Third Issue

McGill

Contents

Letter From The Editors

Dear students of Mathematics, Science and Arts,

In the month of April of 2006, as the final exams were nearing at a frantic pace, a couple of math undergraduates decided to create a new journal to rival those of the Faculty of Arts, and, more importantly, to serve as a medium for the flow of mathematical ideas within the student community of Burnside Hall. Thus, the Delta-Epsilon was born.

Three years later, with the winding clock of time pointing back to April, at a time of rebirth and renewal symbolized by spring, the descendants of that founding Editing Team were slightly behind schedule but on their way to release the third issue of the Delta-Epsilon, more formally known as the McGill Undergraduate Mathematics Magazine. The journal itself, which you now hold in your hands, was reborn and acquired a new philosophy and a new form.

This magazine differs significantly from its predecessors in that it returns to the very basics: it focuses uniquely on the research articles and on interviews with the Faculty. No more jokes, reviews, anecdotes. It is simplistic and stripped down to its core; the contents of each article become the sole purpose. Another important objective to us was to acquaint the new students and veterans alike with the science professors at McGill, and particularly the Department of Maths & Stats. In this year's issue, we will meet professors Claude Crépeau and Dmitry Jakobson.

The Delta-Epsilon is intended as a place to publish summer research by undergraduates and the journal contains eight papers from varied areas of mathematics: probability and statistics, mathematical modeling, analysis and partial differential equations, mathematical physics and also number theory. There is a little hint of a computer science flavor as well. Some of the papers are more accessible than others, and some require a number of more advanced courses to understand. This magazine is designed to have a long shelf-life, that is, we hope that you will keep it on your bookshelf and return to it later again when your own work draws you back to ideas exposed herein perhaps.

Finally, we wish to strongly encourage all of you undergraduates to engage yourselves in summer research and independent studies and submit your work for next year's issue of the Delta-Epsilon. The Editing Team next year will have many choices to make in defining the direction in which our magazine will evolve, but student research will always remain the nucleus of this journal.

The Delta-Epsilon needs you: become an editor and help maintain this tradition alive.

Enjoy the articles and let us know what you think.

The Delta-Epsilon Editing Team

Interview with Professor Claude Crépeau

Nan Yang

Professor Claude Crépeau is a computer scientist who specializes in cryptography and quantum information. I've had the chance to sit down and ask him about his work.

The Delta Epsilon ($\delta\varepsilon$): *Can you tell us about your fields of research?*

Prof. Crépeau: My main fields of research are cryptography, which is the science of secrecy, and quantum computing — the development of a computing machine based on principles of quantum mechanics. There is a strong link between these two fields because quantum computing was born out of quantum cryptography, which was the first time quantum physics was involved in cryptography.

$\delta\varepsilon$: *What exactly is a quantum computer?*

Prof. Crépeau: A quantum computer is a machine in which the basic units of information are quantum states. In a standard or *classic* computer, we make the effort to have these definite states — zeros and ones — which are electrically very different and are thus very easy to distinguish from one another. All computations are done on these zeros and ones. In a quantum computer, however, the states that we manipulate are essentially continuous values; and while it is the case that we have equivalents of zeros and ones that we can distinguish very reliably, all in-between states are also possible, like an analog computer. But unlike an analog computer, the rules of evolution in a quantum computer are — fittingly — given by those of quantum mechanics, and this is a much richer set of rules than those of classical computers. And it appears that certain computations are faster under these rules. A canonical example of this is the algorithm found in 1994 by Peter Shor which could factor large numbers efficiently on a quantum computer, whereas we don't have such an algorithm on classical computers, at least not yet.

$\delta\varepsilon$: *It is thought that analog computers can never be practically realized because they are very sensitive to noise. How do quantum computers overcome this?*

Prof. Crépeau: I would say that there were two major steps in the history of quantum computing: one was to convince everyone that it was significant — something Shor's algorithm has done — and the other was the discovery of quantum error correction codes. The notion of quantum error correction is extremely intriguing because at first glance we may think that Heisenberg's uncertainty principle forbids us from correcting errors in a quantum system because if we tried to look at it we would disturb the system irreversibly, thus causing more errors. But it turns out that Heisenberg's uncertainty principle does not apply here. Quantum error correcting codes are based on the fact that you can observe *errors* without observing the *data*. So by making the right measurements, you can look for errors and not the information; and although you may end up disturbing the errors you are not disturbing the information. There is now a whole theory of how one can build a quantum computer from components that are imperfect, with quantitative theorems about the level of imperfection we can tolerate.

Prof. Claude Crépeau

$\delta\varepsilon$: *In terms of computability, are there uncomputable classical functions that are computable on a quantum computer?*

Prof. Crépeau: In terms of computability theory they are equivalent. Everything that can be done on a classical computer can be done on a quantum computer, and the other way around, which is more surprising. Essentially if you want to simulate a quantum computer on a classical

computer, you just write down the approximative amplitudes and you compute everything according to quantum physics — it's just incredibly slower.

In terms of efficiency, there *appear* to be certain tasks that are feasible on a quantum computer and not on a classical computer. This is a very big open question. Knowing that we can factor large numbers efficiently is probably the most impressive gap between classical and quantum computation. However, the theory of computation is more than just the notion of speed of computations. There are ways of manipulating quantum information that cannot be replicated classically.

$\delta\varepsilon$: *What is quantum teleportation and what was your role in its discovery?*

Prof. Crépeau: (Laughs) Well, first let me make something clear. Quantum teleportation is only vaguely related to what we usually think of as teleportation. The principle is that if the sender and receiver share a special quantum state that we call an EPR pair, it's possible to make a manipulation on the sender's end using half of the EPR pair and a state S that he's trying to send to the receiver Ű– make a manipulation, make a measurement –Ű and communicate over a classical channel the result of that measurement which gives the other party the description of an operation he can apply to his half of the EPR pair that results in S. It's essentially a mechanism that allows you to send a quantum state without having a quantum channel.

Now this is important if the sender does not know where the receiver is; quantum states cannot be cloned, so sometimes it's not possible to broadcast a state (for example, if the state is given to the sender by a third party, who does not wish to disclose it). With quantum teleportation, however, as long as the sender and receiver have arranged to share an EPR pair, the sender can broadcast the classical result and the receiver can pick it up from wherever he is, complete the teleportation process, and end up with the desired state.

$\delta\varepsilon$: *Can you elaborate on this* ***no-cloning theorem****?*

Prof. Crépeau: The no-cloning theorem was discovered in the 80's. It says that if you agree to the rules of quantum mechanics, then the operation of starting from one arbitrary quantum state and producing even one copy of that state is not a valid one. So any process that will try to copy a quantum state will fail with some probability and will produce something which is not quite right. It is not a limitation due to imperfect equipment — no device which corresponds to the laws of quantum mechanics can perform this task. An important consequence of this theorem is that in general the information embedded in a quantum state can only be at one place. Thus one way of demonstration that a certain system does not carry some information is by showing that the information can actually be found elsewhere. This trick is used often in quantum computing and quantum cryptography proofs, and is very elegant.

$\delta\varepsilon$: *How did you first get into this field?*

Prof. Crépeau: Well, first I was interested in number theory, then I read the Scientific American paper that introduced the RSA cryptosystem. Soon after, I realized that one of my professors at Université de Montreal — Gilles Brassard — was working in this area. At the very time that I met him is the time when quantum cryptography was invented by him and Charlie Bennett from IBM, and so my interest for cryptography was, for many years, in classical cryptography, mostly involving number theory, and as time went by I got more and more interested in the quantum aspect Ű- well in particular because I was right there as it was happening. Surprisingly, a lot of people didn't take it very seriously at first, but I was convinced that this was extremely valuable, so I ended up writing my PhD thesis on quantum cryptography, which showed the security of certain cryptographic protocols based on quantum exchanges.

$\delta\varepsilon$: *Is quantum cryptography an active field of research?*

Prof. Crépeau: In a sense it's exploding. Canada has one of the largest set of people working in this field. In particular because a lot of it started here in Montreal, and also the Perimeter Institute in Waterloo has lots of people who are working in quantum computing. It's a very big center — probably the largest center in the world.

$\delta\varepsilon$: *What are the practical difficulties in building a quantum computer?*

Prof. Crépeau: The difficulty lies in the fact that when a quantum system come into contact with its environment it tends to lose its "quantumness" when observed in a large scale such as the one in which we live. In this macroscopic scale we don't see many quantum effects. We have to look at things on a smaller scale and for shorter time periods to actually see most of the quantum effects that are going on. Now if you want to complete a quantum computation that lasts for several seconds — maybe even minutes — and all the quantum states must remain "quantum" all the way through, then you'll have to isolate the system very reliably from its environment. This is mostly where the difficulty comes in because the machinery we have can only isolate a few components and cannot be scaled up.

$\delta\varepsilon$: *Do you think that this is only a temporary technological limitation?*

Prof. Crépeau: Well in principle there's no limitation. At first there were only a few proposals about how quantum computers might be built. Nowadays there are probably 15 to 25 known physical systems that display the right kind of behavior; that none of them can be scaled up will be very surprising to me. I think it's just a matter of finding the right components and finding the right systems, and with sufficient variety in the possibilities we will eventually find the right one and get it to work. But, as always it's hard to predict the future.

$\delta\varepsilon$: *What type of people work in quantum computing? Physicists, computer scientists or mathematicians?*

Prof. Crépeau: It's really a combination of all three worlds. Computer scientists have a good knowledge of computability, efficient computations and so on Ű– looking for new algorithms, new ways of using these computers to do efficient tasks; there are mathematicians, mainly in mathematical physics, that are working on the theoretical grounds of quantum computations; and there are experimental physicists that are trying to develop the components of a quantum computer. There are people from all over these fields collaborating and trying to get all the components together, finding new insights as to how we can harness the power of quantum computing and at the same time get the machine actually built.

$\delta\varepsilon$: *Can you tell us about what you're working on right now?*

Prof. Crépeau: What I'm working on right now is on the verge of quantum information with respect to cryptography. For example, do quantum computers make cryptography harder or easier to achieve? That's the sort of large question that I'm concerned with. The fact that we're theoretically working with a quantum computer shows how much the world of cryptography is changing. When you move on to opponents that are equipped with quantum computers, there are some classical proofs that you must revise, because they may no longer be valid in the face of a quantum computer. So there's a whole range of classical results published in the last 30 years that are suddenly no longer valid; these need to be addressed, and proofs must be found to extend the classical theory of information to quantum information.

$\delta\varepsilon$: *If you could solve one open problem, what would it be? It could be in any field.*

Prof. Crépeau: Find a cure for cancer.

$\delta\varepsilon$: *Would you rather solve a problem that has baffled scientists for centuries, or come up with a problem that would baffle scientists for centuries to come? If you can only do one, which would you rather do?*

Prof. Crépeau: (Laughs) The first one.

The Tetris Randomizer

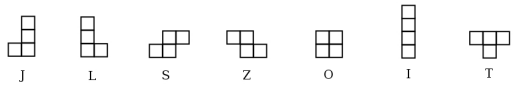
Maya Kaczorowski

How does the game of Tetris generate the sequence of upcoming tetromino pieces? We provide a short reasoning of why the generator is not truly random (or truly pseudorandom) as well as an explanation of how the latest versions of the Tetris randomizer works and what this means for gameplay.

Introduction

Tetris was created in 1985 by the Russian programmer Alexey Pajitnov. Since then, several official and unofficial versions of Tetris have been created on many gaming consoles as the game gains popularity.

The goal of Tetris is to clear rows composed of shapes falling from the top of the screen; the game is lost if the pieces pile up to the top. The player completes these rows by rotating the seven different tetromino pieces, each composed of four blocks, referred to as J, L, S, Z, O, I, and T.



Since the order of upcoming pieces is unpredictable, players do their best to pile pieces without leaving any empty space. If such a space is left and the appropriate piece does not come, usually the I-shaped piece, players will be forced to place another piece, effectively shrinking their playing field and eventually losing the game.

To create more enjoyable gameplay, the programmers of Tetris have, over the years, created a tetromino randomizer designed to produce a more even distribution of the tetromino pieces in the short run.

Not truly random

Is the Tetris randomizer truly random? If tetromino pieces were truly randomly generated, wouldn't there be long streaks of the same piece?

We make the following assumptions about the randomizer:

(1) The selection of pieces is independent

(2) Each shape has an equal probability of being selected

(These do not hold in all versions of Tetris, but are still reasonable assumptions.)

We can now calculate a *lower bound* probability of getting a sequence of at least four of the same tetromino piece out of 1000, an event which we denote A. We split the 1000 piece sequence into 250 sequences of four shapes. For each shape, we denote the events as J_i for a sequence of four J-shaped tetrominoes, L_i, S_i, Z_i, O_i, I_i, and T_i for the other pieces respectively. Note that these cannot occur simultaneously, i.e. in a sequence of four tetromino pieces, we cannot have both four S pieces and four T pieces, so $S_i \cap T_i = \emptyset$.

$$P(A) = 1 - P(A^c) =$$

$$1 - P\left[\bigcap_{i=1}^{250} (J_i^c \cap L_i^c \cap S_i^c \cap Z_i^c \cap O_i^c \cap I_i^c \cap T_i^c)\right]$$

$$= 1 - \prod_{i=1}^{250} P(J_i^c \cap L_i^c \cap S_i^c \cap Z_i^c \cap O_i^c \cap I_i^c \cap T_i^c)$$

by independence

$$= 1 - \prod_{i=1}^{250} P[(J_i \cup L_i \cup S_i \cup Z_i \cup O_i \cup I_i \cup T_i)^c]$$

by de Morgan's law

$$= 1 - \prod_{i=1}^{250} [1 - P(J_i \cup L_i \cup S_i \cup Z_i \cup O_i \cup I_i \cup T_i)]$$

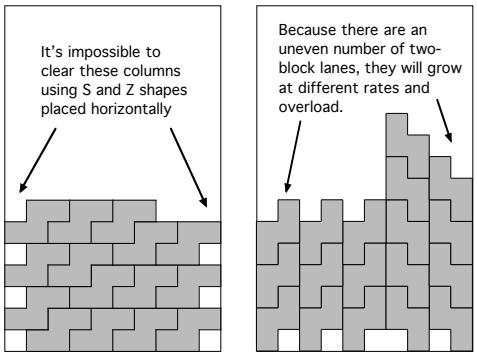
$$= 1 - \prod_{i=1}^{250} [1 - P(J_i) - P(L_i) - P(S_i) - P(Z_i)$$

$$
\begin{aligned}
&\quad - P(O_i) - P(I_i) - P(T_i)] \\
&\text{by inclusion-exclusion} \\
&= 1 - \prod_{i=1}^{250} [1 - \frac{1}{7^4} - \frac{1}{7^4} - \frac{1}{7^4} - \frac{1}{7^4} - \frac{1}{7^4} \\
&\quad - \frac{1}{7^4} - \frac{1}{7^4}] \text{ by counting} \\
&= 1 - \prod_{i=1}^{250} \left[1 - 7\left(\frac{1}{7^4}\right)\right] \\
&= 1 - \left[1 - \frac{1}{7^3}\right]^{250} \\
&= 0.5181.
\end{aligned}
$$

Keeping in mind that this probability is a lower bound, we similarly find a lower bound probability that a run of three pieces occurs out of 1000 is 0.9990. However, in playing a recent version of Tetris, we find that in an experimental run of 1000 pieces, we obtained twelve pairs of the same tetromino piece, but no triples or quadruples. It is then unlikely that the Tetris randomizer selects pieces randomly and independently.

Effects on strategy

Getting a long streak of the same tetromino piece makes play much more difficult. Players whose strategy involves waiting for a certain piece are put at a disadvantage. Furthermore, the uneven distribution of tetromino pieces in the short run will cause pileups, making it difficult for the player to clear rows.



A clearer example of the problem of a truly random Tetris game can be seen if we consider a long run of just Z and S pieces. Note that both the Z and S pieces are three blocks wide and two blocks high, whereas the Tetris playing board is $2n$ blocks wide, with n an odd number. Usually $n = 5$, so the playing board is 10 blocks wide.

If a long sequence of Z and S pieces are placed so that they are three blocks wide, each will consist of one block on the left, one on the right, and two in the middle. If a row is cleared from the board, we will still end up with one more block where the middle of the Tetris piece landed than before the piece was placed, which means that in the long run, the middle columns will always have more blocks in them than the outer ones, leading to pile-up of blocks and a loss of the game.

If instead we place the Z and S pieces so that they are two blocks wide, they must be placed in one of five two block wide lanes that evenly divide the board. The Z and S pieces must be stacked by piece in each lane to prevent empty spaces. As we have an odd number of such lanes, there must be an unequal number of Z and S lanes, growing at unequal rates, so eventually, we must create empty spaces, and then lose the game.

So we see that a very long run of just Z and S pieces, although it has relatively low probability, could arise if tetrominoes were truly randomly generated. However, such a sequence would hasten a loss of the game [1].

The current randomizer

Prior to 2001, the tetromino pieces were generated using a pseudorandom generator. Long runs of the same piece could still occur, although were less likely than if a truly random generator was used.

The Tetris Grand Master game, introduced in 1998 for hyper competitive Tetris gameplay, uses a different randomizer than the original Tetris game. The randomizer maintains a history of the four most recently generated pieces. In generating a piece, it chooses at random one of the seven tetrominoes. If this piece is not found in the history, it is given; if it is in the history, it randomly picks one of the seven tetrominoes again. If a piece outside of the history is not generated after six attempts, it settles for the most recently generated piece [2]. Such a randomizer ensures an even distribution of pieces

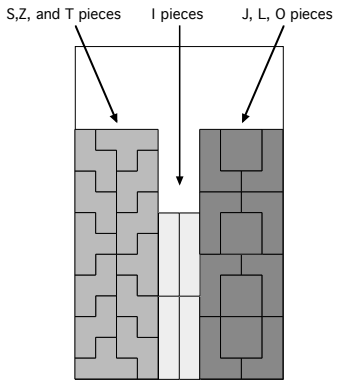
with an unpredictable sequence, and makes it highly unlikely for there to be a long sequence of the same piece.

Since 2001, the Tetris tetromino pieces have been randomized using its own "Random Generator", which generates all seven tetrominoes, one of each, at the same time, then randomizes the sequence [3]. This is analogous to filling a bag with one of each piece, then drawing the pieces out of the bag randomly. When the bag empties, it refills to continue. This randomization guarantees the player an even distribution in the short run. Furthermore, it can never generate a sequence of more than two identical pieces, which occurs if one tetromino is the last in a bag and the second tetromino is the first in the next bag. This rule also limits the waiting time for a specific piece to at most twelve pieces, where the worst case scenario occurs if one piece is the first in a bag and the second the last in the next bag.

Effects on strategy

Avid Tetris players have been able to develop a strategy which theoretically allows infinite gameplay for post-2001 Tetris games which also have the Hold feature, allowing the player to delay the fall of one piece, and at least three piece previews [4].

The field is split into three sections, each filled with different types of pieces. Given the Hold feature and piece preview, this is always possible.
Since the centre field clears more slowly, after a certain number of cycles, a different pattern is followed to clear the field completely.



Due to the even short run distribution of the tetromino pieces, the playing field can be divided into three sections, filling each with only certain types of pieces. Given the Hold feature, players can always fill each section at the same speed and so continually clear rows.

Conclusion

Since games of Tetris use a randomizer that is not truly random, long sequences of the same piece are unlikely to occur. In fact, in post-2001 Tetris games, the tetromino pieces have an even distribution in the short run. In order to extend the game, players should create strategies that do not require many of the same tetromino piece, and if playing a recent version of the game, can rely on the next piece of the same type coming within a maximum of twelve pieces.

References

[1] Burgiel, Heidi. "How to Lose at Tetris." *The Mathematical Gazette*. 81.490 (1997): 194-200.

[2] PetitPrince. "Tetris the Grand Master – a gameplay essay." *B612*. 23 July 2007. 17 December 2008. *http://bsixcentdouze.free.fr/tc/tgm-en/tgm.html.*

[3] "Random Generator." *Tetris Concept*. 13 June 2008. 17 December 2008. *http://www.tetrisconcept.com/wiki/index.php/Random_Generator.*

[4] "Playing forever." *Tetris Concept*. 3 June 2008. 17 December 2008. *http://www.tetrisconcept.com/wiki/index.php/Playing_forever.*

A Short Introduction to Brownian Motion

Phil Sosoe

We define the standard, one dimensional Wiener process ("Brownian motion"), prove its existence, and discuss some basic properties of the sample paths.

Introduction

A stochastic process is an indexed family of random variables. A simple, yet interesting example of a discrete stochastic process is provided by the *symmetric random walk*. In one dimension, this is the process $\{S_n\}_{n\in\mathbb{N}}$ defined by

$$S_n = \sum_{k=1}^{n} X_k,$$

where the X_k are independent, identically distributed random variables defined on a common sample space Ω, taking the values ± 1, each with equal probability. The classical intuitive interpretation of the process S_n is in terms of gambling. Suppose someone repeatedly tossed a fair coin, giving a dollar every time it lands on heads, and asking you to pay a dollar whenever it lands on tails. Assuming you play by the rules, S_n represents your gain after n coin flips. It is a random variable, as it should, for its value is a function of the number of "heads" that occurred up to the n-th coin flip. One can view $S_n(\omega)$ both as the family of random variables $\{S_1, S_2, \ldots\}$ indexed by n, or as a random sequence $(S_n)(\omega)$ for $\omega \in \Omega$.

In this survey I will introduce the continuous analogue of the symmetric random walk, the Wiener process, named for Nobert Wiener, who in his 1923 paper *Differential Space* was the first to give a rigorous construction of the process. This stochastic process is also commonly referred to as Brownian motion, because it serves as a mathematical model for the movement of particles suspended in fluids, as described by botanist Robert Brown in 1827.

Definition of the Wiener Process

A real-valued stochastic process $\{B(t) : t \geq 0\}$ is said to be a (one-dimensional) standard Brownian motion process if it has the following properties:

1. $B(0) = 0$ almost surely.
2. $B(t) - B(s)$ has a Gaussian distribution with mean 0 and variance $t - s$:

   $$B(t) - B(s) \sim N(0, t-s), \quad 0 \leq s < t.$$

   In particular, the distribution of $B(t) - B(s)$ depends only on the difference $t - s$. B is said to have *stationary increments*.
3. B_t has *independent increments*. That is, for any $0 \leq t_1 < \ldots < t_n < \infty$, the random variables

   $$B(t_1) - B(0), B(t_2) - B(t_1), \ldots,$$

   $$B(t_n) - B(t_{n-1})$$

   are independent.
4. $t \mapsto B(t)$ is almost surely continuous.

Here $B(t, \omega)$ is both an uncountable random variable indexed $\{B(t, \omega) : t \geq 0\}$ by the "time" $t \geq 0$, and a random function $t \mapsto B(t, \omega)$. For a fixed ω, the function $B(t, \omega)$ is called a *sample path* of Brownian motion. Hence property 4. above means that almost every sample path of Brownian motion is a continuous function. The first property is a mere convention: we "start" the Brownian motion at the origin. The conditions that the increments be stationary and independent link Brownian motion to the discrete random walk mentioned earlier; the discrete analogues of these conditions are clearly

satisfied by the symmetric random walk. Processes with stationary and independent increments form the important class of Lévy processes, which we will not discuss further. The normal distribution of $B(t) - B(s)$ is in fact a consequence of the continuity of the process, the independent and stationary increments, together with the central limit theorem.

Finally, we make the important observation that

$$\mathrm{Cov}(B(t), B(s)) = \mathbb{E}B(t)B(s) = \min\{s, t\}.$$

To see this, let $s \leq t$, and write

$$\begin{aligned}\mathrm{Cov}(B(t), B(s)) = &\mathrm{Cov}(B(t) - B(s), B(s)) \\ &+ \mathrm{Cov}(B(s), B(s)).\end{aligned}$$

The first term is zero by the independence of increments assumption

$$\mathrm{Cov}(B(t) - B(s), B(s)) =$$

$$\mathrm{Cov}(B(t) - B(s), B(s) - B(0)) = 0.$$

The second term is equal to s by property 2:

$$\mathrm{Cov}(B(s), B(s)) = \mathrm{Var}(B(s) - B(0)) = s.$$

Since Gaussian random vectors are characterized by their covariances, the second and third properties above are equivalent to

2' For $s, t \geq 0$, $\mathrm{Cov}(B(s), B(t)) = \min\{s, t\}$.

3' For $t \geq 0$, $B(t) \sim N(0, t)$.

Existence of Brownian Motion

The Problem

It is not at all clear that the definition given above is not vacuous. How do we know that a stochastic process with the properties listed above even exists? This is obviously an important question, but it may nevertheless seem a bit surprising if you are used to dealing with more elementary and tangible random variables and processes defined in terms of their distribution. In such cases existence issues are largely unproblematic and are usually swept under the rug. When investigating the properties of the sample paths Brownian motion, we come across expressions of the type:

$$\mathbb{P}[B(t) \text{ is differentiable at } 5] =$$

$$\mathbb{P}[B \in \{f : \mathbb{R}^+ \to \mathbb{R} : f \text{ is differentiable at } 5\}],$$

or more generally

$$\mathbb{P}[B \in A] = \mathbb{P}B^{-1}(A),$$

where A is some subset of $C(\mathbb{R}^+)$ of continuous functions on the positive half-line. The distribution $\mathbb{P}B^{-1}$ of Brownian motion is a probability measure on the space $C(\mathbb{R}^+)$, an infinite-dimensional vector space. The probability measures encountered in basic courses on probability are smooth (except possibly at a few points), weighted versions of Lebesgue measure on $\mathbb{R}$. That is, measures $\mathbb{P}$ of the form:

$$\mathbb{P}(A) = \int_A f \, \mathrm{d}x$$

with f a "nice" function. For such distributions, it is a triviality to construct a sample space Ω and a random variable X with distribution $\mathbb{P}$. The reader can check that the random variable

$$X(\omega) = \sup\{x : F(x) < \omega\},$$

where $F(x) = \mathbb{P}(-\infty, x] = \int_{-\infty}^{x} f \mathrm{d}t$ and $\omega \in \Omega = [0, 1]$ has distribution $\mathbb{P}$. No such approach will work in the case of Wiener measure, the distribution of Brownian motion: the "random element" $B(t, \omega)$ takes values in $C(\mathbb{R}^+)$ (as opposed to $\mathbb{R}$), where, among myriads other technical difficulties, no obvious analogues of the probability density function, translation-invariant Lebesgue measure, or even the distribution function F are available.

Levy's Construction of Brownian Motion

We now present Paul Lévy's inductive construction of Brownian motion, following [3]. Brownian motion is constructed as a random element (a $C(\mathbb{R}^+)$-valued random variable) on $[0, 1]$ by ensuring that the properties 1-4 in the definition are satisfied, where we restricted s and t (in property 2) and the t_i (property 3) to the dyadic points:

$$D_n = \{k/2^n : 0 \leq k \leq 2^n\}.$$

Note that $D_n \subset D_{n+1}$. We interpolate linearly between these points; Brownian motion on $[0, 1]$

is realized as uniform limit of these continuous, polygonal paths. Define the set D of all dyadic points

$$D = \bigcup_n D_n.$$

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection of independent, standard normally distributed random variables Z_d, $d \in D$. We set $B(0) = 0$ and $B(1) = Z_1$. For $n \in \mathbb{N}$, B is defined at $d \in D_n$ in such a way that

1. For all $r < s < t$ in D_n, $B(t) - B(s) \sim N(0, t-s)$, and $B(t) - B(s)$ is independent of $B(s) - B(r)$.
2. The vectors $(B(d))_{d \in D_n}$ and $(Z_t)_{t \in D_n \setminus D_{n-1}}$ are independent.

$B(t)$ is already defined on $D_0 = \{0, 1\}$, and we proceed inductively, letting, for $d \in D_n \setminus D_{n-1}$,

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + \frac{Z_d}{2^{(n+1)/2}}.$$

Notice that $d \pm 2^{-n} \in D_{n-1}$, and so $B(d)$ is independent of $(Z_t : t \in D \setminus D_n)$, so the second inductive condition is satisfied. Consider the difference

$$\Delta_n = \frac{1}{2}\left(B(d + 2^{-n}) - B(d - 2^{-n})\right).$$

By induction, Δ_n depends only on $(Z_t : t \in D_{n-1})$, and is hence independent of Z_d. Δ_n and $Z_d/(2^{(n+1)/2})$ being independent $N(0, 2^{-(n+1)})$ random variables, their sum $B(d) - B(d - 2^{-n})$ and their difference $B(d + 2^n) - B(d)$ are independent $N(0, 2^{-n})$ random variables. Thus all pairs of increments $B(d) - B(d - 2^{-n})$, $B(d + 2^{-n}) - B(d)$ for $d \in D_n \setminus D_{n-1}$ are independent. If $d \in D_{n-1}$, we note that the increments are constructed $B(d) - B(d - 2^{-n})$ and $B(d + 2^{-n}) - B(d)$ are constructed from the (independent, by induction) increments $B(d) - B(d - 2^{-j})$ and $B(d + 2^{-j}) - B(d)$, where j is minimal with the property that $d \in D_j$, and disjoint sets of random variables Z_t, $t \in D_n$. Hence the second property holds.

Define the polygonal paths $F_0(t) = tZ_1, 0 \leq t \leq 1$, and $F_n(t) = 2^{-(n+1)/2} Z_t$ for $t \in D_n \setminus D_{n-1}$; $F_n(t) = 0$ for $t \in D_{n-1}$; and $F_n(t)$ is defined to be linear between points of D_{n-1}. Then each F_n is continuous, and we have

$$B(d) = \sum_{i=0}^{\infty} F_i(d)$$

for $d \in D$, as can be seen by induction. The sum has only n non-zero terms if $d \in D_n$.

The claim is now that the series

$$B(t) = \sum_i F_i(t)$$

converges uniformly for $t \in [0, 1]$. To prove this we will make use of the following:

Lemma 1 (Borel-Cantelli lemma.)**.** *If $\{A_n\}$ is a sequence of events in Ω with*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty,$$

then

$$\mathbb{P}[A_n \text{ } i.o.] = 0.$$

where $[A_n$ $i.o.]$ (A_n infinitely often*) is defined as*

$$[A_n \text{ } i.o.] = \bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_m.$$

The proof is elementary, and can be found, for instance, in [1], p. 59. Now, by since the Z_d have standard normal distribution, we have

$$\mathbb{P}[|Z_d| \geq c\sqrt{n}] \leq \exp(-c^2 n/2)$$

for n large and $c > 0$. Hence

$$\sum_{n=0}^{\infty} \mathbb{P}[|Z_d| > c\sqrt{n} \text{ for some } d \in D_n] =$$

$$\sum_{n=0}^{\infty} \sum_{d \in D_n} \mathbb{P}[|Z_d| \geq c\sqrt{n}] < \infty$$

for $c > (2 \log 2)^{1/2}$. By the Borel-Cantelli lemma, there exists $N(\omega)$ such that $|Z_d| < c\sqrt{n}$ for $n \geq N$. This implies that, for $n \geq N$, we have

$$\|F_n\|_\infty \leq c\sqrt{n} 2^{-n/2},$$

so the series defining $B(t)$ does indeed converge to a continuous limit. That the increments of B have the right distribution follows directly from the continuity of B and the properties of the increments. For example, for $r < s < t$, we can choose dyadic sequences r_n, s_n, t_n converging to r, s and t, respectively. Then $B(s) - B(r)$ and $B(t) - B(s)$, being limits of independent Gaussian random variables (note that eventually, $r_n < s_n < t_n$), will be Gaussian and independent. The argument is identical for larger

partitions. Hence $B(t)$ has independent increments and $B(t) - B(s) \sim N(0, t-s)$ whenever $s < t$. We can now extend the definition on $[0,1]$ to $\mathbb{R}^+$ by letting $\{B^n\}_{n\in\mathbb{N}}$ be a collection of independent Brownian motions on $[0,1]$, and defining

$$B(t) = B^{\lfloor t\rfloor}_{t-\lfloor t\rfloor} + \sum_{0\le i<\lfloor t\rfloor} B^i_1,$$

Hence Brownian motion exists.

Properties of the Sample Paths

As mentioned previously, the standard Brownian motion $B(t)$ shares a lot of properties with the symmetric random walk. Three fundamental theorems give us insight into the growth of the process S_n for $n \to \infty$:

1. The (Strong) Law of Large Numbers:

$$\frac{S_n}{n} \to 0,$$

almost surely.

2. The Central Limit Theorem:

$$\mathbb{P}[S_n/\sqrt{n} \le x] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t$$

3. The Law of the Iterated Logarithm:

$$\limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\log\log n}} = 1,$$

almost surely.

For each of these, we have a corresponding theorem for $B(t)$. To get to the Law of Large Numbers, we consider the *time-inverted process*

$$W(t) = \begin{cases} 0 & t = 0 \\ tB(1/t) & t > 0. \end{cases}$$

$W(t)$ in fact has the same distribution as a standard Brownian motion. Indeed, we have

$$\mathrm{Cov}(W(s), W(t)) = ts\min\{1/s, 1/t\} = \min\{t, s\}.$$

The increments of W have joint normal distribution, so they are independent since they are uncorrelated. Hence, by continuity of the paths, we have

$$\lim_{t\to 0} W(t) = \lim_{t\to 0} tB(1/t) = \lim_{s\to\infty} \frac{B(s)}{s} = 0.$$

Thus the proof of the Law of Large Numbers for Brownian motion is surprisingly easier than the classical result.

Corresponding to the Central Limit Theorem, we have *Donsker's Invariance Principle*, a central limit theorem for stochastic processes. There are many variants; a simple formulation in terms of random walks is as follows. Define

$$S(t) = S_{[t]} + (t - [t])(S_{[t]+1} - S[t]).$$

Here $[t]$ denotes the integer part of t, and

$$S_n = \sum_{k=1}^{n} X_k,$$

with X_k *any* random variables with mean 0 and variance 1. $S(t)$ is the continuous function obtained by interpolating linearly between the values of S_n, drawing lines between successive discrete values. Then

$$\Sigma_n(t) = \frac{S(nt)}{\sqrt{n}}$$

converges in distribution in the space $C([0,1])$ to a standard Brownian motion $B(t)$ on $[0,1]$. This result is intuitively appealing as it is the perfect analogue of the central limit theorem. However, one has to be careful how to define convergence in distribution when dealing with random functions rather than random variables. In introductory probability courses, one is told that X_n converges in distribution to X if the distribution function F_n of X_n converges to the distribution F of X at every point of continuity. When dealing with random variables taking values in a functional space, this definition is clearly inadequate. It turns out that the right abstract definition for the concept is weak convergence. A sequence of random elements X_n with values in a metric space (E, d) converges weakly to X if

$$\mathbb{P}[X_n \in A] \to \mathbb{P}[X \in A]$$

whenever $\mathbb{P}[X \in \partial A] = 0$.

As for the Law of the Iterated Logarithm, we have the two results:

$$\limsup_{t\to\infty} \frac{|B(t)|}{\sqrt{2t\log\log t}} = 1.$$

and

$$\limsup_{t\to 0} \frac{|B(t)|}{\sqrt{2t\log\log(1/t)}} = 1.$$

The second result follows from the first one by time inversion. It also highlights a difference between the discrete random walk and Brownian motion; in the discrete case, there is no asymptotic behavior at 0 to speak of.

Beyond the growth properties, one can ask how regular the paths of Brownian motion are. The paths are continuous by definition, and hence uniformly continuous on $[0,1]$. This means that for some (a priori) random function $\epsilon(h)$ with $\epsilon(h) \to 0$ as $h \downarrow 0$,

$$\limsup_{h\to 0} \sup_{0\le t\le 1-h} \frac{|B(t+h)-B(t)|}{\epsilon(h)} \le 1.$$

$\epsilon(h)$ is referred to as the modulus of continuity of B (on [0,1]). A careful examination of Lévy's construction shows that ϵ is not in fact random. If $h > 0$ is sufficiently small, and $0 \le t \le 1-h$, we have

$$C_1\sqrt{h\log(1/h)} \le |B(t+h)-B(t)|$$
$$\le C_2\sqrt{h\log(1/h)}.$$

As a corollary, the sample paths of Brownian motion can be shown to be Hölder continuous with exponent α for every $\alpha < 1/2$. A function f is said to be Hölder continuous if

$$|f(x)-f(y)| \le |x-y|^\alpha.$$

The variation properties of Brownian are quite bad. For instance, $B(t)$ is monotone on no inteval. Indeed, if $B(t)$ is monotone on $[a,b]$ for $0 < a < b < \infty$, then for any partition

$$a = t_0 < \ldots < t_n = b,$$

all the increments $B(t_i) - B(t_{i-1})$ must have the same sign. Since the increments are independent, this event has $2\cdot 2^{-n}$. Letting $n \to \infty$, we see that with probability one, $B(t)$ is not monotone on $[a,b]$. Considering all intervals with rational endpoints, we see that with probability 1, $B(t)$ is monotone on no interval. It is not too hard to show that, fixing any point $t_0 \in \mathbb{R}$, $B(t)$ is almost surely not differentiable at t_0. A harder result, due to Paley, Wiener, and Zygmund is that almost surely, $B(t)$ is *nowhere differentiable*. Note that the former result does not imply the latter, because, even though t_0 is arbitrary in the first result, there are uncountably many $t \in \mathbb{R}$.

Coda

We have only been able to give a very superficial overview of the theory of Brownian motion. Important topics we have left completely untouched are the study of Brownian motion as a continuous-time martingale, and the Markov property. The multidimensional process and geometric aspects are also of great interest; just as one can study the transience and recurrence of random walks on the lattice $\mathbb{Z}^d$, one can ask the same questions about sets in $\mathbb{R}^d$ and Brownian motion. Another important aspect of the theory is the close relation between harmonic functions and Brownian motion. A famous theorem of Kakutani characterizes the solution of the Dirichlet problem on a domain U with continuous boundary data φ as the expectation

$$u(x) = \mathbb{E}_x[\varphi(B(\tau_{\partial U})],$$

where $\tau_{\partial U} = \inf\{t : B(t) \in \partial U\}$ is the first time B hits the boundary and $\mathbb{E}_x$ is expectation with respect to a measure making $\{B(t) : t \ge 0\}$ a Brownian Motion started at $x \in U$. All these topics are of great relevance to current research, but they require a certain amount of analytic machinery for their study. For anyone with a solid understanding of basic probability and an interest in the subject, the excellent book [2] is a good place to start.

References

[1] Patrick Billingsley, *Probability and Measure*, Third Edition, Wiley, 1995.

[2] Peter Mörters, Yuval Peres. *Brownian Motion*, 2008.

[3] Yuval Peres. *An Invitation to the Sample Paths of Brownian Motion*, 2001.

Available at: **www.stat.berkeley.edu/~peres/bmbook.pdf**.

[4] Daniel W. Stroock, *Probability Theory: An Analytic View*, Cambridge University Press, 1999.

The Navier-Stokes Equations

Daniel Shapero

We present a derivation of the Navier-Stokes partial differential equations, which describe the flow of incompressible fluids. We then outline reasons that make the problem of proving existence for all time of a smooth solution, given smooth initial data, very difficult.

Introduction

In this article we present a derivation of the Navier-Stokes equations of fluid flow and show some basic results related to them. George Gabriel Stokes was the first mathematician to correctly derive the equations that bear his name in 1858. More than a century later, fundamental questions about the Navier-Stokes equations have yet to be answered: one of the six remaining Clay Millenium Prize problems is to prove or disprove that, given a smooth initial velocity field, a solution exists which is defined and differentiable for all times. Despite the broad applicability of fluid dynamics in describing phenomena from blood flow to meteorology to astrophysics, these questions, which are a basic sanity check for the validity of any mathematical model, have yet to be answered. Furthermore, modern physics has yet to satisfactorily describe the phenomenon of fluid turbulence.

Preliminaries

The mathematical tool of which we will make greatest use is the divergence theorem: let $\vec{F} : \mathbb{R}^3 \to \mathbb{R}^3$ be a smooth vector field, Ω a domain in $\mathbb{R}^3$ with smooth boundary $\partial\Omega$ and unit outward normal $\hat{n}$. Then the net flux of $\vec{F}$ through $\partial\Omega$ is equal to the volume integral of the divergence of $\vec{F}$ over all of Ω, or

$$\int_{\partial\Omega} \vec{F} \cdot \hat{n} \mathrm{d}\sigma = \int_{\Omega} \nabla \cdot \vec{F} \mathrm{d}\tau.$$

We can, using the Einstein summation convention that a repeated index implies a sum over that index, write $\vec{F} = F_j \hat{\mathbf{e}}_j$; in this form, the divergence theorem states that

$$\int_{\partial\Omega} F_j n_j \mathrm{d}\sigma = \int_{\Omega} \frac{\partial}{\partial x_j} F_j \mathrm{d}\tau.$$

This will be useful when we have to apply the divergence theorem in a slightly modified form to tensor fields. A rigorous treatment of tensors is beyond our scope, but if you are unfamiliar with them you can think of tensors as higher-dimensional generalizations of scalars, vectors and matrices. Every tensor has a number called its rank associated to it: a scalar has rank 0, a vector rank 1 and a matrix rank 2. We will not have to consider tensors of rank greater than two, but higher-rank tensors do arise in fields such as general relativity. Much as you can consider a vector field in some domain of $\mathbb{R}^n$ and do calculus with these vector fields, you can apply familiar analytic tools to the study of tensor fields. You can think of a rank 2 tensor field as associating to each point of $\mathbb{R}^n$ a matrix, and the divergence theorem still holds in this context: if $\overline{T}$ is some smooth tensor field on $\mathbb{R}^3$ with components T_{ij}, then

$$\int_{\partial\Omega} \overline{T}\hat{n} \mathrm{d}\sigma = \int_{\Omega} \nabla \cdot \overline{T} \mathrm{d}\tau,$$

or, in components,

$$\int_{\partial\Omega} T_{ij} n_j \mathrm{d}\sigma = \int_{\Omega} \frac{\partial}{\partial x_j} T_{ij} \mathrm{d}\tau.$$

The divergence of a rank 2 tensor field – one could call it matrix field – is a vector field, and is defined in components by $(\nabla \cdot \overline{T})_i = \frac{\partial}{\partial x_j} T_{ij}$, the pointwise divergence of the matrix's rows. We will have cause to use this machinery when considering the stress tensor of a fluid.

Finally, we will use the fundamental lemma of the variational calculus frequently: if $F : \mathbb{R}^n \to \mathbb{R}$ is smooth and, for every domain $\Omega \subset \mathbb{R}^n$, $\int_\Omega F \mathrm{d}\tau = 0$, then F is identically zero.

Mass Conservation

To begin, we let ρ be the density of the fluid at a given time and position, and $\vec{u} = (u, v, w)$ the fluid's velocity. Our goal is to find the partial differential equations which will govern the evolution of the velocity field $\vec{u}$ over time. First we derive the mass conservation equation. The total mass of fluid in a region Ω at time t is given by

$$\int_\Omega \rho \mathrm{d}\tau,$$

and the mass flux through $\partial\Omega$ is

$$\int_{\partial\Omega} \rho\vec{u} \cdot \hat{n} \mathrm{d}\sigma.$$

The rate of change of mass in Ω must be equal to minus the mass flux through $\partial\Omega$, if there are no sources or sinks:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \rho \mathrm{d}\tau = -\int_{\partial\Omega} \rho\vec{u} \cdot \hat{n} \mathrm{d}\sigma.$$

We can apply the divergence theorem to the right-hand side, and differentiate the left-hand side under the integral sign:

$$\int_\Omega \frac{\partial\rho}{\partial t} \mathrm{d}\tau = -\int_\Omega \nabla \cdot (\rho\vec{u}) \mathrm{d}\tau.$$

Finally, rearranging the terms of the last equation applying the fundamental lemma of the variational calculus to the volume integrals implies that

$$\frac{\partial\rho}{\partial t} + \nabla \cdot (\rho\vec{u}) = 0,$$

the conservation of mass equation. From now on we will have to assume that ρ is identically a constant in order to make any progress. In this case, the flow is called incompressible. While no fluid is truly incompressible – such a fluid could not transmit sound – this is often a reasonable assumption, in that the longitudinal compressions of the fluid are minute compared to the length scale of the flow. In this case, the mass conservation equation reduces to the statement that $\nabla \cdot \vec{u} = 0$.

Momentum Conservation

The Navier-Stokes equations come from applying Newton's second law $F = ma$ to the fluid. In our case, the mass times acceleration term will be $\rho\frac{\mathrm{D}\vec{u}}{\mathrm{D}t}$, the total derivative with respect to time. Since the velocity of a fluid element at (x, y, z) is (u, v, w), using the chain rule we have

$$\begin{aligned}\frac{\mathrm{D}\vec{u}}{\mathrm{D}t} &= \frac{\partial\vec{u}}{\partial t} + \frac{\partial\vec{u}}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial\vec{u}}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial\vec{u}}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t} \\ &= \frac{\partial\vec{u}}{\partial t} + \frac{\partial\vec{u}}{\partial x}u + \frac{\partial\vec{u}}{\partial y}v + \frac{\partial\vec{u}}{\partial z}w \\ &= \frac{\partial\vec{u}}{\partial t} + \vec{u} \cdot \nabla\vec{u}.\end{aligned}$$

The differential operator $\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \vec{u} \cdot \nabla$ is called the material derivative for the flow $\vec{u}$, and the non-linearity of the second term – the inertial term – is the source of the difficulty in solving the Navier-Stokes equations.

In accordance with Newton's laws, $\rho\frac{\mathrm{D}\vec{u}}{\mathrm{D}t}$ should be equal to the sum of all forces acting on the fluid. We would expect to account for body forces like gravity or electrostatic forces, but since we are describing a continuum we must also consider the forces of one fluid element on another. We represent these surface forces by a tensor $\overline{T}$ with components T_{ij}, where the i,j component of this stress tensor at a point (x, y, z) is the surface stress acting on the i-th face in the $\hat{\mathbf{e}}_j$-direction of an infinitessimal tetrahedron δV surrounding (x, y, z). For our purposes, the stress tensor is symmetric: $T_{ij} = T_{ji}$. Situations where this is not the case are rare and we will be content to exclude them. A field where this assumption does not hold is magnetohydrodynamics, the study of conducting fluids. As if the Navier-Stokes equations were not difficult enough, here they must simultaneously be solved with Maxwell's equations of electromagnetism!

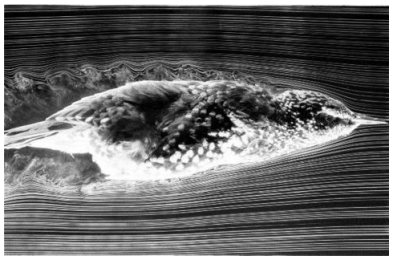
In any case, letting $\vec{f}$ be the body forces acting on the fluid, the momentum conservation relation can be stated as

$$\begin{aligned}\int_\Omega \rho\frac{\mathrm{D}\vec{u}}{\mathrm{D}t}\mathrm{d}\tau &= \int_{\partial\Omega} \overline{T}\vec{n}\mathrm{d}\sigma + \int_\Omega \vec{f}\mathrm{d}\tau \\ &= \int_\Omega \left(\nabla \cdot \overline{T} + \vec{f}\right)\mathrm{d}\tau,\end{aligned}$$

where we applied the divergence theorem to the surface integral. Using the fundamental lemma of the variational calculus, we arrive at

$$\rho\frac{\mathrm{D}\vec{u}}{\mathrm{D}t} = \nabla \cdot \overline{T} + \vec{f}, \tag{1}$$

the differential form of momentum conservation for fluids. Note how similar this looks to $F = ma$, but for the inclusion of the $\nabla \cdot \overline{T}$ term, to account for the forces that one fluid element exerts on another.



Starling in a wind tunnel, shedding vortices off its back.

Navier-Stokes Equations

We have not yet made any assumptions about the relation between the stress tensor T_{ij} and the components of the fluid velocity u_i; in fact, our treatment thus far has been general and can be applied to incompressible solids. The pressure p is the force per unit area exerted normal to a fluid element at a given point. Let

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

be the Kronecker symbol. Stokes derived that, for some constant μ called the dynamic viscosity,

$$T_{ij} = -p\delta_{ij} + \mu\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right) \qquad (2)$$

from three hypotheses. First, the components of the stress tensor should be linear functions of $\frac{\partial}{\partial x_j}u_i$; second, each T_{ij} should be zero if there is no deformation of the fluid; and finally, the fluid is isotropic, which means that there is no "preferred" direction for the stress of a fluid element to point. First, note that $\frac{\partial}{\partial x_j}\left(\frac{\partial u_j}{\partial x_i}\right) = \frac{\partial}{\partial x_i}\left(\frac{\partial u_j}{\partial x_j}\right)$ is the i-th partial derivative of $\nabla \cdot \vec{u}$, so that this term is zero because the flow is incompressible. Substituting our assumptions of equation (2) about the stress tensor into the momentum equation (1) yields the Navier-Stokes equations

$$\rho\frac{\mathrm{D}\vec{u}}{\mathrm{D}t} = \rho\left(\frac{\partial \vec{u}}{\partial t} + \vec{u}\cdot\nabla\vec{u}\right) = -\nabla p + \mu\nabla^2\vec{u} + \vec{f},$$

where the Laplacian is taken component-wise. Letting $\nu = \mu/\rho$, the equations in their full glory are

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + \\
&\quad \nu\nabla^2 u + f_x \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial y} + \\
&\quad \nu\nabla^2 v + f_y \\
\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + \\
&\quad \nu\nabla^2 w + f_z \\
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} &= 0.
\end{aligned}$$

The quantity ν is actually of more physical significance than μ; ν is called the kinematic viscosity, and measures the viscosity per unit density. One can immediately see the gross nonlinearities present in these equations coming from the inertial term $\vec{u}\cdot\nabla\vec{u}$. The few instances where the Navier-Stokes equations are exactly solvable are generally those in which $\vec{u}\cdot\nabla\vec{u}$ is zero or small enough to be negligible.

The Reynolds number and turbulence

One of the most fascinating phenomena in fluid dynamics is turbulence. Accordingly, we want to be able to quantify the relative effects of the fluid's inertia and the viscous dissipation of momentum. This ratio will dictate many of the qualitative properties of the fluid flow, including the transition to turbulence. To this end we define the dimensionless quantity called the Reynolds number.

Let U denote some typical velocity, L a length scale over which the velocity can change of order U and ν the kinematic viscosity. The choices of U and L are somewhat subjective, but this is only to give us a rough idea of the flow regime and total precision is unnecessary. For example, airflow over a plane wing would have L between two and eight meters and U roughly 300 meters per second. The Reynolds number is defined as $R = \frac{UL}{\nu}$. To see that this should be important, note that the inertial term $\vec{u}\cdot\nabla\vec{u}$ is of the order of

$\frac{U^2}{L}$, and the viscous term $\nu\nabla^2\vec{u}$ has units of $\frac{\nu U}{L^2}$. Then the ratio of the inertial term to the viscous term is of the order of the Reynolds number R. Two flows with the same Reynolds number, even if they have different viscosities, length scales or velocities, are dynamically similar. In the example of the airplane, the kinematic viscosity of air is $\nu = 0.15\frac{\text{cm}^2}{\text{s}}$, so $R \geq 40,000,000$.

Flows with Reynolds number less than 1 are dominated by the effects of viscosity, and display a number of characteristic properties. One of them is a high degree of reversibility. A famous experiment goes as follows: pour glycerin between two concentric cylinders, and when the fluid has come to rest inject a small blob of dye between the cylinders with a syringe. Turn the outer cylinder four times clockwise; the viscosity of the glycerin will shear out the dye blob into a ring. Turn the outer cylinder four times counter-clockwise, and the dye will return slightly blurred to its original position.

1980 Mt. St. Helens explosion, showing turbulent flow.

Reynolds number flows above 4000 are characterized by turbulent, chaotic motion. Turbulent fluid flow is still one of the most baffling phenomena in physics, even after hundreds of years of inquiry. A high Reynolds number flow is unstable to small perturbations, so that a minute disturbance in the initial condition of a flow yields an entirely different evolution. Instability is one of the major obstructions to accurate computer simulation of fluid flows, thus making it difficult to gain insight via numerical experiments.

But, the most striking features of turbulence are the vast spectrum of length scales on which complex time-dependent motion is observed, and the rapid and tempestuous changes in pressure and velocity through time and space. In a fluid with small viscosity, energy concentrated in eddies and vortices is dissipated only at minute lengths. Eddies and self-similar structures can be observed at nearly all sizes, as can be seen in the photo of Mt. St. Helens where billowing clouds of smoke contain almost fractal-like copies of themselves. Kolmogorov made tremendous conceptual contributions to the theory by postulating the natural time, length and velocity scales of turbulent flow, near which small vortices shed their energy into heat. The large range of relevant length scales is another difficulty encountered in numerical analysis: the number of mesh points needed in a finite volume method analysis would have to be gargantuan.

The problem and some partial results

It has yet to be proven that on a torus or in all space - let alone inside some arbitrary smooth surface - a solution of the Navier-Stokes equations exists which is smooth for all times given a smooth divergence-less initial flow field. The Clay Mathematics Institute has offered a USD$1,000,000 prize for a correct proof or a counter-example of the claim.

Why has this not been solved? The usual paradigm of non-linear PDE theory is to use functional analysis to find weak solutions of the PDE, which satisfy the differential equation in the mean rather than pointwise, and use calculus and measure-theoretic estimates to show that the weak solutions are smooth. Jean Leray proved in 1934 that weak solutions to the Navier-Stokes equations exist and these solutions are smooth up to some time T, but was unable to demonstrate regularity for all times.

This approach requires finding quantities that dictate the solution's behaviour in some sense, such as upper bounds or asymptotic growth rates. For example, if u is harmonic in a domain Ω, u and its derivatives satisfy a certain growth condition which allows one to conclude that harmonic functions are, in fact, analytic! These controlling quantities will vary depending on the problem, and are very diverse.

This approach is unsuccessful in high Reynolds number flows due to the lack of strong controlled quantities. Global energy conservation does not sufficiently bound kinetic energies at small length scales, leading to singularities. Kinetic energy diffuses to smaller length scales before it can be dissipated by the fluid's viscosity, so this concentration of energy at small scales is a phenomenon that one has to worry about. No one has found any other sufficiently coercive controlled quantities; finding them is no easy task, since turbulent flows are highly asymmetric.

At low Reynolds number, which corresponds to high viscosity or initial velocities of small magnitude, viscous dissipation of energy prevents any singularities from forming. In this case the solution remains smooth for all times, and can closely resemble the heat equation with a small perturbation. Regularity has also been proven for flows in only two dimensions or with a symmetry about some axis, where energy conservation does prevent blow-up. Generalizing this approach to three dimensions has proven fruitless.

The most recent development concerns the size in space-time of any blow-up that does occur. Caffarelli, Kohn and Nirenberg proved in 1982 that, if a blow-up does occur in the solution, it cannot fill out a curve in space-time: the one-dimensional Hausdorff measure of the flow's singular set is zero. The result has not been improved and is the forefront of our progress towards the full result. While partial regularity is not a full resolution of the problem, the result is encouraging in that a singularity can only occupy a very small set.

Why do we care?

The Navier-Stokes equations purport to be a valid mathematical model of a physical phenomenon. As such, one would certainly expect that a unique solution exists for all times and that it is smooth. Otherwise, we would have to question whether our model were correct, as this cannot describe reality.

As a parallel, in electromagnetism one can demonstrate that the electrostatic potential V satisfies Laplace's equation $\nabla^2 V = 0$ where there is no charge density. So long as the divergence theorem can be applied to the domain, it is easy to show that the solution is unique. In dimension two, the existence and differentiability of solutions to Laplace's equation on a simply-connected domain are guaranteed by the Riemann mapping theorem and other tools of complex analysis. The general theory of elliptic operators comes into play in higher dimensions. But, the end result is the same: our mathematical model always predicts precisely one smooth solution, and questions about its validity will be based on physical rather than mathematical grounds.

The Navier-Stokes equations have yet to fulfill these sanity checks. In spite of this unfortunate state of affairs, experiment has demonstrated that, to the best accuracy that modern numerical analysis can discern, the Navier-Stokes equations describe the motion of viscous incompressible fluids remarkably well. Numerical analysis for non-linear PDE and especially Navier-Stokes is notoriously difficult, and is needed in many fields of science and engineering. A resolution of the existence-smoothness question would likely shed some light on the very practical issue of how to obtain approximate solutions.

Finally, fluid turbulence is plainly visible to the naked eye and yet physics has yet to provide a truly satisfactory description of it. An apocryphal quote attributed both to Werner Heisenberg and to Horace Lamb has him asking God, "Why relativity? And why turbulence?", being hopeful about the former. As a tantalizing problem of practical and theoretical significance which has thus far defied our best efforts, its resolution will require exciting and novel ideas of mathematics and physics.

References

[1] D.J. Acheson, *Elementary Fluid Dynamics*, Oxford University Press: Oxford, 1990.

[2] L. Caffarelli, R. Kohn and L. Nirenberg, *Partial regularity of suitable weak solutions of the Navier-Stokes equations*, Communications on Pure and Applied Mathematics, 35 (1982), pp. 771-831.

[3] O. Gonzales and A.M. Stuart, *A First Course in Continuum Mechanics*, Cambridge University Press: Cambridge, 2008.

[4] O.A. Ladyzhenskaya, *Mathematical Theory of Viscous Incompressible Flow*, Gordon and Breach: New York, 1969.

[5] J. Leray, *Sur le mouvement d'un liquide visqueux emplissant l'espace*, Acta Mathematica, 63 (1934), pp. 193-248.

[6] T. Tao, "Why global regularity for Navier-Stokes is hard", *What's New* 17 Mar. 2008. 1 Feb. 2009. http://terrytao.wordpress.com/2007/03/18/why-global-regularity-for-navier-stokes-is-hard/

Glimpse of Infinity: A Brief Overview of Asymptotic Analysis

Ioan Filip

Asymptotic analysis provides powerful tools to investigate the behavior of ordinary and partial differential equations at limiting points, and is hence of great interest in physics and modeling problems, but also in the analysis of algorithms for instance. In fact, questions of limiting behavior pervade much of mathematics and thus asymptotic analysis is an essential study in its own right. In this paper, we briefly introduce some fundamental notions in asymptotic analysis and illustrate the WKB method to approximate second-order differential equations. Our motivation for asymptotic analysis comes from studying the solutions of the eigenvalue problem of the Laplacian on the ellipse.

Eigenfunctions of the Laplacian on the Ellipse

The context is as follows. Let $\Omega \subset \mathbb{R}^2$ be the ellipse defined by the equation

$$x^2 + \frac{y^2}{1-a^2} = 1,\ 0 \le a < 1, \tag{1}$$

with foci at $(\pm a, 0)$. Recall that the *Laplace operator*, denoted by Δ, is the differential operator given by

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \tag{2}$$

We are interested in the *eigenvalue problem* for this operator: finding non-trivial solutions of

$$\Delta u(x,y) + \lambda_j^2 u(x,y) = 0 \text{ in } \Omega,$$

for constants λ_j^2, called the *eigenvalues* of Δ (we know the eigenvalues are positive). We also impose the *Neumann boundary condition*:

$$\frac{\partial u}{\partial \nu} = 0 \text{ on } \partial\Omega,$$

where $\frac{\partial}{\partial \nu}$ is the derivative in the exterior normal direction to Ω. To solve the problem, we apply separation of variables. First, define elliptical coordinates (ϕ, ρ) as follows:

$$(x,y) = (a\cos\phi\cosh\rho, a\sin\phi\sinh\rho),$$

where

$$\begin{cases} 0 \le \rho \le \rho_{\max} = \cosh^{-1} a^{-1}, \\ 0 \le \phi \le 2\pi. \end{cases}$$

Note that the lines $\rho = const$ are confocal ellipses and $\phi = const$ confocal hyperbolae. Moreover, the foci occur at $\phi = 0, \pi$ and the origin is at $\rho = 0, \phi = \pi/2$. Expressing the Laplace operator in the coordinates (ϕ, ρ), we obtain

$$\frac{1}{a^2(\sin^2\phi\cosh^2\rho + \cos^2\phi\sinh^2\rho)} \times \\ \times \left(\frac{\partial^2 u}{\partial\phi^2} + \frac{\partial^2 u}{\partial\rho^2}\right) + \lambda^2 u = 0.$$

Assume u is of the form $u = f(\phi)g(\rho)$, we plug this product into the above equation. Since

$$\frac{\partial^2 u}{\partial\rho^2} = f(\phi)g''(\rho),\ \frac{\partial^2 u}{\partial\phi^2} = f''(\phi)g(\rho)$$

the equation becomes

$$\frac{1}{a^2(\cosh^2\rho - \cos^2\phi)}(gf'' + fg'') + \lambda^2 fg = 0$$

or, equivalently

$$\frac{f''}{f} - a^2\lambda^2\cos^2\phi = -\frac{g''}{g} - a^2\lambda^2\cosh^2\cosh^2\rho.$$

Because the left-hand and the right-hand sides are functions of different variables, both sides of the equations must be equal to the same constant. Introducing a constant of separation C, we get the system of second-order ordinary differential equations:

$$\begin{cases} \frac{f''}{f} - a^2\lambda^2\cos^2\phi = -C \\ \frac{g''}{g} + a^2\lambda^2\cosh^2\rho = C \end{cases}$$

or, equivalently

$$f''(\phi) + f(\phi)(C - a^2\lambda^2\cos^2\phi) = 0 \tag{3}$$
$$g''(\rho) - g(\rho)(C - a^2\lambda^2\cosh^2\rho) = 0. \tag{4}$$

Equations (3) and (4) are known as *Mathieu's equations* named after the French mathematician Émile Léonard Mathieu who first introduced them. The solutions of these equations are called the ordinary Mathieu functions (or the angular functions for (3)) and the *modified* Mathieu functions (or the radial functions for (4)). The theory of Mathieu functions is well understood and we refer the reader to [4]. We are mainly interested in the behavior of the solutions to (3) and (4) as the parameter λ, or the eigenvalue of the Laplacian, tends to infinity. We are thus naturally led to the analysis of asymptotics.

The WKB method

In this section, we follow [5]. We begin with a few definitions from [2].

Definition. Let $f(z)$ and $g(z)$ be two (complex-valued) functions defined on a domain D with $z_0 \in \overline{D}$. We write

$$f(z) = o(g(z)) \text{ as } z \to z_0 \text{ from } D$$

if for any $\epsilon > 0$, there exists some $\delta(\epsilon) > 0$ such that $|f(z)| \leq \epsilon |g(z)|$ for $z \in D$ and $0 < |z - z_0| < \delta(\epsilon)$.

Definition. A sequence of functions $\{\phi_n(z)\}_{n=0}^{\infty}$ is an *asymptotic sequence* as $z \to z_0$ from the domain D if we have that $n > m \Rightarrow \phi_n(z) = o(\phi_m(z))$ as $z \to z_0$. We allow $z_0 = \infty$.

Definition. Let $\{\phi_n\}_{n=0}^{\infty}$ be an asymptotic sequence as $z \to z_0$. Then the sum $\sum_{n=0}^{N} a_n \phi_n(z)$ is an asymptotic approximation as $z \to z_0$ of a function $f(z)$ if $f(z) - \sum_{n=0}^{N} a_n \phi_n(z) = o(\phi_N(z))$ as $z \to z_0$. If $\{a_n\}_{n=0}^{\infty}$ is a sequence such that the above holds for all N, then the formal series $\sum_{n=0}^{\infty} a_n \phi_n(z)$ is called an *asymptotic series* and it is an asymptotic expansion of $f(z)$ as $z \to z_0$. We write

$$f = \sum_{n=0}^{\infty} a_n \phi_n(z) \text{ as } z \to z_0.$$

We sometimes write $\sim$ instead of equality in the above expansion.

Our objective is to study the asymptotic theory (as the parameter $\lambda \to \infty$) of ordinary homogeneous linear differential equations of second order in standard form

$$y'' + q(x, \lambda) y = 0. \tag{5}$$

Note that the Mathieu equations fall within this family, but dealing in full generality here is advantageous. We assume that $q(x, \lambda)$ has the form

$$q(x, \lambda) = \sum_{n=0}^{\infty} q_n(x) \lambda^{2k-n},$$

where $q_n(x)$ are independent of λ and $k \in \mathbb{N}^{\times}$ is fixed. (We are essentially saying that the asymptotic expansion of $q(x, \lambda)$ in terms of λ does not exhibit a 'severe' singularity, but only a pole, as $\lambda \to \infty$.) Further suppose that $q_0(x)$ does not vanish in the domain of x we consider. This assumption is crucial in our derivations which follow. The case when q_0 vanishes is discussed, in certain particular cases, in the section titled *Transition Points*. First, we do some computations formally below, and then we proceed to deal with the convergence issues.

Formal Solutions Assume that the solution to (5) has an expansion of the form

$$y(x, \lambda) = \exp\left\{ \sum_{0}^{\infty} \beta_n(x) \lambda^{k-n} \right\}. \tag{6}$$

Substituting this expression into (5), we obtain

$$\sum \beta_n(x)'' \lambda^{k-n} + \left(\sum \beta_n(x)' \lambda^{k-n} \right)^2 + \\ + \sum q_n \lambda^{2k-n} = 0.$$

Grouping the coefficients of λ^{2k-n} we get the following relations

$$\beta_0'^2 + q_0 = 0 \tag{7}$$

$$2\beta_0' \beta_n' + q_n + \sum_{m=1}^{n-1} \beta_n' \beta_{n-m}' = 0, \\ \text{for } n = 1, \ldots, k-1 \tag{8}$$

$$2\beta_0' \beta_n' + q_n + \sum_{m=1}^{n-1} \beta_m' \beta_{n-m}' + \beta_{n-k}'' = 0, \\ \text{for } n = k, k+1, \ldots \tag{9}$$

We obtain two independent formal solutions of this type. Note also that we have assumed $q(x, \lambda)$ has a pole of even order at $\lambda = \infty$, and if it had a pole of odd order, then we would expand in powers of $\lambda^{1/2}$ instead of λ. We now prove that the solutions of (5) can indeed be asymptotically represented in the above form.

Asymptotic Solutions Fix $N \in \mathbb{N}^{\times}$ and set

$$Y_j = \exp\left\{\sum_{n=0}^{2k+N-1} \beta_{nj}(x)\lambda^{k-n}\right\}, \text{ for } j = 1, 2, \tag{10}$$

where $\beta'_{01} = -\beta'_{02}$ and for each j, the β_{nj} satisfy the recurrence relations (7), (8) and (9) listed above. Observe that the coefficients β_{nj} are completely determined by $q_0, \ldots, q_{2k+N-1}$ and certain derivatives. We say that the q_n are *sufficiently often differentiable* if all the derivatives to determine the β_{nj}'s exist and are continuous. Let now $x \in I := [a, b]$ and let λ vary over a domain S defined by $|\lambda| \geq \lambda_1, \phi_0 \leq \arg(\lambda) \leq \phi_1$. We show the following.

Theorem. *Suppose that for each $\lambda \in S$ $q(x, \lambda)$ is continuous over I. Assume also that the $q_n(x)$ are sufficiently often differentiable in I, and that*

$$q(x, \lambda) = \sum_{0}^{2k+N-1} q_n(x)\lambda^{2k-n} + O(\lambda^{-N})$$

holds uniformly in x and $\arg(\lambda)$, as $\lambda \to \infty$ in S. Let also

$$Re\{\lambda^k[-q_0(x)]^{1/2}\} \neq 0$$

for $\lambda \in S$ and $x \in I$. Then the differential equation (5):

$$y'' + q(x, \lambda)y = 0$$

has a system of linearly independent solutions $y_1(x), y_2(x)$ satisfying

$$\begin{aligned} y_j &= Y_j[1 + O(\lambda^{-N})] \\ y'_j &= Y'_j[1 + O(\lambda^{-N})] \end{aligned}$$

uniformly in x and $\arg(\lambda)$, as $\lambda \to \infty$ in S.

Proof. Since $Re\{\lambda^k[-q_0(x)]^{1/2}\} \neq 0$, and from (7), $\beta_0'^2 + q_0 = 0$, we may choose β_{01} and β_{02} so that for each $\lambda \in S$ $Re\{\lambda^k\beta_{01}(x)\}$ and $Re\{\lambda^k\beta_{02}(x)\}$ are increasing and decreasing functions of x respectively. From (10), we conclude that

$$|Y_1| = \left|\exp\left\{\sum_{0}^{2k+N-1} \beta_{n1}(x)\lambda^{k-n}\right\}\right|$$

is increasing for λ sufficiently large, and similarly $|Y_2(x)|$ is decreasing. We substitute

$$y_1(x) = Y_1(x)z(x)$$

in the equation (5) to obtain the new equation

$$\begin{aligned} &z'' + 2\frac{Y_1''}{Y_1}z' + F(x, \lambda)z = 0 \iff \\ &\frac{d}{dx}\left[Y_1^2(x)\frac{dz}{dx}\right] + Y_1^2(x)F(x, \lambda)z = 0, \end{aligned}$$

where

$$\begin{aligned} F(x, \lambda) = \frac{Y_1''}{Y_1} + q = \sum_{n=0}^{2k+N-1} \beta''_{n1}(x)\lambda^{k-n} + \\ + \left(\sum_{0}^{2k+N-1} \beta'_{n1}\lambda^{k-n}\right)^2 + q = O(\lambda^{-N}) \end{aligned}$$

from the assumptions of the theorem and the recurrence relations (7), (8), (9). Integrating the second form of the new equation twice and changing the order of integration we obtain a *Volterra equation*:

$$z(x) = 1 - \int_a^x K(x, t)F(t, \lambda)z(t)dt, \tag{11}$$

where $K(x, t) = \int_t^x Y_1^2(t)Y_1^{-2}(s)ds$. Since $|Y_1(x)|$ is increasing, we know $|Y_1(t)| \leq |Y(s)|$ and hence that $|K(x, t)| \leq b - a$. The existence of $z(x)$ can be established by successive approximations using (11). We know $F(x, \lambda) = O(\lambda^{-N})$ and we can write

$$\begin{aligned} z(x) &= 1 - \int_a^x K(x, t)F(t, \lambda)z(t)dt \\ &\leq 1 + \left|\int_a^x K(x, t)F(t, \lambda)z(t)dt\right| \end{aligned}$$

$$\leq 1 + O(\lambda^{-N})M(b - a) = 1 + O(\lambda^{-N}),$$

uniformly in x and $\arg(\lambda)$ as $\lambda \to \infty$. $z(x)$ is also differentiable because

$$z'(x) = -\int_a^x Y_1^2(t)Y_1^{-2}(x)F(t, \lambda)z(t)dt = O(\lambda^{-N})$$

and thus

$$\begin{aligned} y'_1(x) &= Y'_1(x)\left[z(x) + \frac{Y_1(x)}{Y'_1(x)}z'(x)\right] \\ &= Y'_1(x)[1 + O(\lambda^{-N})]. \end{aligned}$$

The result follows for $j = 1$. The second solution with $j = 2$ is analogous. □

Liouville's Equations We now restrict our study to second-order differential equations of the form below, known as *Liouville's equations*:

$$y'' + [\lambda^2 p(x) + r(x)]y = 0 \tag{12}$$

for λ large and positive, $x \in [a, b]$, $p(x)$ twice continuously differentiable and $r(x)$ continuous. Note that (12) has the form (5) with $k = 1$, $q_0 = p$, $q_2 = r$ and $q_n = 0$ for $n \neq 0, 2$. Recall that the Mathieu equation (3) that motivated our study of asymptotic behaviors is of this type: because of the following asymptotic expansion $C = \lambda^2 \sum_{i=0}^{\infty} t_i \lambda^{-i}$, our Mathieu equation is of type Liouville with $r(x) = C - t_0\lambda^2$ and $p(x) = t_0 - a^2 \cos^2 x$. As an aside, it is worth mentioning that, in fact, the coefficient t_0 can be interpreted as an *energy level* E of a particle ($C \sim E\lambda^2$) with a one dimensional time-independent Schrödinger equation given by (3).

To obtain asymptotic expansions, proceed as follows. Substitute $\xi = \int p(x)^{1/2} dx$, $\eta = p(x)^{1/4} y$. We get a new interval $\alpha \leq \xi \leq \beta$ and a new differential equation in the variable ξ

$$\frac{d^2\eta}{d\xi^2} + \lambda^2 \eta = \rho(\xi)\eta,$$

where $\rho(\xi) = \frac{1}{4} \cdot \frac{p''}{p^2} - \frac{5}{16} \cdot \frac{p'^2}{p^3} - \frac{r}{p}$, a continuous function of ξ. The solutions of the new equation satisfy, again, a Volterra integral equation and can be written as

$$\eta(\xi) = c_1 \cos\lambda\xi + c_2 \sin\lambda\xi + \lambda^{-1} \int_{\gamma}^{\xi} \sin\lambda(\xi - t)\rho(t)\eta(t)dt,$$

where $\alpha \leq \gamma \leq \beta$ and $c_1, c_2 \in \mathbb{R}$. The full solution can be obtained by successive approximations of the form

$$\eta(\xi, \lambda) = \sum_{0}^{\infty} \eta_n(\xi, \lambda),$$

with $\eta_0(\xi, \lambda) = c_1 \cos\lambda\xi + c_2 \sin\lambda\xi$ and $\eta_{n+1}(\xi, \lambda) = \lambda^{-1} \int_{\gamma}^{\xi} \sin\lambda(\xi - t)\rho(t)\eta_n(t, \lambda)dt$. Note that if $|\rho(\xi)| \leq A$, is bounded, then the series expression for $\eta(\xi, \lambda)$ converges uniformly on the domain of ξ for λ large enough, so that indeed it is an asymptotic expansion of η. Observe that for this procedure to hold, the function $p(x)$ is assumed to be non-zero on the interval of x. Near zeros of $p(x)$, the technique breaks down and the asymptotic behavior of the solutions differs significantly in such situations. We would like, however, to generalize the method and study the asymptotics even when $p(x)$ admits zeros on the domain of x for (12).

Definition. A zero of $p(x)$ is called a *transition point* of (12).

Transition Points Assume then that $p(x)$ has a simple zero (for simplicity, to start with) at $x = c$ and no other zero in $[a, b]$. Suppose that $p'(c) > 0$ so that $p(x)$ is negative on $[a, c)$. From our previous discussion, we know that in an interval $x \in [c + \epsilon, b]$ for some $\epsilon > 0$ where $p(x) > 0$, the solution of (12) are asymptotically given by

$$\begin{aligned} &c_1[p(x)]^{-1/4} \cos\{\lambda \int [p(x)]^{1/2} dx\} + \\ &c_2[p(x)]^{-1/4} \sin\{\lambda \int [p(x)]^{1/2} dx\}, \end{aligned} \tag{13}$$

and, in $[a, c - \epsilon]$ where $p(x) < 0$ the solutions are computed in a similar way as

$$\begin{aligned} &c_3[-p(x)]^{-1/4} \exp\{\lambda \int [-p(x)]^{1/2} dx\} + \\ &c_4[-p(x)]^{-1/4} \exp\{-\lambda \int [-p(x)]^{1/2} dx\}. \end{aligned} \tag{14}$$

Recall that for these solutions to hold, $p(x)$ cannot have any zero in $[a, b]$. Observe also that the asymptotic behavior changes from one side of the transition point at c to the other: to the left of $x = c$, where $p(x) < 0$, (14) is monotonic while to the right, where $p(x) > 0$, (13) is oscillatory. As detailed in [5], there are two fundamental problems to deal with when $p(x)$ vanishes on $[a, b]$:

1. finding the connection between the constants c_1, c_2 from the expansion to the right of $x = c$ and constants c_3, c_4 from the expression to the left; combining them is necessary to describe the solution on $[a, b]$;
2. determining the asymptotic behavior in a neighborhood of c: $[c - \epsilon, c + \epsilon]$.

There are various approaches to obtain the desired *connection formulas* relating the coeffi-

[1] Named after Wentzel, Kramers and Brillouin who developed these methods in the 1920's. Jeffreys also independently established these techniques for approximating solutions to linear second order differential equations and so WKB is often replaced with WKBJ.

cients c_1, c_2 and c_3, c_4, and the general methods are known under the name of the WKB method[1]:

1. One way to relate the two sets of constants from both expansions c_1, c_2 and c_3, c_4 is to approximate $p(x)$ by $(x-c)p'(c)$ near $x = c$ and obtain an asymptotic form in terms of Bessel functions of order $\pm 1/3$, and then to compare with the expressions to the left and right of $x = c$.
2. Another way is to use complex analysis instead and integrate the differential equation along a contour in $\mathbb{C}$ consisting of the real intervals $(a, c-\epsilon)$, $(c+\epsilon, b)$ and a semi-circle through the point (c, ϵ) avoiding $x = c$ altogether.

Exercise. It is left as an exercise to the reader to apply the above results to the case of the Mathieu functions obtained in the first section, (3) and (4), in order to obtain approximations valid outside the transition region only (because $p(x)$ has zeros in the domain of x!).

Finally to obtain asymptotic solutions valid in the transition region, the idea is to transform our equation (12), by a change of variables, into an equation which is close to

$$\frac{d^2y}{dx^2} + \lambda^2 xy = 0, \tag{15}$$

whose solutions are well understood and exhibit a transition point at $x = 0$. Expansions in the transition region of solutions of this simpler equation will in turn yield expansions for solutions of (12) and the latter will involve the *Airy functions* $Ai(x)$ and $Bi(x)$. The analysis can also be extended to zeros of $p(x)$ of higher order. We do not pursue this direction any further. A useful method of estimating such functions is the *method of steepest descent*. For a more detailed discussion of these notions and procedures, we refer the reader to [2] and [5].

References

[1] Jost, J. "Partial Differential Equations." *Springer-Verlag*, New York, 2002, 344 pages.

[2] Miller, P. D. "Applied Asymptotic Analysis." *American Mathematical Society*, 2006, 467 pages.

[3] Ince, E. L. "ODE's." *Dover Publications*, 1956, 558 pages.

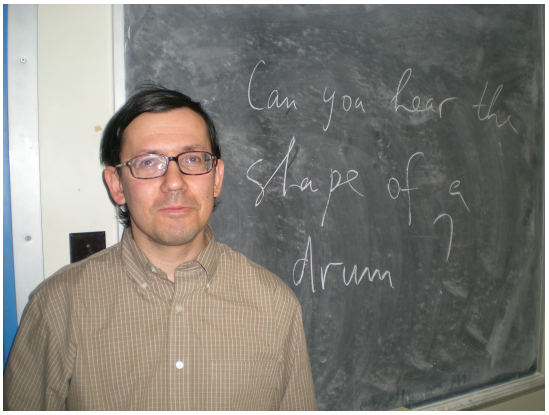
[4] McLachlan, N. W. "Theory and Applications of Mathieu Functions." *Dover Publications*, 1964, 401 pages.

[5] Erdélyi, A. "Asymptotic Expansions." *Dover Publications*, New York, 1956, 108 pages.

Interview with Professor Dmitry Jakobson

Phil Sosoe

The Delta Epsilon ($\delta\varepsilon$): *First, I am going to ask you what your current research is about. What kind of mathematician would you describe yourself as?*



Prof. Dmitry Jakobson

Prof. Jakobson: Well, I am an analyst. My main interest is spectral theory, which concerns things like vibrations of a drum, vibrations of a string. In math, those are eigenfunctions of the Laplacian: in $\mathbb{R}^n$, it's the sum of the second derivatives. Examples are sines and cosines, or spherical harmonics if we look at the sphere. In the disk, it would be Bessel functions. In many cases, it is difficult to write things down precisely, but they are interesting objects which people use to study heat and wave equations, and they occur in applications.

I am also interested in geometry, how these things relate to geometry. [Eigenfunctions of the Laplacian] also come up in mathematical physics. There are also discrete versions of these eigenfunctions, when we consider graphs. In this case we just consider the nearest-neighbour discretization of the Laplacian. That's another example of something I am interested in.

$\delta\varepsilon$: *What about your earlier research? I know that you started your career as an analytic number theorist, working under Sarnak at Princeton...*

Prof. Jakobson: Yes. I started off studying these eigenfunctions in the hyperbolic plane, which is geometry in negative curvature, "Lobachevsky geometry". The kind of results that I was proving you could try to prove on any manifold, and on any surface, but on the surfaces on which I was working, you could prove a little bit more, because there was more structure on these surfaces, called arithmetic hyperbolic surfaces. The structure essentially came from a big group of symmetries these surfaces have. There are many symmetries acting on spaces of functions, which people study in number theory and these are called Hecke symmetries. If you take a function which is invariant or changes in a nice way under all these symmetries then this function somehow has much more structure than just any arbitrary function you could come up with. The subject is called arithmetic quantum chaos. The keyword here is arithmetic. That was one half of my thesis. The other half was on Fourier series on the square torus in high dimension and there, I also used some algebraic number theory, but used it to reduce the dimension by two, essentially. You can imagine that things in dimension 2 are a little bit easier than things in dimension four, say.

As an undergrad I studied symmetry group-invariant solutions of some differential equations. Examples of model problems include dragging a chain on a rough plane. This was modeled by some system of differential equations. I would look at the symmetry group and use the symmetry group to construct group-invariant solutions, so it was about Lie groups and Lie algebras. That was different stuff.

$\delta\varepsilon$: *You mention your undergraduate work. When did you decide to go into mathematics, what drew you to mathematics?*

Prof. Jakobson: In grade 6, I suppose, I went to a competition in Moscow, in Russia. I did reasonably well, and loved it. I think it was called "tournament of cities"; there exists a version in Canada as well, in Toronto. There are enough Russians teaching math in other countries to export this type of thing. It was certainly lower-level than Putnam, but it's still a type of math contest. They mentioned that there was a school where they teach math, a sort of specialized math school, and I eventually attended. That's

where I started seriously learning math as well. I am trying to re-create something of that nature. I am organizing lectures for CEGEP students, so we will see how that goes. At the moment we've had three lectures and I want to keep it going.

This school had many graduates. Every year, maybe 30 to 60 people interested in math would graduate. Of course, not all of them would continue to do math, but many of them would. So there were several schools like that in Moscow, in St Petersburg, and in large cities in Russia, and there is a bit of a community there. People continued later on to university. It gave rise to a social network. It was nice because you would interact with people who are also interested in math and that was a good motivation.

$\delta\varepsilon$: *Tell me about your later education. At some point I believe you moved to the United States...*

Prof. Jakobson: Yes. After my freshman year, I moved to the States. I was at the University of Maryland, College Park, and the last two years I finished at MIT. I attended graduate school at Princeton. Post-doc at Caltech and the IAS, Princeton. And then one year at Chicago and then I came here. You know, 2 oceans and Chicago on this continent, and then I moved to the St Laurent, which is not as big as an ocean, but a large body of water nevertheless. I like to live in a large city, I suppose. I prefer it to a small town, but that's very personal. It depends on what various people like. Nothing to do with math.

$\delta\varepsilon$: *My last question is about mathematics in Russia, and especially mathematical education. Are there any significant differences between the way it is done here and in Russia?*

Prof. Jakobson: I would say that people in Russia used to start learning advanced things a little bit earlier than they do here. I also think that in Russia, a lot of very strong people went into math because it was good option. Many good options here like finance, law, or medicine were not as attractive in Russia at the time, when I was a student, as they are in the West, or as they are in Russia now. My parents are also math graduates, so for me it was following what my parents did. It was the path of least resistance: it's in the family.

There are a certain number of strong people who would do well in many different kinds of science. Then the question is, do they want to do math or do they want to do something a little bit different. Maybe they prefer economics or they prefer physics, or electrical engineering. I think in Russia at the time math was kind of a good option, because the technology was not so advanced, and in math you don't need so much technology. It doesn't depend so much on the equipment available.

There were lots of research institutes of some kind or other which existed in Russia at the time. After graduating from university, a mathematician would be employed, for example, by the Institute of Beekeeping or Medical Equipment, or similar things, and would do algebraic geometry on a very high level. He would be one of the top ten algebraic geometers, "studying beekeeping". I don't think the beekeeping industry in Russia profited so much from this, but it was a great place to be employed at. Now, I think the country just cannot afford as many of these places.

Lots of people who would sort of stay back. People moved a lot less than they do now, and than they do in the West, so there are sort of community relations. People would go back to their old school to teach and to give lectures. Some of them went back as teachers; good people would go back as teachers.

In contrast, Montreal, is a nice place to live, and many people like to stay. Unfortunately, in academia, most of the time you go elsewhere to do your PhD, and then you would go all over the place to do a post-doc. Whether you end up in your old city or not, depends on the job market and what openings there are. People end up in very different places and it takes a little bit of time before they can start developing new connectioins and start teaching themselves.

Some of the early math education goes back to the 1920s and 30s, when they were trying to make things very democratic and so on. A lot of math competitions. A lot of it sort of continues in this tradition.

In the long run, if what you want is to continue doing math research, and finish a PhD and so on, it doesn't matter so much whether you learn things during your junior year in col-

lege or during your junior year in high school. Of course, it's nice, and it gives you a lot more self-confidence if you did it junior year of high school than if you did it in junior year of college, but after 5 years, when you learned a particular thing, it doesn't matter so much. What matters is how you learned it. Can you go on learn new things on your own? Are you able to use the stuff you have learned? How well are you able to use it?

A lot of people who start very early become very self-confident and they sort of taper off, and they don't work much. I have seen examples like that, a lot. They are not stimulated because they know all the freshman and sophomore material already but then they don't work. It becomes like the last year of high school in the US: people just party and wait until they go to college. Then it really depends on how disciplined someone is. There could be sort of a flip side, that people get over-confident and don't work. It's good to start early, but on the other hand, everyone has their own pace. Some people are extremely quick and just catch things like that. Some people are quite slow, but they think deeply. It is very difficult to see. A lot of it depends on luck. You end up at some university, and you talk to someone, you talk to some advisor who is working on some problem. Whether this is the right problem for you, whether this problem is interesting, whether it's doable, how good the advisor is, how good the matching is. Eventually, by the law of large numbers, you will hit the lucky problem, but it may take time. Don't be discouraged that the problem seems boring, and not so interesting.

$\delta\varepsilon$: *Well, thank you very much.*

Prof. Jakobson: You're welcome.

A Theorem in the Probabilistic Theory of Numbers

Maksym Radziwill

Let φ denote the Euler-phi function. In the 20's Schoenberg proved that $\varphi(n)/n$ posses a 'limiting distribution'. This means that given a $0 \leqslant t \leqslant 1$ the proportion of $n \leqslant N$ for which $\varphi(n)/n \leqslant t$ tends to a finite limit as $N \to \infty$. The theorem is of course intuitively appealing, the limiting function being the 'probability' that '$\varphi(n)/n \leqslant t$'. In this note we prove this theorem (modulo a reference to a theorem of Erdős) using only basic analysis and some elementary number theory.

Let $\varphi(n)$ denote the number of integers $1 \leqslant k \leqslant n$ coprime to n. In this note we want to investigate the average behaviour of $\varphi(n)$. For instance is $\varphi(n)$ usually about n (maybe within a constant multiple) ? If yes, given $0 \leqslant \alpha < \beta \leqslant 1$ how often does $\alpha \leqslant \varphi(n)/n \leqslant \beta$ hold ? To answer this question consider the quantity,

$$Q_x(\alpha,\beta) = \frac{1}{x} \cdot \# \left\{ n \leqslant x : \alpha \leqslant \frac{\varphi(n)}{n} \leqslant \beta \right\}.$$

We will prove the following theorem.

Theorem 1. *There is a function $V(x)$ such that for any fixed $0 \leqslant \alpha \leqslant \beta \leqslant 1$,*

$$\lim_{x\to\infty} Q_x(\alpha,\beta) \quad = \quad V(\beta) - V(\alpha). \qquad (1)$$

A few properties of V are easy consequences of (1). For instance, for any $0 \leqslant \alpha \leqslant \beta \leqslant 1$ the left hand side of (1) is positive hence $V(\beta) - V(\alpha) \geqslant 0$ and it follows that V is increasing. Another simple property is that $V(1) - V(0) = 1$ because for all integers n we have $0 \leqslant \varphi(n)/n \leqslant 1$. Less trivially, V is a continuous function. This is the content of Theorem 2.

Theorem 2. *$V(x)$ is continuous.*

An amusing consequence of Theorems 1 and 2 is that $\{\varphi(n)/n : n \in N\}$ is dense in $[0,1]$, hence $\{n/\varphi(n) : n \in N\}$ is dense in $[1,\infty)$. Imagine proving this directly ! The result (Theorems 1 and 2) belongs to a branch of number theory called Probabilistic number theory and is due (originally) to Schoenberg (1928).

Proof of Theorem 1

In order to show that the limit in (1) exists we will start by showing that for each $k \in N$ there is a C_k such that,

$$\frac{1}{x} \cdot \sum_{n\leqslant x} \left(\frac{\varphi(n)}{n} \right)^k \quad \longrightarrow \quad C_k.$$

Therefore, 'by additivity', for each polynomial $f \in R[x]$ there is a constant $C(f)$ such that $(1/x)\sum_{n\leqslant x} f(\varphi(n)/n)) \to C(f)$. Then, using Weierstrass's theorem we approximate $I(x;\alpha,\beta)$ - the indicator function of the interval $[\alpha,\beta]$ - by polynomials, and obtain the existence of the limit

$$\lim_{x\to\infty} \frac{1}{x} \cdot \sum_{n\leqslant x} I\left(\frac{\varphi(n)}{n}; \alpha, \beta \right).$$

Thus the limit (1) exists. Then, taking

$$V(\alpha) := \lim_{x\to\infty} Q_x(0,\alpha)$$

yields the assertion of the Theorem. As announced, we start with the following lemma.

Lemma 1. *For each $k \in N$ there is a C_k such that,*

$$\lim_{x\to\infty} \frac{1}{x} \cdot \sum_{n\leqslant x} \left(\frac{\varphi(n)}{n} \right)^k \quad = \quad C_k.$$

Proof. An elementary property of φ is that φ is a multiplicative function, that is $\varphi(mn) = \varphi(m)\varphi(n)$ for $(m,n) = 1$ coprime. Further $\varphi(p^a) = p^{a-1}\cdot(p-1)$ for p prime and $a \in N$ (note that the values taken by a multiplicative function on the prime powers determine it uniquely). We can write,

$$\left(\frac{\varphi(n)}{n} \right)^k \quad = \quad \sum_{d|n} h(d) \qquad (2)$$

with h a multiplicative function given by $h(p) = (1-1/p)^k - 1$ and $h(p^\ell) = 0$ for prime p and $\ell \geqslant 2$ (to see this : check that the sum on the

right of (2) is a multiplicative function, then it's enough to check that equality in (2) holds on prime powers; which is easy). Therefore, interchanging summation,

$$\begin{aligned}
&\sum_{n\leqslant x}\left(\frac{\varphi(n)}{n}\right)^k \\
&= \sum_{n\leqslant x}\sum_{d|n} h(d) \\
&= \sum_{d\leqslant x} h(d) \sum_{\substack{n\leqslant x \\ d|n}} 1 \\
&= \sum_{d\leqslant x} h(d)\left\lfloor \frac{x}{d}\right\rfloor \\
&= x\sum_{d\leqslant x}\frac{h(d)}{d} + \sum_{d\leqslant x} h(d)\left\{\frac{x}{d}\right\}. \qquad (3)
\end{aligned}$$

Let us assume for now (and we'll prove later) that $|h(n)| \leqslant C(k)\cdot n^{-\beta}$ where $C(k)$ and $\beta = \beta(k) > 0$ are two (positive !) constants, depending only on k. If that is true then the series in (3) is absolutely convergent, and

$$|\sum_{d\leqslant x} h(d)\cdot\left\{\frac{x}{d}\right\}| \leqslant \sum_{d\leqslant x}|h(d)| \leqslant C(k)\sum_{d\leqslant x} d^{-\beta}$$

is bounded by $B\cdot x^{1-\beta}$ with some B depending only on k. Dividing both sides of (3) by x and taking $x\to\infty$ we get,

$$\lim_{x\to\infty}\frac{1}{x}\cdot\sum_{n\leqslant x}\left(\frac{\varphi(n)}{n}\right)^k = \sum_{d=1}^{\infty}\frac{h(d)}{d}.$$

Of course the function h, hence the limit depends on k. To complete the proof of the lemma it remains to prove that $|h(n)| \leqslant C(k)\cdot n^{-\beta}$ holds for all $n \geqslant 1$ with some $C(k)$ and $\beta = \beta(k)$ depending only on k. For primes $p \leqslant k$ we have $|h(p)| \leqslant 1$ (in fact this is true for all primes p), while for primes $p > k$ we have,

$$|h(p)| = 1-(1-1/p)^k \leqslant k/p \leqslant p^{-\beta}$$

where $\beta = \beta(k) = 1/(k^2+1)$. The first inequality follows from $(1-1/p)^k \geqslant 1-k/p$ while the second from $k^{k^2+1} \leqslant k^{k^2} + k^2\cdot k^{k^2-1} \leqslant (k+1)^{k^2} \leqslant p^{k^2}$. Therefore, since $|h(n)|$ is multiplicative, we obtain, for squarefree n,

$$\begin{aligned}
|h(n)| &\leqslant \prod_{\substack{p|n \\ p>k}} p^{-\beta} \leqslant \prod_{p\leqslant k} p^{\beta}\cdot\prod_{p|n} p^{-\beta} \\
&= C(k)\cdot n^{-\beta}
\end{aligned}$$

where $C(k) = \prod_{p\leqslant k} p^{\beta}$ is a constant depending only on k. When n is not squarefree we have $h(n) = 0$. Hence $|h(n)| \leqslant C(k)\cdot n^{-\beta}$ holds for all $n \geqslant 1$, as desired. □

An immediate corollary of Lemma 1 is that the limit $(1/x)\cdot\sum f(\varphi(n)/n)$ exists for polynomials $f(x) \in R[x]$.

Corollary 1. *For each polynomial $f(x) \in R[x]$ there is a $C(f)$ such that*

$$\lim_{x\to\infty}\frac{1}{x}\cdot\sum_{n\leqslant x} f\left(\frac{\varphi(n)}{n}\right) = C(f).$$

We will use the next lemma to approximate the indicator function $I(x;\alpha,\beta)$ of the interval $[\alpha,\beta]$ by polynomials.

Lemma 2. *Let $0 \leqslant \alpha \leqslant \beta \leqslant 1$ be given. For any $1 \geqslant \varepsilon > 0$ there is a polynomial $P_\varepsilon(x) \in R[x]$ such that*

$$|P_\varepsilon(x) - I(x;\alpha,\beta)| \leqslant \varepsilon + E(x;\alpha,\beta,\varepsilon)$$

for all $x \in [0,1]$. Here, $E(x;\alpha,\beta,\varepsilon)$ is the sum of two indicator functions

$$I(x;\alpha-\varepsilon,\alpha+\varepsilon) + I(x;\beta-\varepsilon,\beta+\varepsilon).$$

Furthermore, $|P_\varepsilon(x)| \leqslant 4$ for all $x \in [0,1]$.

Proof. Let $f_\varepsilon(x)$ be a continuous function defined as follows :

$$f_\varepsilon(x) = \begin{cases} 0 & , x\in[0,\alpha-\varepsilon]\cup[\beta+\varepsilon,1] \\ \text{linear} & , x\in[\alpha-\varepsilon,\alpha]\cup[\beta,\beta+\varepsilon] \\ 1 & , x\in[\alpha,\beta] \end{cases}$$

where by 'linear' it is meant that f_ε is a linear function on that interval (chosen so that the continuity of f_ε is preserved). By the Weierstrass theorem given $\varepsilon > 0$, there is a $P_\varepsilon(x)$ such that $|f_\varepsilon(x) - P_\varepsilon(x)| \leqslant \varepsilon$. By construction, we have $|f_\varepsilon(x) - I(x;\alpha,\beta)| \leqslant E(x;\alpha,\beta,\varepsilon)$. Therefore, the result follows by the triangle inequality:

$$|P_\varepsilon(x) - I(x;\alpha,\beta)| \leqslant \varepsilon + E(x;\alpha,\beta,\varepsilon).$$

Furthermore,

$$|P_\varepsilon(x)| \leqslant I(x;\alpha,\beta) + \varepsilon + E,$$

which is less than $1+1+2 = 4$. □

Lemma 2 is saying that $I(x;\alpha,\beta)$ can be approximated uniformly by polynomials except in a small neighborhood of the points $\{\alpha,\beta\}$. This is as it should be since $I(x;\alpha,\beta)$ has discontinuities at $x=\alpha$ and $x=\beta$. To handle the term $E(x;\alpha,\beta,\varepsilon)$ we will need one last technical lemma.

Lemma 3. *There is an absolute constant D such that for $\varepsilon>0$ and $\alpha,\beta>0$,*

$$\frac{1}{x}\cdot\sum_{n\leqslant x}E\left(\frac{\varphi(n)}{n};\alpha,\beta,\varepsilon\right)\quad\leqslant\quad\frac{D}{\log(1/\varepsilon)}.$$

Proof. This lemma is saying that on average $\varphi(n)/n$ rarely concentrates in small intervals. This is essentially 'continuity'. For our aims, we don't need the full force of Lemma 3 (in fact any term decaying to 0 as $\varepsilon\to 0$ on the right hand side would do). For a proof of this lemma see www.math−inst.hu/∼p_erdos/1974-19.pdf. Of course, we are 'cheating' since the theorem quoted is more involved (in terms of thinking) that what we aim at proving. Nonetheless, my aim was to not assume knowledge of probability in this note and the proof I referred to does not make use of it. □

We are now ready to prove Theorems 1 and 2.

Proof. Given $k\in N$, by Lemma 2 there is a polynomial $P_k(y)$ such that

$$|P_k(y)-I(y;\alpha,\beta)|\leqslant\frac{1}{k}+E\left(y;\alpha,\beta,\frac{1}{k}\right)$$

for all $y\in[0,1]$. Since

$$Q_x(\alpha,\beta)=\frac{1}{x}\sum_{n\leqslant x}I\left(\frac{\varphi(n)}{n};\alpha,\beta\right)$$

by the triangle inequality,

$$\begin{aligned}&\left|\frac{1}{x}\sum_{n\leqslant x}P_k\left(\frac{\varphi(n)}{n}\right)-Q_x(\alpha,\beta)\right|\\ \leqslant\;&\frac{1}{x}\sum_{n\leqslant x}\left|P_k\left(\frac{\varphi(n)}{n}\right)-I\left(\frac{\varphi(n)}{n};\alpha,\beta\right)\right|\\ \leqslant\;&\frac{1}{x}\sum_{n\leqslant x}\left(\frac{1}{k}+E\left(\frac{\varphi(n)}{n};\alpha,\beta,\varepsilon\right)\right)\\ \leqslant\;&\frac{1}{k}+\frac{D}{\log k}\leqslant\frac{D+1}{\log k}\end{aligned}$$

by Lemma 3. Therefore,

$$\begin{aligned}&C(P_k)-\frac{D+1}{\log k}\\ =\;&\liminf_{x\to\infty}\left(\frac{1}{x}\sum_{n\leqslant x}P_k\left(\frac{\varphi(n)}{n}\right)-\frac{D+1}{\log k}\right)\\ \leqslant\;&\liminf_{x\to\infty}Q_x(\alpha,\beta)\leqslant\limsup_{x\to\infty}Q_x(\alpha,\beta)\\ \leqslant\;&\limsup_{x\to\infty}\left(\frac{1}{x}\sum_{n\leqslant x}P_k\left(\frac{\varphi(n)}{n}\right)+\frac{D+1}{k}\right)\\ =\;&C(P_k)+\frac{D+1}{\log k}.\end{aligned}\tag{4}$$

By Lemma 2 we have $|P_k(\varphi(n)/n)|\leqslant 4$ and hence $|C(P_k)|\leqslant 4$. Thus, by Bolzano-Weierstrass there is a subsequence n_k such that $C(P_{n_k})\longrightarrow\ell$ for some ℓ. Let $k\to\infty$ in (4) through the subsequence n_k. We get,

$$\ell\leqslant\liminf_{x\to\infty}Q_x(\alpha,\beta)\leqslant\limsup_{x\to\infty}Q_x(\alpha,\beta)\leqslant\ell$$

hence the limit in (1) exists and the function $V(x)$ is given by

$$V(\beta)\quad=\quad\lim_{x\to\infty}Q_x(0,\beta).$$

□

Proof of Theorem 2

As it turns out Theorem 2 is an easy consequence of Lemma 3 and Theorem 1.

Proof. Since

$$I\left(x;\alpha,\alpha+\varepsilon\right)\leqslant E\left(x;\alpha,\beta,\varepsilon\right),$$

we obtain by Lemma 3 that

$$0\leqslant\frac{1}{x}\sum_{n\leqslant x}I\left(\frac{\varphi(n)}{n};\alpha,\alpha+\varepsilon\right)\leqslant\frac{D}{\log(1/\varepsilon)}$$

or in a different notation,

$$0\leqslant Q_x(\alpha,\alpha+\varepsilon)\leqslant\frac{D}{\log(1/\varepsilon)}.\tag{5}$$

Let $x\to\infty$ and use Theorem 1 to conclude $0\leqslant V(\alpha+\varepsilon)-V(\alpha)\leqslant D/\left(\log(1/\varepsilon)\right)$. Therefore $V(\alpha+\varepsilon)\longrightarrow V(\alpha)$ when $\varepsilon\to 0^+$. Thus V is right continuous. To prove left continuity replace α by $\alpha-\varepsilon$ in (5) and take the limit $x\to\infty$. This gives us that $0\leqslant V(\alpha)-V(\alpha-\varepsilon)\leqslant D/\left(\log(1/\varepsilon)\right)$. Hence V is left continuous. □

Conclusion

In the proof given above there was no reference to probability theory. However, the interaction with probability is quite strong, and in fact once that Lemma 1 is known the conclusion of Theorem 1 is immediate by a theorem in probability theory (the 'method of moments') that was not used here. In fact using probability theory one can prove that,

$$V(t) = P\,\mathrm{rob}\left(\prod_{p\mathrm{prime}}\left(1-\frac{1}{p}\right)^{Z_p} \leqslant t\right), \quad (6)$$

where Z_p are independent random variables distributed according to

$$P(Z_p = 1) = \frac{1}{p} \text{ and } P(Z_p = 0) = 1 - \frac{1}{p}.$$

There is a heuristic reason to expect (6), which I am going to explain now. Since $\varphi(n)/n$ is a multiplicative function with $\varphi(p^\ell)/p^\ell = 1-1/p$, we can write

$$\frac{\varphi(n)}{n} = \prod_{\substack{p|n \\ p\mathrm{prime}}}\left(1-\frac{1}{p}\right). \quad (7)$$

Given a 'random' integer n, the probability that $p|n$ is intuitively $1/p$, while the probability that $p \nmid n$ is $1-1/p$. (If this is not clear : what is the probability that a random integer is even ? Intuitively it is 1/2.) Furthermore, for two primes $p \neq q$ the event $p|n$ and $q|n$ can be seen as independent (none has any influence on the other; however, for composite numbers this is no longer true : if $6|n$ then $3|n$ necessarily). Therefore for a 'random' integer n, the probability that in (7) a $1-1/p$ appears in the product is $1/p$, while the probability that the term $1-1/p$ does not appear is $1-1/p$. Hence we expect the product in (7) (that is $\varphi(n)/n$) to behave as the random variable in (6) (in (6) the Z_p essentially stands for 'does p divide a random n ?'). Indeed, when $Z_p = 1$ a $1-1/p$ appears in (6) and the probability of $Z_p = 1$ is $1/p$, in agreement with our intuition about the likelihood of the event '$p|n$' and its 'action' on $\varphi(n)/n$.

The heuristic explained previously is a powerful idea, and is due to Mark Kac. To implement the idea in practice one has to compare two distinct measures. There are general theorems that are doing just that, for instance the so-called 'Kubilius model'. For more information on the interaction between number theory and probability theory a good starting point are the notes from a talk by Jean-Marc Deshouillers to be found in algo.inria.fr/seminars/sem96-97/deshouillers.pdf. A (more involved) survey by Kubilius can be found at www.numdam.org/numdam-bin/fitem?id=SDPP_1969-1970__11_2_A9_0.
The standard textbooks are due to Tenenbaum, *Introduction to analytic and probabilistic number theory* and to Elliott, *Probabilistic number theory Vol I, II*. The former is more accessible and is available in the SUMS library.

Is Implied Volatility Incremental to Model-Based Volatility Forecasts?

Tigran Atoyan

Improving forecast of future volatility can greatly improve the accuracy of the option pricing models based on the Black-Scholes original model. Furthermore, the exponential increase of computing power in the last decades has unlocked a whole range of tools useful to forecasting. The main goal of our study is to mimic the work done by Becker(2007), i.e. to check if implied volatility contains any incremental information to that obtained from historical models such as GARCH, SV, and ARMA. This could help establish the link between mathematical volatility models used for producing forecasts and the intuitive forecast made by the market.

Introduction

Volatility in Financial Markets

What exactly is volatility? We will first start to define what we mean by price volatility of financial assets. Let us define $P(t)$ to be the spot price of an asset. We can then define the return to be:

$$R(t) \;=\; \log P(t) - \log P(0), \quad t > 0 \qquad (1)$$

In financial theory, $R(t)$ can be represented by the following stochastic continuous time process:

$$\mathrm{d}R(t) \;=\; \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t), \quad t > 0 \qquad (2)$$

We call $\mu(t)$ the drift process, $\sigma(t)$ the spot volatility, and $W(t)$ is the standard Brownian motion process. We can often omit the drift process $\mu(t)$ from equation (2). Finally, we define the actual (or daily) volatility for the n^{th} day by:

$$\sigma_n^2 \;=\; \int_{n-1}^{n} \sigma^2(s)\mathrm{d}s \qquad (3)$$

Now that we have defined what we mean by spot volatility and actual volatility, we can go on to discussing estimates of the volatilities.

The simplest unbiased estimate of daily volatility is squared daily returns. Indeed, if we set $\mu(t) = 0$, we see that by taking the integral of equation (2) from $n-1$ to n along $\mathrm{d}t$, squaring the result, and finally taking expected values, we get that the expected value of $(\log P(n) - \log P(n-1))^2$, i.e. daily squared returns, equals σ_n^2. However, daily squared returns are not the most efficient estimators available. It has been shown (see Poon and Granger, 2003) that summed intradaily squared returns, called realized volatility (RV), is another unbiased estimate of daily volatility which is more efficient than daily squared returns.[1]

Realized Volatility

One of the papers which thoroughly covers the definition and properties of realized volatility is Andersen et al. (2001). According to Andersen, we can define daily realized volatility as:

$$RV_n \;=\; \sum_{j=1}^{m} (r_m(m\cdot(n-1)+j))^2 \qquad (4)$$

where

$$r_m(t) \;=\; \log P_m(t) - \log P_m(t-1), \quad t \geq 0 \qquad (5)$$

Here, we assume the series P(t) is the set of intradaily asset prices with m data points per day.

However, we should always keep in mind that theory does not always perfectly describe the real behavior of financial markets. According to theory, realized volatility converges to the underlying volatility as $m \to \infty$. However, if we go beyond a certain frequency for intradaily data sampling, financial microstructures (e.g. uneven time spacing of tick-by-tick prices) can

[1] However, in practice, daily RV are not completely unbiased due to the non-zero correlation of the return series, as will be briefly discussed later on.

affect results by inducing negative autocorrelation in the interpolated return series (Andersen et al. 2001). Thus, we must find a frequency which is large enough so that the daily realized volatility measurements are largely free from measurement errors, but small enough so that market microstructures don't significantly affect results. It has been empirically found that 5-minute intradaily data is a good choice for computing realized volatility.

Implied Volatility

As mentioned above, the Black-Scholes model (and its variants) use the estimated forecast of the future volatility of the underlying asset returns as part of its input to compute the current price of an option. Furthermore, if we denote the price function for the option as P and the the estimated forecast of the future volatility as σ_f, then P is a strictly increasing function of σ_f. This means that given a price $\overline{P}$, we can find the corresponding estimate $\overline{\sigma}_f$ by using the inverse function of $P(\sigma_f, .)$. This is called implied volatility (IV). Thus, implied volatility is the measure of the market's best estimate of future volatility.

There are however some inconsistencies in implied volatility estimates. For example, we should theoretically get the same estimate of σ_f for each asset. It has however been noted (see Poon and Granger, 2003) that options with the same time to maturity but with different strike prices yield different IV estimated volatility forecasts for the same underlying asset. This is often called volatility smile.[2].

Data and VIX

We use the S&P 500 Composite Index, presented at 5-minute intervals, from February 9^{th} 2006 to June 6^{th} 2008, as the base of our study. After the Dow Jones, the S&P 500 is the second most watched index of US large-cap stocks and is also considered to be a bellwether of the US economy, i.e. to be influential on trends and informative on the state of the economy. As for the IV index, we used the VIX index provided by the CBOE[3]. The VIX is a weighted average of the implied volatility of options of a wide range of strike prices. We have chosen to include VIX in our study since it is a popular measure of the implied volatility of the S&P 500 index.

Volatility Models

In this section, we will describe the ARFIMA, GARCH, and SV classes of models. In each case, we will give general definitions and specifications of the models, then discuss some results on their properties, and finally provide some of their pros and cons.

ARFIMA(p,d,q) Models

Many common time series models are included in the general ARFIMA(p,d,q) model. The latter may be represented as:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t \;=\; \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{6}$$

where p,q are integers, $d \geq 0$, L is the lag operator, ϕ_i are the parameters of the autoregressive part, θ_i are the parameters of the moving average part, and ε_t are error terms (usually taken to be i.i.d Normal). If d belongs to a certain set of non-integer values, it has been shown that the ARFIMA model can exhibit long-range dependence, and thus would be well suited for long-memory time series. If d is a positive integer, then the ARFIMA(p,d,q) model reduces to an ARIMA(p,d,q) model. Finally, if $d = 0$, then the ARFIMA(p,d,q) model reduces to the ARMA(p,q) model. In this study, we mainly deal with the ARMA(2,1) model.

[2]The graph of IV vs strike price is u-shaped and hence looks like a smile.
[3]Chicago Board Options Exchange

GARCH Models

In the analysis of volatility, the most commonly used autoregressive model would be the GARCH model and its many variants. Let us first define the following:

$$r_t = \mu + \varepsilon_t, \qquad \varepsilon_t = \sqrt{h_t} z_t, \qquad z_t \sim N(0,1) \tag{7}$$

Here, r_t represents the returns series. The only term which we haven't defined yet is the term h_t, corresponding to the underlying return volatility in the GARCH model. For the GARCH(1,1) process, which is the basis of many of the commonly used GARCH models, we have:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{8}$$

If the modeled time series is stationary, we must have $\alpha_1 + \beta \in (0,1)$.

Since it has been empirically found that volatility series behave differently depending on the sign of the reruns, the GARCH GJR process is an enhanced version of GARCH which takes into account the sign of the returns. We will call this a non-symmetrical process. The h_t term in GARCH GJR is defined as follows:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 s_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{9}$$

where

$$s_t = \begin{cases} 1 & \text{if } \varepsilon_t < 0 \\ 0 & \text{if } \varepsilon_t \geq 0 \end{cases}$$

Thus, the term with the α_2 coefficient is non-symmetric as desired. Note that GJR yields the GARCH(1,1) process back if we set $\alpha_2 = 0$.

The last GARCH model we will describe here is an extension of the GARCH GJR process. Since it has been shown (e.g. by Andersen, 2001) that RV is a good estimator of volatility, it may be worth incorporating RV data into the GARCH model. One of the variations of the GJR doing this is the GARCH GJR+RVG process. For the latter, h_t is defined as follows:

$$h_t = h_{1t} + h_{2t} \tag{10}$$

$$h_{1t} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 s_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{11}$$

$$h_{2t} = \gamma_1 h_{2t-1} + \gamma_2 \mathrm{RV}_{t-1}. \tag{12}$$

Note that this yields the GARCH GJR process if $\gamma_1 = \gamma_2 = 0$ and the GARCH(1,1) process if $\gamma_1 = \gamma_2 = \alpha_2 = 0$.

Even though the studies on GARCH performance have been inconclusive, there are some definite advantages and disadvantages of using GARCH models. These are :

- Pros:
  Unlike most simple historical methods, some variants of GARCH models such as the non-symmetric GJR model can separate volatility persistence from volatility shocks. This is because the GJR model reacts differently to positive and to negative returns. This is a useful property for volatility models because of the strong negative relationship between volatility and shocks.
- Cons:
  Because volatility persistence in GARCH GJR models changes relatively quickly when the sign of returns changes, GJR models underforcast volatility with a higher probability than simpler models such as EWMA, which might be problematic in some settings. It has also been empirically found that parameter estimation becomes unstable when the data period is short or when there is a change in the volatility levels.

Stochastic Volatility Models

The last class of models we will consider are the stochastic volatility models. The key characteristic of stochastic volatility models is the inclusion of a stochastic term in the volatility equation. The basic SV model can be represented in the following manner:

$$r_t = \mu + \sigma_t u_t, \qquad u_t \sim N(0,1) \tag{13}$$

where

$$\log(\sigma_t^2) = \alpha + \beta \log(\sigma_{t-1}^2) + w_t, \quad w_t \sim N(0, \sigma_w^2) \tag{14}$$

We should also note that in the estimation of the parameters, it is often easier to deal with $\eta = \frac{\alpha}{1-\beta}$. In this study, it is η and not α that

we are estimating (of course, given $\beta < 1$, there is a one-to-one correspondence between α and η).

As in the GARCH case, we wish to incorporate RV into the basic model (We will call this the SV-RV model). We will thus incorporate the RV component as an exogenous variable in the volatility equation and get:

$$\log(\sigma_t^2) = \alpha + \beta \log(\sigma_{t-1}^2) + \gamma(\log(\mathrm{RV}_{t-1}) - E_{t-1}[\log(\sigma_{t-1}^2)]) + w_t \tag{15}$$

where, as in eq. (14), $w_t \sim N(0, \sigma_w^2)$. It is worth noting that this augmented model nests the basic SV model if $\gamma = 0$.

As mentioned above, what makes the SV models innovative is the stochastic component used in the volatility equation. It is only in the mid 1990's that SV models caught the interest in the area of volatility analysis, mainly due to its high computational requirements. It has been shown to fit returns better than ARCH models and to have residuals closer to the standard normal. But the total body of studies comparing performances of SV with that of other models is yet inconclusive.

It is worth noting that SV models are harder to extend than GARCH models, at least from the technical point of view. Because the likelyhood of SV models can't be computed in closed form due to the presence of two stochastic processes, we must use methods such as Markov Chain Monte Carlo, the method of moments, quasi-maximum likelyhood methods, etc. for parameter estimation, which are harder to deal with than the straighforward maximum likelyhood estimation that can be performed on simpler models.

Thus, here are the main pros and cons of SV models:

- Pros:
  SV models fit returns of some asset classes better, have residuals closer to standard normal, have fat tails, and are closer to theoretical models in finance, especially in derivative pricing, compared to other common financial returns models.

- Cons:
  The estimation of the parameters of SV models can be somewhat involved, especially in the case of extended SV models. The computational requirements are also higher.

Selected Results

In this section, we will cover some of the results obtained thus far in our study.

Parameter Estimates

The first important model we examine is the GARCH GJR model. Here, we compute parameters by maximizing the log-likelihood function. Note that we use h_0 to be equal to the variance of the return series. This is the most natural choice, since it is essentially an "average volatility" estimate if $\mu \approx 0$. The parameters we obtained are:

μ	α_0	α_1	α_2	β	log(L)
$1.4 \cdot 10^{-4}$	$2.7 \cdot 10^{-6}$	0.041	0.281	0.81	1814

The GARCH GJR+RVG model parameters are also computed using the maximum likelihood procedure. However, assigning optimal values to h_{10} and h_{20} is not trivial. As above, we can set $h_0 = h_{10} + h_{20}$ to be equal to the variance of the return series. Then we have to decide what ratio to use for h_{10} and h_{20}. So far, we have decided on using a 1:1 ratio, meaning $h_{10} = h_{20}$. This yielded the following parameter estimates:

μ	α_0	α_1	α_2
$2.6 \cdot 10^{-4}$	$7.0 \cdot 10^{-6}$	0.13	0.29

β	γ_1	γ_2	log(L)
0.67	-0.10891	-0.16	1804

However, we can most likely improve the results considerably by assigning a better ratio of h_{10} vs h_{20}. This is something which needs to be worked on in the future.

Finally, the SV model parameters were estimated using Markov Chain Monte Carlo, denoted shortly by MCMC.[4] Using an Accept-Reject Metropolis-Hastings algorithm with the initial volatility series set equal to the square returns series, we found the following parameter estimates:

β	η	σ_w^2	log(L)
0.942	-9.9	0.101	249

With associated standard errors:

s.e.(β)	s.e.(η_0)	s.e.(σ)
0.0090	0.25	0.0094

Here we assume that $\mu = 0$.

Volatility Results and Properties

The volatility and log-volatility plots are given below (at the end of the article) for the GARCH GJR, GARCH GJR+RVG, and SV volatility series.

Next, we examine the series $(r_t - \mu)/\sigma_t$ for the GARCH GJR, GARCH GJR+RVG, and SV models. Their plots are given below (at the end of the article).

If the models are correct, the latter series should be equivalent to a $N(0,1)$ process. We computed the means and variances of the series and the means of the square of the series. Note that the latter should be approximately equal to $E(\chi^2_{(1)}) = 1$ if our assumption about normal residuals is correct.

	mean(residuals)	variance(residuals)
GJR	0.01939	0.9430
GJR+RVG	0.02001	0.9534
SV	0.06682	1.048

	mean(residuals^2)
GJR	0.9416
GJR+RVG	0.9520
SV	1.051

The above results indicate that the scaled residuals are indeed approximately $N(0,1)$ distributed.

Next, we look at the autocorrelation functions of each volatility series (see the end of the article for the figures).

All three series have correlation functions which decay slowly. It is interesting to note that the correlation for the SV series starts increasing after a lag of approximately 15 days. If time allows it, this is a result which would be worth analyzing further.

Finally, we examine the crosscorrelations of the volatility series with respect to each other, to the squared daily returns series (SDR), and to the VIX series.

	GJR	GJR+RVG	SV
GJR	1	0.9692	0.6110
GJR+RVG	0.9692	1	0.5974
SV	0.6110	0.5974	1
SDR	0.4514	0.4760	0.5274
VIX	0.6535	0.5747	0.6253

	SDR	VIX
GJR	0.4514	0.6535
GJR+RVG	0.4760	0.5747
SV	0.5274	0.6253
SDR	1	0.3992
VIX	0.3992	1

We see that the 3 historical volatility models are much more correlated to the VIX index than is the squared daily returns (SDR) series, which was expected since the former should be more accurate estimators of true volatility than the raw SDR series.

Concluding Remarks

To summarize, we have thus far done the following:

- RV COMPUTATION:
  We computed the RV series, and began observing the behavior of the RV series as the frequency of the intradaily data was changed. We also studied the effect of

[4] We can't use the regular ML methods since that the likelihood function is not known for SV models (because the unobserved volatility series has a stochastic behavior).

using data from the opening and closing hours of each trading day.[5]

- Parameter Estimates:
  We have thus far computed the parameters for the ARMA(2,1), GARCH GJR, GARCH GJR-RVG, and SV series. The methods used were maximum likelihood estimation for the cases where the likelihood could be computed (all but the SV models) and Markov Chain Monte Carlo (MCMC) estimation for the SV models.
- Volatility Series Analysis
  Using the parameters found in the step above, we found the volatility series for each model. We then analyzed these series by computing the ACF (autocorrelation function) and the cross-correlations, and also by checking for normality in the scaled residuals series. What we found was what we roughly expected based on the empirical results in the literature.

What yet has to be done for achieving the goal of this study is the following:

- RV issues:
  Resolve some inconsistencies encountered during the RV computations. These inconsistencies include the unexpected scaling issues encountered previously, low correlation with the historical volatility and VIX series, etc.
- Finish model estimation:
  Compute the parameters for the SV model which incorporates RV and find a better ratio for the initial values of the GARCH GJR+RVG model. We base the need for a better ratio on the fact that the log likelihood of the GARCH GJR model is higher than that of the GARCH GJR+RVG model, which should not be the case since the latter model has more parameters.
- Average volatility series:
  Compute an appropriate weighted average of the volatility series obtained from each of the above models. This will yield our best estimate of volatility as given by the historical data models.
- VIX vs historical volatility
  Do a regression of VIX onto the historical volatility series obtained from the above step, and see if the VIX contains any pertinent information incremental to that from the historical volatility series. The details of the methodology for doing this are given by Becker (2007).

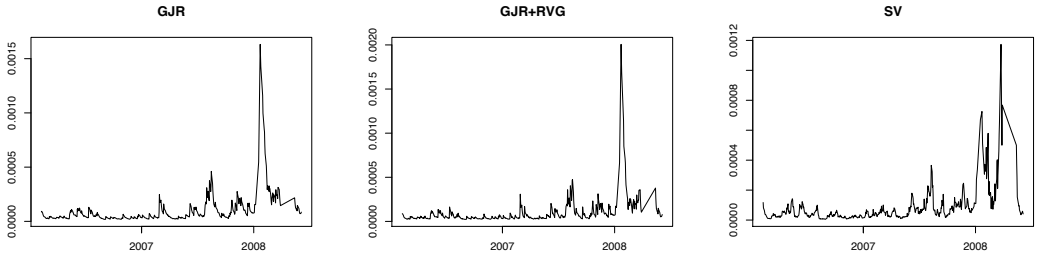
References

[1] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42-55.

[2] Becker, R., Clements, A.E., White, S.I, 2007. Does implied volatility provide any information beyond that captured in model-based volatility forecasts? Journal of Banking & Finance 31, 2535-2549.

[3] Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance 12, 445-475.

[4] Poon, S.-H., Granger, C.W.J., 2003. Forecasting volatility in financial markets: a review. Journal of Economic Literature 41, 478-539.



Volatility Series

[5] The results and graphs from the RV vs frequency and data filtering study are available upon request, but have not been attached to this report since they didn't concern the main objective of our study.

More figures:

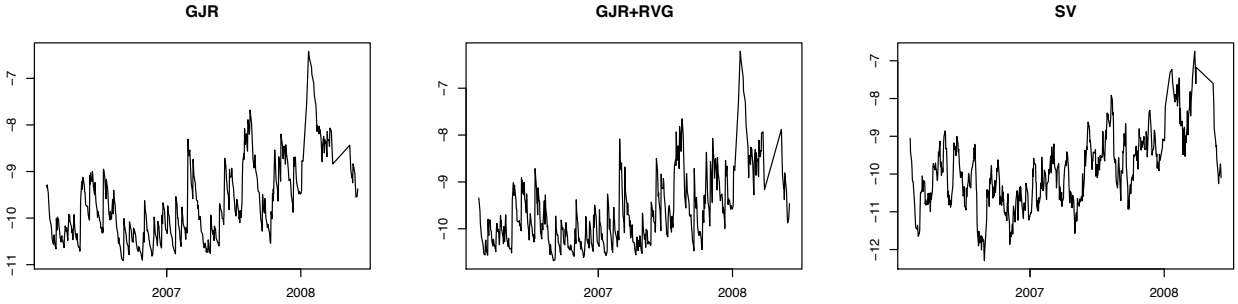


Figure 1: Log Volatility Series

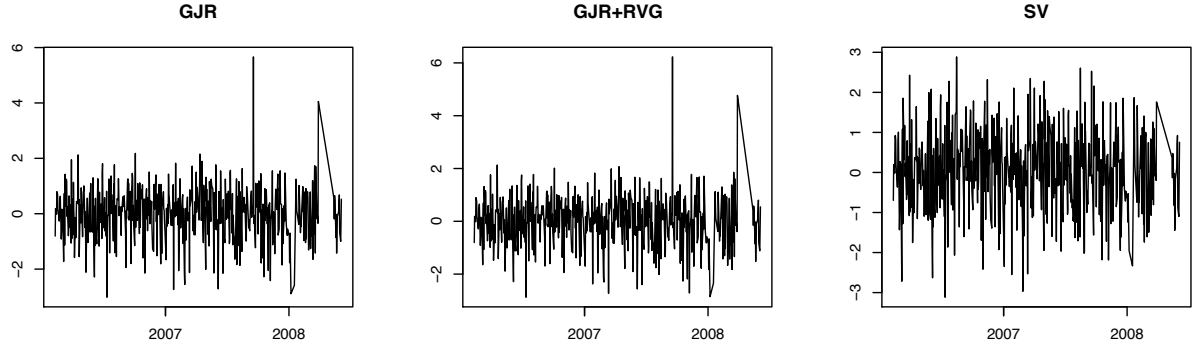


Figure 2: Scaled Residuals

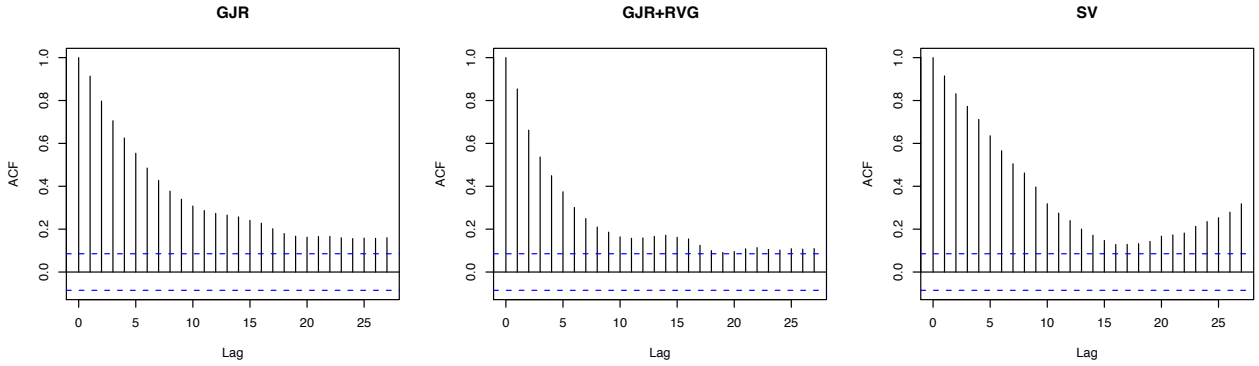


Figure 3: Autocorrelation Function

Definition of Entropy in Hamiltonian Mechanics Based on Lectures by Prof. Vojkan Jaksic

Alexandre Tomberg

We provide an elementary introduction to the definition of entropy and entropy production of open Hamiltonian systems.

Introduction

On the scales where quantum and relativistic effects are negligible, Newtonian mechanics present an acurate and a relatively simple framework for describing physical systems. However, the step from Newton's equations of motion to the concept of entropy and temperature is not trivial. The aim of this document is to provide an elementary introduction to the definition of entropy and entropy production of open systems in the context of Hamiltonian systems. Although a rigorous approach to these definitions for infinite systems usually involves advanced concepts from Measure Theory and Functional Analysis, we will try to avoid them in our discussion.

Hamiltonian System

A classical (as opposed to quantum) system is described by a phase space and a Hamilton function on that phase space. For example, a system consisting of k particles in which N independant directions of movement are defined is said to have N degrees of freedom, and the phase space is typically $M = \mathbb{R}^{kN} \oplus \mathbb{R}^{kN}$ with variables

$$x = (q_1, ..., q_k, p_1, .., p_k),$$

where

$$q_i = (q_{i1}, \dots, q_{iN})$$

is the position of the ith particle, and

$$p_i = (p_i 1, \dots, p_{iN})$$

is its momentum.

Given (q, p),$q = (q_1, ..., q_k), p = (p_1, ..., p_k)$ the energy of the system is described by the Hamilton function $H(q, p)$, $H : M \to \mathbb{R}$. For example, for $N = 1$, we can easily write Hamiltonians for a few very simple systems. A free particle $(k = 1)$ of mass m in a potential field V:

$$H(q, p) = \frac{1}{2m}p^2 + V(q)$$

Setting $V(q) = \frac{\omega q^2}{2}$, we get a harmonic oscillator of frequency ω, and so on.

Equations of motion

We assume H is C^2. Let $(q_t, p_t) = x_t$ be a phase point at time t. Then

$$\begin{cases} \dot{q_t} = (\nabla_p H)(q_t, p_t) \\ \dot{p_t} = (-\nabla_q H)(q_t, p_t) \end{cases} \tag{1}$$

is a system of differential equations with initial conditions $(q_\circ, p_\circ)$ (often simply written as (q, p)). The solutions to this system are curves in the phase space describing the motion. In Proposition we state a sufficient condition for the existence and uniquieness of solutions to (1), and we will always assume that a unique solution exists for all t.

Lemma 1. *Conservation of energy. That is for all t, $H(q_t, p_t) = H(q, p)$*

Proof. $H(q_t, p_t)$ function of t. Then using the Hamilton equations (1) we have:

$$\begin{aligned}\frac{d}{dt}\big(H(q_t, p_t)\big) =& \dot{q_t} \cdot [(\nabla_p H)(q_t, p_t)] \\ &+ \dot{p_t} \cdot [(-\nabla_q H)(q_t, p_t)] = 0.\end{aligned}$$

□

The Hamilton equations of motion have a global solution, i.e. for any initial data (q, p), a unique solution exists for all t, under the following two conditions.

$$H(q, p) \geq 0 \; \forall (q, p)$$

$$\lim_{||(q,p)||^2 \to \infty} H(q,p) = +\infty$$

The proof consists of finding a solution on a finite interval through Picard iteration method, and then extending it to all t using the Energy Conservation Lemma. However, due to its length, the proof is ommitted here.

Let x_t be the Hamilton flow, let D be a region (open set) in M, and define

$$D_t = \{x_t : x_\circ \in D\}$$

Then $Vol(D_t) = Vol(D)$.

Before we proceed to the proof of the theorem, let us state a proposition that is a generalization of Liouville theorem for a more general family of differential equations. Suppose we are given a system of ODEs

$$\dot{x} = f(x), \text{ where } x = (x_1, \ldots, x_n),$$
$$\text{and } f : \mathbb{R}^n \to \mathbb{R}^n$$

and that a global solution exists. Let ϕ_t be the corresponding flow

$$\phi_t(x) = x + f(x) \cdot t + O(t^2), \ (t \to 0). \qquad (2)$$

Let $D(0)$ be a region in $\mathbb{R}^n$, with $V(0)$ its volume. Then $V(t)$ = volume of $D(t)$, where $D(t) = \{\phi_t(x) : x \in D(0)\}$.

$$\left(\frac{d}{dt}V(t)\right)\Big|_{t=0} = \int_{D(0)} \operatorname{div} f \, dx$$
$$= \int_{D(0)} \operatorname{div} f \, dx_1 \cdots dx_n$$

Proof. Since $D(t) = \phi_t(D(0))$, the change of variables formula yields:

$$V(t) \overset{def}{=} \int_{\phi_t(D(0))} dy = \int_{D(0)} |\det \phi_t'| \, dx$$

Using (2) to calculate $|\det \phi_t'|$, we get:

$$\phi_t' = Id + f' \, t + O(t^2) \text{ as } t \to 0$$

Since $\det(Id + A\,t) = 1 + t \ \operatorname{tr}(A) + O(t^2)$ for any matrix A,

$$|\det \phi_t'| = 1 + t \ \operatorname{tr}(f') + O(t^2)$$
$$= 1 + t \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i} + O(t^2).$$

Hence,

$$V(t) = \int_{D(0)} 1 + t \operatorname{div} f + O(t^2) \, dx, \text{ and thus,}$$

$$\left(\frac{d}{dt}V(t)\right)\Big|_{t=0}$$

$$= \left(\frac{d}{dt}\int_{D(0)} 1 + t \operatorname{div} f + O(t^2) \, dx\right)\Big|_{t=0}$$

$$= \left(\frac{d}{dt}\int_{D(0)} dx + \frac{d}{dt}\int_{D(0)} t \operatorname{div} f \, dx\right.$$

$$\left. + \frac{d}{dt}\int_{D(0)} O(t^2) \, dx\right)\Big|_{t=0}$$

$$= \int_{D(0)} \operatorname{div} f \, dx$$

$$\frac{d}{dt}V(t) = \int_{D(0)} \operatorname{div} f \, dx \qquad \square$$

Proof of Liouville theorem. Now, let ϕ_t denote the Hamilton flow. Since the global solution exists for all t, equation (2) becomes:

$$\phi_t(q,p) = (q,p) + f(q,p)t + O(t^2) =$$

$$(q,p) + [(\dot{q} + \dot{p})(q,p)]t + O(t^2)$$

$$\Rightarrow f = (\dot{q} + \dot{p})$$

And so, using the Hamilton equations

$$\operatorname{div} f = \operatorname{div}(\dot{q} + \dot{p}) = \operatorname{div} \ \dot{q} + \operatorname{div} \ \dot{p}$$

$$\overset{(1)}{=} \operatorname{div} \left(\nabla_p H\right) + \operatorname{div} \left(-\nabla_q H\right) =$$

$$\nabla_q \left(\nabla_p H\right) - \nabla_p \left(\nabla_q H\right) \equiv 0$$

Hence by Proposition ,

$$\frac{d}{dt}V(t) = 0 \Rightarrow \forall t, \ Vol(D_t) = Vol(D) \qquad \square$$

Observables

Given the phase space, an *observable* is a C^2 function $f : M \to \mathbb{R}$. For example, the k^{th} coordinate functions of p and q, p_k and q_k, are observables; the Hamilton function itself is an observable.

For an observable $f(x)$, set $f_t(x) = f(x_t)$, where $t \to x_t$ is the Hamilton flow with $x_\circ = x$. Then, $t \to f_t$ is the Hamilton flow on functions (observables).

$$\frac{d}{dt}\big(f_t(x)\big) = \frac{d}{dt}\big(f(x_t)\big) = \frac{d}{dt}\big(f(q_t, p_t)\big)$$

$$= \nabla_q f \cdot \dot{q}_t + \nabla_p f \cdot \dot{p}_t \overset{(1)}{=}$$

$$\underbrace{(\nabla_q f \cdot \nabla_p H - \nabla_p f \cdot \nabla_q H)}_{\text{Poisson bracket}}(q_t, p_t) =: \{H, f\}(x_t)$$

Hence $\frac{d}{dt}(f_t) = \{H, f\}_t$.

States of classical systems

We are given a C^2 function $\rho(q,p)$, $\rho : M \to \mathbb{R}_+$, $\rho(q,p) \geq 0$ s.t.

$$\int_M \rho(q,p)\, dq\, dp = 1$$

Then ρ is the initial density of positions and momenta. If B is a box in M, the probability that initially the system has a particle in B, is $\int_B \rho(q,p)\, dq\, dp$.

The classical system is initially at *inverse temperature* β ($\mathrm{T} = \frac{1}{\beta}$ is the physical temperature) if

$$\rho(q,p) = \frac{e^{-\beta H(q,p)}}{Z}, \tag{3}$$

where $Z = \int_M e^{-\beta H(q,p)}\, dq\, dp < \infty$. The right hand side of equation (3) is referred to as *Gibbs cannonical ensemble*. Now, given $\rho(q,p)$ (the initial density), we can define the density at time t as $\rho_t(q,p) = \rho(q_{-t}, p_{-t})$, because by Liouville's theorem, $\int_M \rho_t(q,p)\, dq\, dp = 1$. Then the expected value of the observable f at time t given the initial state ρ is

$$\int_M f_t(q,p)\rho(q,p)\, dq\, dp = \int_M f(q,p)\rho_t(q,p)\, dq\, dp.$$

Gibbs cannonical ensemble is invariant under the flow since $H(q_t, p_t) = H(q,p)$ for all t. If $\rho(q,p) = F\big(H(q,p)\big)$, then ρ is also invariant under the flow.

In the non-equilibrium case, the initial measure ρ is *not* invariant under the flow.

A First Look at Entropy Production

We start with the phase space $M = \mathbb{R}^N \oplus \mathbb{R}^N$, Hamilton flow associated to H, $\phi^t(x) := x_t$, and initial state $\rho(q,p)\, dq\, dp$, $\rho(q,p) > 0$, $\forall p, q$. We shall assume that ρ is *not* invariant under the flow. Then, state at time t is

$$\rho_t(q,p)\, dq\, dp = \rho(q_{-t}, p_{-t})\, dq\, dp$$

Note that by Liouville theorem,

$$\int_M \rho_t(q_{-t}, p_{-t})\, dq\, dp = 1$$

Consider the Radon-Nykodym derivative $\frac{\rho_t}{\rho} = h_t$. We have, for any observable f,

$$\int_M f\ \rho_t\, dq\, dp = \int_M f\ h\ \rho d\, dq\, dp$$

We have,

$$h_t(q,p) = \frac{\rho_t(q,p)}{\rho(q,p)} = e^{\ln \rho_t(q,p) - \ln \rho(q,p)}$$

Then, we define

$$l_t = \ln \rho_t(q,p) - \ln \rho(q,p)$$

the total entropy produced by the system in the time interval $[0,t]$. The function $\sigma = \frac{d}{dt} l_t\big|_{t=0}$ is called the *entropy production observable* of the system.

Let us now compute σ:

$$\sigma = \frac{d}{dt}\big(\ln \rho_t(q,p)\big) - \frac{d}{dt}\big(\overbrace{\ln \rho(q,p)}^{\text{constant term}}\big)$$

$$= \frac{d}{dt}\big(\ln \rho(q_{-t}, p_{-t})\big) + 0$$

$$= \frac{1}{\rho(q,p)}\Big((\nabla_q\rho)(-\nabla_p H) + (\nabla_p\rho)(\nabla_q H)\Big)$$
$$= \frac{-1}{\rho(q,p)}\Big((\nabla_q\rho)(\nabla_p H) - (\nabla_p\rho)(\nabla_q H)\Big).$$

And thus,

$$\sigma = \frac{-1}{\rho(q,p)}\,\{H,\rho\} \tag{4}$$

Open classical systems

Take k hamiltonian systems, $M_j = \mathbb{R}^{N_j} \oplus \mathbb{R}^{N_j}$, H_j - Hamiltonian, $\rho_j(q,p) = \frac{e^{-\beta_j H_j(q,p)}}{Z_j}$. That is the j^{th} system is in thermal equilibrium at inverse temperature β_j.

A coupled system, in *absence of interaction* is

$$M = (\mathbb{R}^N_1 \oplus \ldots \oplus \mathbb{R}^N_k) \oplus (\mathbb{R}^N_1 \oplus \ldots \oplus \mathbb{R}^N_k),$$

$H = \sum_{j=1}^k H_j$, because each H_j depends on its own variables only. Initial state

$$\rho = \prod_{j=1}^k \rho_j = \frac{e^{-\beta_1 H_1 - \cdots - \beta_k H_k}}{Z}.$$

If the inverse temperatures are different, then the initial state is the non-equilibrium state. The interaction is another Hamiltonian, $V : M \to \mathbb{R}$, which depends (in general) on all variables. The full Hamiltonian $H_V = H + V$, so V is the interaction term that allows for the energy transfer between the systems.

Note that ρ is invariant under the flow induced by H, but (in general) *not* for the flow induced by H_V.

Now, by Equation (4),

$$\sigma \overset{(4)}{=} \frac{-1}{\rho(q,p)}\,\{H,\rho\} = \frac{-1}{\prod_{j=1}^k \rho_j}\,\{H+V, \prod_{j=1}^k \rho_j\}$$
$$= \frac{-\prod_{j=1}^k Z_j}{\prod_{j=1}^k e^{-\beta_j H_j}}\,\{H+V, \prod_{j=1}^k \frac{e^{-\beta_j H_j(q,p)}}{Z_j}\}$$

Canceling Z_j's and using distributivity of the Poisson bracket we get:

$$\frac{-1}{\prod_{j=1}^k e^{-\beta_j H_j}}\Big[\sum_{j=1}^k \underbrace{\{H_j, \prod_{j=1}^k e^{-\beta_j H_j(q,p)}\}}_{=A_j} +$$
$$+\{V, \prod_{j=1}^k e^{-\beta_j H_j(q,p)}\}\Big].$$

Let us compute A_j for an arbitrary j:

$$A_j = \{H_j, \prod_{j=1}^k e^{-\beta_j H_j(q,p)}\} \overset{def}{=}$$
$$\nabla_q \prod_{j=1}^k e^{-\beta_j H_j(q,p)} \cdot \nabla_p H_j$$
$$-\nabla_p \prod_{j=1}^k e^{-\beta_j H_j(q,p)} \cdot \nabla_q H_j$$
$$= \big(e^{\sum_{j=1}^k -\beta_j H_j(q,p)}\big)\times$$
$$\times\big[\nabla_q(\sum_{j=1}^k -\beta_j H_j(q,p)) \cdot \nabla_p H_j$$
$$-\nabla_p(\sum_{j=1}^k -\beta_j H_j(q,p)) \cdot \nabla_q H_j\big]$$
$$= \big(e^{\sum_{j=1}^k -\beta_j H_j(q,p)}\big)\times$$
$$\times\big[\sum_{j=1}^k -\beta_j \nabla_q(H_j) \cdot \nabla_p H_j$$
$$-\sum_{j=1}^k -\beta_j \nabla_p(H_j) \cdot \nabla_q H_j\big] = 0$$

Hence, $\sum_{j=1}^k A_j = 0$, and we have

$$\sigma = \frac{-1}{\prod_{j=1}^k e^{-\beta_j H_j}}\,\{V, \prod_{j=1}^k e^{-\beta_j H_j(q,p)}\} \overset{def}{=}$$
$$\frac{-1}{\prod_{j=1}^k e^{-\beta_j H_j}}\,(\nabla_q \prod_{j=1}^k e^{-\beta_j H_j} \cdot \nabla_p V -$$
$$\nabla_p \prod_{j=1}^k e^{-\beta_j H_j} \cdot \nabla_q V).$$

Using the expansion of the Poisson bracket from our computation of A_j,

$$\sigma = \frac{-\prod_{j=1}^k e^{-\beta_j H_j}}{\prod_{j=1}^k e^{-\beta_j H_j}}\times$$
$$\times\big[\sum_{j=1}^k -\beta_j \nabla_q(H_j)\cdot\nabla_p V - \sum_{j=1}^k -\beta_j \nabla_p(H_j)\cdot\nabla_q V\big]$$

$$= -1 \cdot \Big[\sum_{j=1}^{k} -\beta_j (\nabla_q (H_j) \cdot \nabla_p V - \nabla_p (H_j) \cdot \nabla_q V) \Big]$$

$$= -\sum_{j=1}^{k} \big(-\beta_j \{V, H_j\} \big).$$

Therefore,

$$\sigma = -\sum_{j=1}^{k} \beta_j \{H_j, V\} \tag{5}$$

Now define $\Phi_j = \{H_j, V\}$, then,

$$\sigma = -\sum_{j=1}^{k} \beta_j \Phi_j \tag{6}$$

The physical meaning of Φ_j is the energy flux out of the j^{th} subsystem.

Proof. H_j - the energy (Hamiltonian) of the j^{th} subsystem.

$$H_{jt}(q,p) = H_j(q_t, p_t) \neq const.,$$

because of the term V.

$$\frac{d}{dt}\big(H_{jt}(q,p)\big) = \frac{d}{dt}\big(H_j(q_t,p_t)\big)$$

$$= \nabla_q H_j(q_t,p_t) \cdot \dot{q_t} + \nabla_p H_j(q_t,p_t) \cdot \dot{p_t} \overset{(1)}{=}$$

$$\nabla_q H_j \nabla_p (H+V) + \nabla_p H_j (-\nabla_q (H+V))$$

$$\overset{distributing\ \nabla}{=} (\nabla_q H_j)(\nabla_p H)$$

$$+(\nabla_q H_j)(\nabla_p V) - (\nabla_p H_j)(\nabla_q H)$$

$$-(\nabla_p H_j)(\nabla_q HV).$$

Since each H_j depends only on its own variables, $\nabla_q H_j = \nabla_{q_j} H_j$, and $\nabla_p H_j = \nabla_{p_j} H_j$. Furthermore, $(\nabla_{q_j} H_j) \cdot (\nabla_p H) = (\nabla_{q_j} H_j) \cdot (\nabla_{p_j} H_j)$, because all other non-$j$ coordinates are 0 in the first vector. Hence,

$$\frac{d}{dt}\big(H_{jt}(q,p)\big) = (\nabla_{q_j} H_j) \cdot (\nabla_{p_j} H_j) +$$

$$(\nabla_{q_j} H_j) \cdot (\nabla_{p_j} V) - (\nabla_{p_j} H_j) \cdot (\nabla_{q_j} H_j) -$$

$$(\nabla_{p_j} H_j) \cdot (\nabla_{q_j} V) = (\nabla_{q_j} H_j) \cdot (\nabla_{p_j} V) -$$

$$(\nabla_{p_j} H_j) \cdot (\nabla_{q_j} V) = \{H_j, V\}_t = \Phi_t.$$

□

Conclusion

For Hamiltonian systems in the non-equilibrium case, the state measure ρ is *not* invariant under the flow, and thus, the entropy production observable is non trivial. One then studies the large time ($t \to +\infty$) of the system, trying to understand various phenomena like the approach to equilibrium, the flow of heat, etc.

Mathematically, to understand these phenomena, one needs idealizations:

1. Systems must be infinite (size $\to \infty$).
2. Phenomena emerges only in the limit as $t \to \infty$.

Computation of such limits, especially the second, is too hard for general systems, so one usually looks on particular examples, where thses limits can be taken.

References

[1] Arnold, V. I. *Mathematical methods of classical mechanics* translated by K. Vogtmann and A. Weinstein. New York: Springer-Verlag, 1989.

[2] Landau, L. D. and Lifshitz, E.M. *Mechanics* translated by J.B. Sykes and J.S. Bell. New York: Pergamon Press, 1969.

Any Integer Is the Sum of a Gazillion Primes

Maksym Radziwill

At the turn of the century, the Russian mathematician Schnirelman proved that there is a constant $C > 0$ such that any integer $n \geqslant 2$ can be written as a sum of at most C primes. The aim of this note is to reproduce his elementary (!) proof.

Goldbach's famous conjecture states that every even integer $n \geqslant 2$ can be written as a sum of at most two primes. In particular, if Goldbach is true then every integer $n \geqslant 2$ can be written as a sum of at most three primes (because even $+3$ is odd). The problem being notoriously difficult, Edmund Landau asked at the beginning of the century if one could actualy prove that there is constant C such that any integer $n \geqslant 2$ is the sum of at most C primes. The answer came in the 30's from a Russian mathematician Schnirelman, and quite remarkably his proof was completely elementary. Schnirelman's original approach yielded a rather huge[1] $C \approx 10^9$. Schnirelman method was further refined in recent times to yield $C = 7$ (see [Ram]). However, a few years later than Schnirelman, by a completely different method Vinogradov succeeded in proving that any sufficiently large integer is a sum of at most 4 primes (here sufficiently large can be taken to mean $\geqslant \exp(\exp(9.75))$). Nonetheless, the distinct advantage of Schnirelman's method is that it is simpler, elementary and yields an "effective" result (i.e one that is true for all integers). Its weakness is of course in the size of the constant C. In this note, I propose to prove a variation of Schnirelman's theorem which is the following.

Theorem 1. *There is a $C > 0$ such that every integer $n \geqslant 2$ is a sum of at most C primes.*

Let us start by introducing some preliminary notation.

(Important !) Notation

Given two subsets $A, B \subseteq N$ we define their sum,

$$A + B \quad := \quad \{a + b : a \in A, b \in B\}.$$

In particular, we will write $2A$ to mean $A + A$, and in general kA to mean the sum of A with itself k times. Note that if $0 \in B$ then $A \subseteq A + B$. Thus 0 holds a special position and we will usually assume that the sets we will be dealing with contain 0. Further, given an arbitrary set $A \subseteq N$ we define,

$$A(n) \quad = \quad |A \cap [1, n]|.$$

That is, $A(n)$ is the number of elements in A that are less than n. A natural concept of 'density' for a subset of integers is the so-called natural density,

$$d(A) \quad = \quad \liminf_{n \to \infty} \frac{A(n)}{n}.$$

Thus the sets of even and odd integer have both density $1/2$, which is consistent with our intuition. However, as Schnirelman pointed out, if we are interested in set addition, a better notion of density is the so-called Schnirelman density.

Definition. *Given a set $A \subseteq N$, define its Schnirelman density*

$$\delta(A) \quad = \quad \inf_{n=1,2,\ldots} \frac{A(n)}{n}.$$

Note that there is something very peculiar about $\delta(A)$. Namely if $1 \notin A$, then $A(1)/1 = 0$, hence $\delta(A) = 0$. You may wonder about the utility of such a weird density but as it will turn out this is the *right* concept.

[1] According to Ramare, $C \approx 10^9$ is due to Klimov. There is a lot of contradictory information as to what Schnirelman proved in his paper (indicating that nobody reads it anymore !). There are three kind of claims in the literrature : Schnirelman didn't exhibit any particular $C > 0$ (I believe this one), Schnirelman got $C \approx 10^{10}$ (maybe) and the last claim being that Schnirelman got $C \approx 20$ (I don't believe this one). It would be worthwile to take a look at his original paper. Unfortunately it's in German.

Plan of the proof.

We will prove two theorems from which Schnirelman's theorem will follow. First, we prove Schnirelman's theorem on set addition.

Theorem 2. *Let $A \subseteq N$. Suppose that 0 and 1 belong to A. If $\delta(A) > 0$ there is a k such that $kA = N$. Further, k can be taken to be any integer $> -\log 4/\log(1-\delta(A))$.*

In other words, if $\delta(A) > 0$ then any integer $n \geqslant 1$ can be written as a sum of at most (recall that $0 \in A$!) k elements from A. In light of the theorem, it is now clear why Schnirelman's density makes sense : If $E = \{k : k \text{ even } \geqslant 0\}$ then $\delta(E) = 0$ because $1 \notin E$; and this is really how it should be because $kE = E$ for all $k \geqslant 1$. On the other hand, if $O = \{k : k \text{ odd}\} \cup \{0\}$, then $O + O = N$ and $\delta(O) > 0$. To prove Schnirelman's theorem (Theorem 1) it would be enough to have $\delta(P) > 0$ where $P = \{p : p \text{ prime}\} \cup \{0, 1\}$. However, it is well-known that $\delta(P) = 0$ so this 'naive' approach will not work. Schnirelman's second genius insight (the first was the definition of Schnirelman density) is that $\delta(P+P) > 0$ and that this can be proven ! (Of course we expect $\delta(P+P) = 1/2$ by Goldbach's conjecture.) Thus, our second 'preparatory' theorem reads as follows.

Theorem 3. *If $P = \{p \text{ prime}\} \cup \{0, 1\}$, then $\delta(P + P) > 0$.*

Together Theorems 2 and 3 prove the existence of a k such that $k \cdot (P + P) = N$. Hence any integer n can be written as a sum of at most $2k$ primes and at most k 'ones'. To prove Theorem 1, it remains to write the sum of those ℓ 'ones' $(1 \leqslant \ell \leqslant k)$ as a sum of primes. If $\ell \geqslant 2$ and ℓ is even, write $\ell = 2+\ldots+2$ with $\ell/2$ 'two' . If $\ell \geqslant 2$ and ℓ is odd, write $\ell = 2 + \ldots + 2 + 3$. Finally, when $\ell = 1$ we use a 'trick'. So suppose that we have a representation of the integer n as a sum of at most $2k$ primes and a 1. The integer $n-2$ can be written as a sum of at most $2k$ primes and $k \geqslant a \geqslant 0$ 'ones'. Thus the integer n is a sum of at most $2k$ primes and $a + 2$ 'ones' and now we can use the earlier procedure to write $a + 2$ as a sum of primes! It follows that that every integer can be written as a sum of at most $3k$ primes, and k can be chosen to be any integer bigger than $-\log 4/\log(1 - \delta(P + P))$. An explicit estimate for $\delta(P+P)$ would give an estimate for the constant C appearing in the statement of Theorem 1. Now, since we've shown how to deduce Theorem 1 from Theorems 2 and 3 it remains to prove the latter theorems.

Proof of Theorem 2

The proof is delightful. Let us start with the following lemma.

Lemma 1. *Let $A, B \subseteq N$. Suppose that $0 \in B$ and $1 \in A$. Then*

$$\delta(A + B) \quad \geqslant \quad \delta(A) + \delta(B) - \delta(A)\delta(B).$$

Proof. Let n be an integer and k be the number of elements of A that are less than n (that is $k = A(n)$). Name and order the elements as

$$a_1 < a_2 < \ldots < a_k.$$

Consider $L_i = \{a_i + 1, \ldots, a_{i+1} - 1\}$ the i-th gap between a_i and a_{i+1}. Note that if $b \in B$ and $1 \leqslant b \leqslant |L_i|$ then $(a_i + b) \in (A + B) \cap L_i$. Furthermore, any distinct $b \in B$ with $1 \leqslant b \leqslant |L_i|$ yields a distinct $a_i + b$. Therefore each gap $|L_i|$ contributes at least $B(|L_i|)$ elements to $A + B$ (recall that $B(|L_i|)$ denote the number of elements of $b \in B$ that are less than $|L_i|$). Since $0 \in B$, we also know that $A \subseteq A+B$. Therefore,

$$(A + B)(n) \quad \geqslant \quad A(n) + \sum_{i=1}^{k} B(|L_i|).$$

By the definition of Schnirelman density, $B(n)/n \geqslant \delta(B)$ for all integers. Hence $B(n) \geqslant \delta(B)n$ for all integers $n \geqslant 1$. Also note that the gaps $L_1 \cup \ldots \cup L_k = [1, n] \backslash A$, and since they are disjoint $|L_1| + \ldots + |L_k| = n - A(n)$. With those two observations in mind we see that our earlier sum (and hence $(A + B)(n)$) is at least

$$\begin{aligned}
(A + B)(n) \quad &\geqslant \quad A(n) + \delta(B) \cdot \sum_{i=1}^{k} |L_i| \\
&= \quad A(n) + \delta(B) \cdot (n - A(n)) \\
&= \quad A(n) \cdot (1 - \delta(B)) + \delta(B)n \\
&\geqslant \quad \delta(A) \cdot (1 - \delta(B))n + \delta(B)n \\
&= \quad (\delta(A) + \delta(B) - \delta(A)\delta(B))n .
\end{aligned}$$

Dividing by n and taking the min we obtain $\delta(A + B) \geqslant \delta(A) + \delta(B) - \delta(A)\delta(B)$. □

A simple consequence of the lemma is the following corollary.

Corollary 1. *Let $A \subseteq N$. Suppose that $0, 1 \in A$. Then,*

$$\delta(kA) \geqslant 1 - (1 - \delta(A))^k .$$

Proof. The corollary is proven by induction on k. The case $k = 2$ is exactly the statement of Lemma 1. For the general case, by Lemma 1, we find that $\delta(kA)$ is bigger than

$$\begin{aligned} & \delta(A) + \delta((k-1)A) - \delta(A)\delta((k-1)A) \\ = \ & \delta(A) + (1 - \delta(A))\delta((k-1)A) \\ \geqslant \ & \delta(A) + (1 - \delta(A))(1 - (1 - \delta(A))^{k-1}) \\ = \ & 1 - (1 - \delta(A))^k. \end{aligned}$$

□

Finally, we can conclude and prove the theorem.

Proof. Take an integer k so large so as to make,

$$\delta(kA) \geqslant 1 - (1 - \delta(A))^k > \frac{1}{2}.$$

Fix an arbitrary $n \geqslant 1$; we will show that $n \in 2kA$. Since $\delta(kA) > 1/2$, the two sets

$$\begin{aligned} S_n &= \{a : a \in kA, a \leqslant n\} = kA \cap [1, n] \\ S'_n &= \{n - a : a \in kA, a \leqslant n\} \end{aligned}$$

have both $> n/2$ elements. For S_n, this follows from $|S_n| = (kA)(n) \geqslant \delta(kA)n > n/2$. As for S'_n, it is in bijection with S_n and thus $|S'_n| = |S_n| > n/2$. Note that both S_n and S'_n are subsets of $[1, n]$. If they were disjoint, we would obtain $|S_n| + |S'_n| \leqslant n$, a contradiction, because as we've just shown that both S_n and S'_n are have cardinality $> n/2$! Therefore S_n and S'_n are not disjoint, and hence there are elements a and b in kA (both a, b are $\leqslant n$ but this is not relevant) such that $a = n - b$. Hence $n = a + b \in 2kA$. Since n was arbitrary, we conclude $2kA = N$. □

Proof of Theorem 3.

We will need two lemmas from number theory that we will not prove here.

Lemma 2. *Let $P = \{p : p \text{ prime}\} \cup \{0, 1\}$. There is a constant B such that for all $n \geqslant 2$ we have $P(n) \geqslant Bn/\log n$.*

This lemma is known as Chebyscheff's bound. The constant B could be taken to be 0.92. The second lemma is more involved and although elementary, it is harder to prove.

Lemma 3. *Let $p_2(n)$ denote the number of representations of n as a sum of two elements from $P = \{p : p \text{ prime}\} \cup \{0, 1\}$. There is constant $C > 0$ such that for all $n \geqslant 2$ we have,*

$$p_2(n) \leqslant C \cdot \frac{n}{(\log n)^2} \cdot \prod_{p|n} \left(1 + \frac{2}{p}\right).$$

Here, $\prod_{p|n}$ is a product over the prime divisors of n.

Now we are ready to give the proof of Theorem 3.

Proof. **(Proof of Theorem 3)** Let $P_2 = P + P$ where $P = \{p : p \text{ prime}\} \cup \{0, 1\}$. Let also $p_2(n)$ denote the number of representations of n as a sum of two elements from P. Note that if $n \notin P_2$ then $p_2(n) = 0$. Using this and Cauchy-Schwarz's inequality, we obtain

$$\begin{aligned} \sum_{k \leqslant n} p_2(k) &= \sum_{\substack{k \leqslant n \\ k \in P_2}} p_2(k) \\ &\leqslant P_2(n)^{1/2} \cdot \left(\sum_{k \leqslant n} p_2(k)^2 \right)^{1/2}. \end{aligned}$$

Therefore,

$$P_2(n) \geqslant \left(\sum_{k \leqslant n} p_2(k) \right)^2 \cdot \left(\sum_{k \leqslant n} p_2(k)^2 \right)^{-1}.$$

We will lower bound the first sum by a $Bn^2/(\log n)^2$ (with some constant $B > 0$) and upper bound the second sum by a $Cn^3/(\log n)^4$ (again with some constant $C > 0$). Inserting those bounds in the above inequality will yield

$$P_2(n) \geqslant (B/C) \cdot n.$$

Hence $\delta(P_2) = \delta(P + P) \geqslant B/C > 0$ and that will finish the proof. Thus, it is enough to prove the above stated upper/lower bounds. First, we prove that

$$\sum_{k \leqslant n} p_2(k) \geqslant B \cdot \frac{n^2}{(\log n)^2}.$$

Indeed, note that

$$\begin{aligned}\sum_{k\leqslant n} p_2(k) &= \sum_{k\leqslant n}\sum_{\substack{p,\,q\,\in\,P\\ p+q\,=\,k}} 1\\ &= \sum_{\substack{p,\,q\,\in\,P\\ p+q\,\leqslant\,n}} 1. \qquad (1)\end{aligned}$$

If $p \leqslant n/2$ and $q \leqslant n/2$, then $p+q \leqslant n$. Thus any choice of $p \leqslant n/2$ and $q \leqslant n/2$ gives a contribution to the sum in (1). Therefore, the sum (1) is at least $P(n/2)\cdot P(n/2)$ and by Lemma 2 there is a constant K such that $P(n/2) \geqslant Kn/(\log n)$. We conclude that (1) is at least $K^2n^2/(\log n)^2$, as desired (take $B = K^2$). Now, we prove that

$$\sum_{k\leqslant n} p_2(k)^2 \quleqslant C\cdot\frac{n^3}{(\log n)^4}.$$

Since $k/(\log k)^2$ is an increasing function, by Lemma 3 for all $k \leqslant n$ we have,

$$p_2(k) \quleqslant K\cdot\prod_{p|k}\left(1+\frac{2}{p}\right)\cdot\frac{n}{(\log n)^2}$$

for some constant $K > 0$. Therefore the sum $\sum_{k\leqslant n} p_2(k)^2$ is bounded above by

$$\begin{aligned}&\sum_{k\leqslant n} K^2\cdot\frac{n^2}{(\log n)^4}\prod_{p|k}\left(1+\frac{2}{p}\right)^2\\ \leqslant\ & K^2\cdot\frac{n^2}{(\log n)^4}\sum_{k\leqslant n}\prod_{p|k}\left(1+\frac{8}{p}\right)\end{aligned}$$

using the inequality $(1+2/p)^2 \leqslant (1+8/p)$ valid for $p \geqslant 2$. Now we show that the sum on the right hand side is bounded by Cn for some $C > 0$. The proof is a standard argument in analytic number theory and in some sense does not belong to this article. Rather than trying to justify all the steps, I will just write down the argument and hope you take it on faith, if you didn't see those things before.

$$\begin{aligned}\sum_{k\leqslant n}\prod_{p|n}\left(1+\frac{8}{p}\right) &= \sum_{k\leqslant n}\sum_{d|k}\frac{\mu(d)^2}{d}\cdot 8^{\omega(d)}\\ &= \sum_{d\leqslant n}\frac{\mu(d)^2}{d}\cdot 8^{\omega(d)}\sum_{\substack{k\,\leqslant\,n\\ d|k}} 1\\ &\leqslant \sum_{d\leqslant n}\frac{\mu(d)^2}{d}\cdot 8^{\omega(d)}\cdot\frac{n}{d}\\ &\leqslant n\cdot\sum_{d\geqslant 1}\frac{\mu(d)^2}{d^2}\cdot 8^{\omega(d)}\\ &= n\cdot\prod_{p}\left(1+\frac{8}{p^2}\right)\end{aligned}$$

and the latter product converges because $\sum 8/p^2$ does[2]. □

References.

The exact reference to the improvement $C = 7$ of Schnirelman's constant is the following.

[Ram] O. Ramare, *On Schnirelman Constant*, Annali de la Scuola Superior de Pisa, 1995, pages 645-705. Also available at

math . univ . lille 1. fr / ∼ ramare / Maths / Article . pdf

A nice introduction to additive number theory/combinatorics (this is the 'field' to which Schnirelman's theorem belongs to) can be found at

www . math . dartmouth . edu / ppollack / notes . pdf

Some aspects (but beside this much more) of the subject are in the book 'Sequences' by K.F Roth (the Fields medalist !) and H. Halberstam (the 'god of sieves' !).

[2]The notation used above is standard: the Möbius function $\mu(n)$ is defined to be 1 if n is squarefree and has an even number of prime factors, -1 if n is squarefree with an odd number of prime factors, and $\mu(n)$ is 0 if n is not squarefree. Also, $\omega(n)$ is the number of distinct prime factors of n.

Credits

The Delta-Epsilon Editing Team
In alphabetical order

Ioan Filip

Vincent Larochelle

Daniel Shapero

Phil Sosoe

Alexandre Tomberg

Nan Yang

Cover Art & Design

Linda

Acknowledgements

First of all, we wish to thank professors Claude Crépeau and Dmitry Jakobson for giving us their time and for their collaboration with the magazine. Their presence here tremendously increases the quality of the publication.

Second, we must thank all the writers who have submitted articles; without their work and their desire to communicate an enthusiasm for mathematics and science, the Delta-Epsilon would not exist. In particular we would like to thank Tigran Atoyan, Maya Kaczorowski and Maksym Radziwill for their fascinating papers. We also thank Linda for her beautiful cover art.

We are grateful to the ISM (Institut des sciences mathématiques) and to the Department of Mathematics and Statistics of McGill University for providing part of the funding. We also thank Maple for their support.

We end this acknowledgements section in an usual way. Some advice for next year's team: make sure you recruit at least half of your ranks from the first and the second years, if you don't want to have the same problems we've had all year long in trying to keep up with our own editing deadlines...

Maple™ 12

The Essential Tool for Mathematics and Modeling

12 Redefining Usability

Maple's Smart Document Environment is an intuitive user-interface that automatically captures all of your technical knowledge in an electronic form that seamlessly integrates calculations, explanatory text and math, graphics, images, sound, and more. It has many built-in tools to assist you in analysis and solution development.

12 Extensive Collection of Built-In Mathematical Algorithms

Maple 12 combines the world's most powerful mathematical computation engine with an intuitive user interface. It offers over 4000 mathematical functions, cutting-edge solvers for ODEs, PDEs, and DAEs, highly efficient numeric solvers, and world-leading symbolic solvers.

12 A Vital Component of Your Tool Chain

Maple 12 integrates with CAD, Excel®, and other standard tools. It includes code generation tools for C, Fortran and MATLAB®, and is supported by an extensive range of products for various industrial applications including automotive, aerospace, electronics, energy, and finance.

12 What's New in Maple 12

- A new Dynamic Systems package offers a large selection of analytic and graphing tools for linear time-variant systems, which are essential in control systems development.
- New interactive embedded components include dials, gauges and radio buttons.
- The Exploration Assistant lets you create interactive mini applications to explore the parameters of expressions.
- New plot types include dual axis plots, polar axis plots, and specialized engineering plots.
- Improvements to solvers for ODEs, PDEs and DAEs continue to strengthen Maple's world-leading position in numeric and symbolic differential equation solving.
- The Discrete Transforms collection has been expanded to include numerical wavelet transforms.
- CAD connectivity adds important analysis capabilities to CAD systems.
- MATLAB to Maple code translation allows you to leverage existing work.
- …and more!

Licenses of Maple 12 Professional include the Maple Toolbox for MATLAB, MapleNet, and a 1 year subscription for the Elite Maintenance Program.

To learn more about Maple 12 visit the new Maplesoft Demo Center and see Maple 12 in action!

Go to: www.maplesoft.com/products/maple/demo/

© Maplesoft, a division of Waterloo Maple Inc., 2008. Maplesoft, Maple, and MapleNet are trademarks of Waterloo Maple Inc. All the other trademarks are the property of their respective owners.

www.maplesoft.com

ISSN 1911-9003