Listing of the Summaries of the

Rockefeller Mathematics Seminar
Spring 2008

Organized by: J.J.P. Veerman**), M.J. Feigenbaum, F.M. Tangerman.

Summaries by J. J. P. Veerman and F. M. Tangerman

in collaboration with — and with advice from — the speakers, and E. G. D. Cohen.

Special thanks to:
Melanie Lee*),
for her efficient help in the organization and dissemination of this seminar.

*) To quote her: On MJF leaving the building early: "He could not find anyone to play with"; On JJPV being happy with a great speaker on a sunny day: "All is well in the world of Veerman."

**) e-mail: veerman@pdx.edu

Rare Events in Complex Systems

Eric Van Den Eijnden, Courant Institute,

email: eve2@cims.nyu.edu.

March 20, 2pm

The dynamics of a long molecule in a solvent — say water — may be modeled by

$$\dot{x} = -\nabla V(x) + \sqrt{2\epsilon}\,\eta(t) \quad ,$$

where $\eta(t)$ is white noise, $\epsilon = kT$ (T is temperature, k the Boltzmann constant) if the environment is in thermal equilibrium. V is an empirically determined potential which depends in practice not only on the configuration of the long molecule but also on the surrounding water molecules. The number os relevant degrees of freedom may easily number in the thousands, and the energy landscape is known to be very rugged. Denote this space by Ω.

When ϵ is very large the motion is Brownian. When $\epsilon = 0$ the motion is a gradient flow to the local minimum. When ϵ is small enough, the transition probabilities from one minimum to the next can be computed by using the "large deviation limit" which allows us to model this process as a (computable) Markov process (assuming you know V). There are two problems with this. For biologically interesting molecules the landscape V is far too complex to efficiently compute. Second, for given $\epsilon > 0$ the deviations from equilibrium might not be computable from the limit for small ϵ. This occurs in the present context, where the size of ϵ at room temperature overwhelms all but very few of the largest energy differences.

Define a probability density function

$$m(x) \equiv \left(\int e^{-V(x)/\epsilon}\, dx \right)^{-1} e^{-V(x)/\epsilon} \quad ,$$

and consider an 'ergodic' trajectory $X(t)$, that is: a continuous path $[0, \infty) \to \Omega$ such that for all continuous functions ϕ

$$\int \phi(X(t))\, dt \longrightarrow \int \phi(x) m(x)\, dx \quad .$$

Assume for simplicity that V has only two "deep" minima a and b and let A and B be disjoint open sets containing a respectively b. Given the ergodic path X, define R as as the set of intervals $[a, b]$ such that $X : [a, b] \to \Omega$ is a path from $\overline{A}$ to $\overline{B}$ without intersecting A or B. If χ is the usual characteristic function, one can show that there exist a probability density m_R such that

$$\int \phi(X(t)) \chi_R\, dt \longrightarrow \int \phi(x) m_R(x)\, dx \quad .$$

Furthermore, this density decomposes as:

$$m_R(x) = q_+(x) q_-(x) m(x)$$

where $q_+(x)$ (or $q_-(x)$) is the probability that a path starting at x leads first to $\overline{B}$ (or started last in $\overline{A}$). The probability current J_R associated to m_R is used to compute the probability flux through any surface. In this case it turns out to be given by

$$J_R(x) = \epsilon m(x) \nabla q_+(x) \quad .$$

In what amounted to a full talk after the 'official' talk had ended, Van Den Eijnden showed how this current may explicitly be computed for (at least) moderately complex molecules. He employed similar path integral methods as Onsager and Machlup introduced. It turns out that the most likely paths thus constructed can indeed be different from the ones constructed from the "large deviation limit".

Flocks and Formations: ODE's coupled along graphs

J. J. P. Veerman, Portland State Univ. and Rockefeller Univ.,

email: veerman@pdx.edu.

March 27, 2pm

We aim to understand the guidance and control laws for 'flocking behavior', using local steering laws. Every agent controls its motion, acceleration, by observing and weighing the (relative) motion of its neighbors, in a sense to made precise. The orbit of some of these, the leaders, is determined entirely by outside influences.

The communication graph is a (directed weighted) graph whose nodes represent the agents in the flock and such that there is a directed edge from agent j to agent i when agent i uses agent j for its steering law. For any edge from i to j there is a positive weight $w(i,j)$. When i is a leader, all $w(i,j)$ are zero "leaders see no-one"), for all other i we have $\sum_j w(i,j) = 1$. The selection of communication graph and weights might depend on the agent's field of view, range to other agent's etc. These factors effect flock dynamics.

Given a communication graph, let k denote a function that to each node of the graph assigns a position or velocity in $\mathbb{R}^m$. The 'Laplacian' operating on this function k is defined as: $L(k)(i) = \sum_j w(i,j)(k(i) - k(j))$. Now let h be the fixed function defining the positions of the configuration, and x the position of each of the agents (depends on time). The model is given by

$$\ddot{x}(i) = f \cdot ((Lx)(i) - Lh(i)) + g \cdot (L\dot{x})(i) + u(i)$$

Here u models extraneous forces (threat response and noise), or in the case of a convoy the response of the 'leader' to a change in traffic light, and f and g are real constants to be chosen so that the performance is optimized.

An 'in formation' motion is is a motion whereby all agents have the same velocity and their relative positions are constant and given by an a priori selected configuration. The central question is: under what conditions do all initial conditions converge to an in formation solutions, and how fast is that convergence? How is that convergence influenced by the communication graph (for example the size of the flock)?

Theorem 1: *Given L. The parameters f and g can be chosen so that for all desired configurations h, all solutions are asymptotic to an in formation motion* iff *a certain graph theoretical condition holds. In that case f and g are negative.*

Next consider the effect of perturbation on the motion of the leader in a stable flock. The question is how much does this perturbation get amplified during its propagation through the flock and before it dies out. Assume that there is one leader who has index 0 and that all conditions of Theorem 1 are satisfied. Consider a time dependent perturbation of its motion by prescribing $u_0(t)$. Analyze its effect via the Fourier transform $\hat{u}_0(\omega)$. Let $a(\omega) = \hat{x}(\omega)$ be the frequency response function, i.e. the solution of the Fourier transform of the above equation:

$$-\omega^2 a(\omega) = fLa(\omega) + i\omega gLa(\omega) + \hat{u}(\omega) \quad \text{or} \quad \left(L + \frac{\omega^2}{f + i\omega g} I\right) a = \hat{u}(\omega) \quad \text{where} \quad \hat{u}(\omega) = (\hat{u}_0(\omega), 0, ..., 0) \quad .$$

One can show that, except at 0, all eigenvalues of L have positive real part and so this equation has a unique solution $a(\omega)$ when $(\omega \neq 0)$.

We look at convoys on a one lane road $(x_0 > x_1 > .....x_N)$. Every agent looks only at the car directly in front of it and the one directly behind it, except for the first and last car. The last car only sees the car in front of it, and the first car is a leader reacts to traffic lights, so that his orbit is given. The two parameters in the problem are the weight $\rho \equiv w(i, i+1)$ of the car behind and the size N of the graph.

Theorem 2: *For the convoy problem the response function satisfies:*
i) Looking forward or $\rho \in (0, 1/2)$: a_N is exponentially increasing in N for small ω.
ii): Looking backward or $\rho \in (1/2, 1)$: a_N decreases exponentially in N for small ω, except at one peak for very small *ω, where a_N is exponentially increasing in N.*
iii): $\rho = 1/2$: a_N grows at worst linearly in N.

Relative Mapping Class Group Techniques in Surface Dynamics

Michael Handel, CUNY Grad. Center and Lehman College,

email: Michael.Handel@lehman.cuny.edu.

April 4, 2pm

Two homeomorphisms h_0 and h_1 of a surface S to itself are isotopic if there is a continuous $F : S \times [0,1] \to S$ such that $F_0 = h_0$ and $F_1 = h_1$ and all $F_t : S \to S$ are homeomorphisms. this is an equivalence relation. The mapping class group $\mathbf{MCG}^*(S)$ of an orientable surface S of finite type is the group of isotopy classes of orientation preserving homeomorphisms of S. A homeomorhphism h is a representative of its equivalence class $\tau_h \in \mathbf{MCG}^*(S)$.

A well-known result states that τ_h is entirely determined by the action of h on closed curves. As an example, consider the torus T^2, represented as the unit square with the standard identifications. A linear (and orientation preserving) homeomorphism $A : T^2 \to T^2$ can be represented by a matrix with determinant 1. A closed curve goes around p times in the x direction and q in the y direction, where p and q are integers. Naturally A maps closed curves to closed curves, and so the matrix representing A has integer entries. We have derived that $\mathbf{MCG}^*(T^2) = \mathbf{SL}(2,\mathbb{Z})$ the Special Linear Group.

We distinguish 3 types of linear maps on tori: periodic — for example $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, pseudo-Anosov — the standard example is $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, and a Dehn Twist or a reducible map — as in $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Pseudo-Anosov means that the manifold has a 1 dimensional stable and a 1 dimensional unstable foliation (determined by the stable and unstable eigenvectors in the torus example, along which the map contracts or expands). The 'reducible' case can by further analysis be reduced to the other two (no details given). The linear map A is in some sense a canonical representative of its class. For example, A minimizes the number of periodic orbits and the topological entropy in its class.

Thurston's Classification Theorem states that this decomposition holds for any surface S and its leads to the construction of a canonical representative ϕ_τ (a Thurston normal form) of an equivalence class τ. In fact ϕ_τ has an invariant set R in S such that $S \backslash R$ is the disjoint union of surfaces of lower genus S_i. Furthermore on each S_i: $\phi_\tau|_{S_i}$ is either periodic or pseudo-Anosov.

Given a surface S and a set K in it, the Relative Mapping Class Group $\mathbf{MCG}^*(S \backslash K)$is defined as follows. Two orientation preserving homeomorphisms h_0 and h_1 correspond to the same equivalence class if there is a continuous $F : S \times [0,1] \to S$ such that $F_0 = h_0$ and $F_1 = h_1$ and all $F_t : S \to S$ are homeomorphisms *and fix each point in* K. We look at two examples where this notion is useful.

Let $K = \{a, b, c, \infty\}$ be 4 points on the 2-sphere S^2. Given any homeomorphism h of the sphere fixing K. We know that h is isotopic to identity by an isotopy $F : S^2 \times [0,1] \to S^2$. Consider the "track" the image for $t \in [0,1]$ of $F(a,t)$ in $\mathbb{S}^2 \backslash K$ (a punctured surface). This curve tells us what the relative mapping class of h is. One proves that this curve is non-contractible on $\mathbb{S}^2 \backslash K$, then h has a horseshoe, and thus positive topological entropy.

Brouwer's Plane translation Theorem states that if a homeomorphism $h : \mathbb{R}^2 \to \mathbb{R}^2$ has no fixed points, then an orbit $\{h^n(x)\}_{n \in \mathbb{Z}}$ has no accumulation points. Here is a very coarse outline of its proof. If the Theorem false, then there is an orbit that has an accumulation point. By modifying h slightly one can then arrange for h to have a periodic orbit. But the Thurston normal form of a periodic map of the plane (and which fixes infinity) is a rotation which has a fixed point. Therefore h has a fixed point, which is a contradiction.

Optimization of Information Flow in Gene Regulation

William Bialek, Princeton Univ. and Rockefeller Univ.,

email: wbialek@princeton.edu

April 10, 2pm

Gene expression is the process by which inheritable information from a gene, such as the DNA sequence, is made into a functional gene product, such as protein or RNA. A central process of life is the regulation of gene expression. At least in some cases one can think of this process as a transformation from inputs (transcription factor concentrations which control the transfer of genetic information) to outputs (expression levels).

Let X denote the input (transcription factor in units of concentrations or molecule counts) and Y the output (expression levels). The activation graph $\overline{Y}(X)$ is often sigmoidal: flat at low and high X concentrations and steep in the middle. We replace a simple input-output relation (a graph $Y(X)$), by a stochastic representation: a model for the probability distribution $P(Y|X)$ of Y, at a given X concentration. We will assume that $P(Y|X)$ is Gaussian, described by a mean $\overline{Y}(X)$, and a variance $\sigma^2_Y(X)$.

A reasonable class of variance models has this form:

$$\sigma^2_Y(X) = \alpha\overline{Y}(X) + \beta(\frac{\partial\overline{Y}}{\partial X})^2 X$$

The first term on the right captures the effects of "low concentration" noise, and the second captures process noise where the activation curve is steep. This relation will be used as a constraint in what follows.

We thus represent the process by an input-output channel with these data: $P_1(X)$ the distribution of X (independent of Y), the conditional probability $P(Y|X)$ and the output distribution $P_2(Y)$ (independent of X). The mutual information $I(X,Y)$ between the distributions for X and Y is defined as

$$I(X,Y) \equiv \int\int P(X,Y) \ln \frac{P(X,Y)}{P_1(X)P_2(Y)}\, dX\, dY$$

Note that when X and Y are independent then $I(X,Y) = 0$ and recall that $P(X,Y) = P(Y|X)P_1(X)$.

This sets the stage for the application of Shannon's communication theory principles. Examples are:
- Fix $P(Y|X)$ and P_2. Choose $P_1(x)$ to maximize $I(X,Y)$.
- Fix P_1 and $I(X,Y)$. Adjust $P(Y|X)$ to control the output P_2 ("distortion theory").
- Fix P_1 and $P(X,Y)$. Choose $P(Y|X)$ to minimize I and analyze the limiting behavior when σ^2_Y tends to zero.

The limits for the first case (maximize $I(X,Y)$) are analyzed in the low noise limit. These are some of the results:

- If the first term in the standard deviation dominates in the noise model, then the limit has P_2 decreasing in Y.
- If the second term dominates in the noise model then the limiting distribution is bi-modal, with two extrema: one at a low Y concentration and at a high Y concentration. This case provides an information theoretic basis for for the prevalence of 'on-off' chemical activation laws.

Next generalize to the case of a single input X and multiple outputs $Y_1, ..., Y_N$. It is often the case that:

- There are correlations between the outputs $Y_1, ..., Y_N$. In particular some outputs may inhibit others.
- The on-off relations between X and Y_i, found earlier, also suggest on-off relations between $Y_1, ..., Y_N$.

Consider the curve $(\overline{Y}_1(X), ...\overline{Y}_N(X))$ as the noise variance tends to zero. Maximizing mutual information provides a 'length' for this curve in $(Y_1, ...Y_N)$ space. Thus maximization mutual information maximization would favor curves that pass close to as many edges and corners as possible in the $(Y_1, ..., Y_N)$ cube.

References for this talk are:

G Tkacik, CG Callan Jr, W Bialek, *Information flow and optimization in transcriptional regulation*, arXiv:0705.0313 [qbio.MN] (2007).
G Tkacik, CG Callan Jr, W Bialek, *Information capacity of genetic regulatory elements*, arXiv:0709.4209 [qbio.MN] (2007).

Random Matrices are Universal.

Percy Deift, Courant Inst.

email: deift@courant.nyu.edu

April 17, 2pm

In the last decennia Random Matrices (in particular the expectation of the spacings in their spectra) have been found to describe many phenomena in physics and mathematics. In each of these one finds that after a substantial amount of mathematics or experimentation there is quantitative agreement with Random Matrix Theory. But the relation is strictly *a posteriori*: the phenomena discussed below bear no (as yet) apparent mathematical relation to one another. Is there a mathematical theory underlying all these phenomena ?

The Gaussian Unitary Ensemble of degree N (or GUE_N) is the set of N by N unitary ($M^\dagger M = I$) matrices together with a probability measure (or probability density). Typically the measure is given by a Gaussian $Ce^{-\sum |m_{ij}|^2}\, dM$, where m_{ij} are the entries of the matrix M and dM denotes Lebesgue measure on the algebraically independent entries of M. The constant C normalizes the total measure to 1. The Gaussian Orthogonal Ensemble (GOE_N) is the restriction of this to matrices with real coefficients together with the induced measure. The statement that property X holds for a 'random matrix' means (roughly speaking) that the property is true for a set of matrices of full measure in GUE_N or GOE_N.

Among the most salient properties of random matrices is the fact that their spectrum is not at all random. In particular the tendency to bunch up is much less than in an sequence of random points (with the Poisson distribution.) In fact, in the space of symmetric $N \times N$ matrices, the subspace of matrices with double eigenvalues has codimension 2 (Wigner, Von Neumann, 1929). On the other hand, the set of points in $\mathbb{R}^N$ such that two components are equal form $\binom{N}{2}$ codimension 1 hyperplanes. Thus the chances of two eigenvalues being very close to each other is much smaller than the chance of two random points in $\mathbb{R}$ being very close (this is called level-repulsion).

1: Neutron scattering. Neutrons shot at a heavy nucleus are strongly repelled from the nucleus only at certain energies, called scattering resonances. The statistical properties of these resonances are exactly those of spectra of GOE_N as was noted by Wigner in 1957 from physical observations.

2: Riemann's Zeta Function. Rescale the imaginary part of the zeroes $\frac{1}{2} + i\overline{\gamma}_j$ of the Riemann zeta function so their mean distance is 1. Given an interval (a, b) denote by $R(a, b)$ the fraction of distances among $\{\gamma_i - \gamma_j\}$ that is in (a, b). Then (roughly speaking) $R(a, b)$ is the same as for the spectrum of GUE. (Proved by Montgomery in 1973.)

3: Longest Increasing Subsequence. Let $\pi : \{1, \cdots N\} \to \{1, \cdots N\}$ be a permutation and, given π, denote by $q_N(\pi)$ the length of the longest increasing subsequence for a permutation π of N numbers. Then for random permutations π, the quantity Prob $(q_N(\pi) < t)$ (after suitable scaling) statistically behaves exactly like the leading eigenvalue of matrices in GUE_N. (Proved by Baik, Deift, and Johansson, 1999).

4: Deregulated Mexican Buses. In Cuernavaca (close to Mexico DF) Krbálek and Ŝkeba in 2000 observed that bus drivers in Cuernavaca optimizes their bus' income by making use of observers along their route. These would signal them to slow down or hurry up according the time lapsed since the last bus had passed. The distribution of time lapses between consecutive buses was that of the GUE_N spectrum.

5: Vicious Walkers. N walkers are positioned at positions $\{0, 1, 2, \cdots\}$ in $\mathbb{Z}$. At every time step a single walker chosen at random among those walkers who do not have a neighbor at their immediate left, takes one step to the left. The distance d_N of the leftmost walker after N time steps behaves (as $N \to \infty$) like the leading eigenvalue of matrices in GOE_N. (olved by Forrester in 1999).

6: The Aztec Diamond. Cut up the unit square in a grid of $N + 1$ by $N + 1$ little squares. The diamond is the inside of the line-figure connecting the midpoints of each of the four sides of the unit square. Now tile the squares contained within the diamond by dominoes that measure 2 by 1 little squares. Very close to each of the four extremities almost all tilings are 'frozen' into a standard configuration. The boundary between these frozen regions and the variable insides is called the "arctic circle". Fluctuations in the positions of these arctic circles were proved to be related to the the distribution of the leading eigenvalue of GUE_N by Johansson in 2002.

Large Deviations and Statistical Mechanics

Hugo Touchette, Queen Mary College, London.

email: ht@maths.qmul.ac.uk

April 24, 2pm

Consider a combination S_n of a sequence of random variables $\{X_i\}_{i=1}^n$, such as coin tosses. The most obvious example is the sample average: $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Each X_i has the same probability density $p(x)$ (the 'probability' that X equals x) and the X_i may or may not be independent of one another. The object of Large Deviations Theory is the probability density $P_n(a) \equiv P(S_n = a)$.

The probability density $P(S_n = a)$ is said to satisfy the *Large Deviations Principle* or *LDP* if $I(a)$ exists where

$$I(a) \equiv \lim_{n\to\infty} \frac{1}{n} \ln P(a) \geq 0 \quad .$$

One writes $P(S_n = a) \approx e^{-nI(a)}$, and I is called the *rate function*. The goal of the theory is i) to show that an LDP exists and ii) to calculate the rate function.

Two important tools are the Gärtner-Ellis Theorem and the Contraction Principle. The function $\lambda(k) \equiv \lim_{n\to\infty} \frac{1}{n} \ln \int e^{nkS_n}\, dx$ for $k \in \mathbb{R}$ are called *Scaled Cumulant Generating Functions* or *SCGF*.

Gärtner-Ellis Theorem: *If $\lambda(k)$ exists and is differentiable, then $P_n(a) \approx e^{-nI(a)}$ and I is the Legendre transform of λ:*

$$I(a) = \max_k \{ka - \lambda(k)\} \quad .$$

Let $f : X_A \to X_B$ a continuous function. Suppose we have 'cumulant' random variables A_n on X_A and B_n on X_B, such that $P(B_n = b) = \int_{f^{-1}} P(A_n = a)\, da$ (or ($P_n(b)$ is the push-forward of $P_n(a)$.

Contraction Principle: *If the $\{A_n\}$ satisfy the LDP, then $\{B_n\}$ satisfy the LDP and*

$$P(B_n = b) \approx e^{-nI(b)} \qquad \text{where} \qquad I_B(b) = \min_{a:f(a)=b} I_A(a) \quad .$$

The Legendre transform in Gärtner-Ellis Theorem always yields a convex rate function. This is unrealistic for example in the presence of a phase transition where rate functions typically have two local minima. In these cases the Contraction Principle often helps.

Suppose the X_i are iid (independent and identically distributed). it is clear that $\int e^{nkS_n} = \int \Pi^n e^{kX_i} = \left(\int e^k X\right)^n$. Thus for individuals with Gaussian distributions $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, one easily calculates that $\lambda(k) = \frac{1}{2}k^2\sigma^2 + \mu k$ and thus that $I(s) = \frac{(x-\mu)^2}{2\sigma^2}$. If the distributions are Poisson with average μ one derives in a similar way that $I(s) = \frac{s}{\mu} - \ln\frac{s}{\mu} - 1$. The fact that the rate function I possesses a *minimum* whose value equals zero is equivalent with the *Law of Large Numbers*. The fact that the rate function is *quadratic* around its minimum is equivalent with the *Central Limit Theorem*.

Here lies the crux of the matter. The central limit theorem states that the probability density of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ converges to the normal distribution. The sample average given above is scaled by $\frac{1}{n}$ (as opposed to $\frac{1}{\sqrt{n}}$) and is thus a different random variable. Essentially, the central limit theorem looks at the small fluctuations of the sum of random variables — the large deviation theory looks at the large fluctuations. In the two examples given in the previous paragraph, the large deviations are indeed very distinct.

The microcanonical ensemble is the probability space describing possible configurations of a system with a fixed number of particles (n), energy (E), and volume (V). The microstate $\omega = \omega_1\omega_2\cdots\omega_n$ consists of the positions and velocities of each particle. The macrostate $M_n(\omega)$ is some observable such as the velocity distribution. consider the probability density $P_E(M_n = m)$. The microcanonical LDP states that $P_E(m) \approx e^{-nI_E(m)}$. Usually this is written in terms of the entropy $S(m) \equiv -I_E(m)$. the most probable state m is of course the one that minimizes I (or maximizes S). In physics this is called the maximum entropy principle.

Similarly, the canonical ensemble is the probability space describing the states of the system of n particles held at constant volume and constant temperature. Now the canonical LDP states that $P_\beta(m) \approx e^{-nI_\beta(m)}$, where this rate

function is called the 'free energy' and denoted by F. A large number of quantities for equilibrium statistical physics can be interpreted in terms large deviations and this reformulation shows promise also for non-equilibrium statistical physics.

Note: A standard reference is: A. Dembo and O. Zeitouni. Large Deviations Techniques and Applications. Springer, New York, 2nd edition, 1998.

Simulation of Biomolecules on Supercomputers

Yuefan Deng, SUNY Stony Brook and Brookhaven Nat'l Lab.

email: Yuefan.Deng@StonyBrook.edu

May 1, 2pm

The physics underlying molecular biology is essentially understood at small length- and time-scales. The size of an atom is of the order of one ångström or 10^{-10} m. The time it takes light to cross a few thousand atomic diameters is a femtosecond or 10^{-15} seconds (the smallest time scale at which interactions in *molecular dynamics* are immediate). (It is also the timescale for atomic vibration within a typical molecule). However processes important to molecular biology (such as folding of large molecules) take place within time scales of nanoseconds — 10^{-9} s — to microseconds — 10^{-6} s — and involve on the order of 10^6 atoms or more. These processes moreover critically involve the dynamics of many layers of the surrounding solvent (typically water). Even for, biologically speaking, small molecules, this presents a huge computational challenge due to the amount of the interactions among the atoms and that need to be computed and the relatively large time scale over which these computations need to be accurate. Molecular Dynamics (MD) for biological molecules involves the stages of modeling, algorithms, software, and hardware.

Modeling: The starting point is to use an accurate potential $U(x_1, ..., x_N)$ (N is of the order of a million) modeling interactions on the atomic scale. MD is then given by

$$m_i\ddot{(x_i)} = -\nabla_{x_i} U(x_1, ..., x_N) \quad ,$$

where the potential includes: atomic vibrations, Van Der Waals forces, ionization levels (to model PH), Lennart-Jones forces, and Coulomb forces. Temperature is an energy level $E = \sum_i \frac{1}{2} m_i \dot{x}_i^2 + U$

Algorithms: to solve the evolution equation there are two issues: the computation of the potential at any time, and a limiting maximal timestep that can be taken, on the order of femto-seconds. The potential among N atoms involves N^2 terms. One decomposes the potential in two terms, a local (short-range interaction), and the long range interaction: the Coulomb potential (slowly decaying as $1/N$ and thus acting over long distances). The *Ewald summation* to the Coulomb force is employed here. This approximation has two terms, a rapidly time varying, but local term, and a slowly time varying global term. The second problem is to time-step carefully while preserving integrals of motion.

To obtain useful results in the life-time of a researcher the approach chosen is to harness many processors (10,000 minimum), and to carefully task them. This approach cuts the computational volume (where the N atoms live) in a large number of small cubes (typically many more cubes than there are processors), so that the short range interactions for atoms near the center of the cube can be computed within a cube, (atoms on the boundary of the cube require data from adjacent cubes as well). The authors develop a 'Task Mapping' approach, a complex optimization problem to ensure that every processor remains equally busy. Next, time integration, over a femtosecond, again a task that requires 'task mapping'. Atoms will move across cube boundaries, and must be kept track of as well. This involves communication between processors. Network latency (time per request) can be a a particular killer. To solve this requires appropriate network architecture, (3D-6D torus architecture, and the futuristic Tensor Expansion architecture), replacement of generic network software (TCP/IP) by carefully hardware-programmed switches, and a suitable quality CPU per processor. Relevant architectures are IBM's Blue Gene (200,000 processors, 500,000 Gflops) and the Red Neurons Network developed by the authors.

The research presents a preliminary analysis of the conformational structure changes of Botulinum which has about 30,000 atoms and involves more than 120,000 atoms of the solvent water.

Note: References: Y. Chen and Y. Deng, *Task mapping on supercomputers with cellular networks*, Comp. Phys. Comm., Accepted (March 28, 2008).
X. Chen and Y. Deng, *Simulations of a specific inhibitor of the disheveled PDZ domain*, J. Mol. Modeling, Submitted (Dec 13, 2007).
P. Zhang and Y. Deng, *Design and analysis of new interconnection networks for ultra-scalable supercomputers*, IEEE Transactions on Computers. (submitted May 17, 2007).
K. Oh and Y. Deng, *An efficient parallel implementation of the smooth particle mesh Ewald method for molecular dynamics simulations*, Comp. Phys. Comm. 177 (2007) 426-431.

Blow-up of Complex Solutions in Fluid Dynamics

Yakov Sinai, Princeton Univ. and Landau Inst.

email: sinai@math.princeton.edu

May 8, 2 pm

The 3-dimensional incompressible Navier Stokes equation (NS3D) is:

$$\partial_t u + u.\nabla u = \Delta u - \nabla p \quad , \tag{0.1}$$

where u is the velocity and p the pressure. The point of this presentation is to demonstrate that there exist complex-valued solutions of NS3D that blow up in finite time. This is the first time that such solutions have been constructed. The authors suggest that solutions with singularities should not be considered as turbulent. Rather they may manifest the onset of turbulence. To understand whether blowup is possible for real-valued solutions of NS3D is *the major unsolved* problem in fluid dynamics. In two dimensions blowup does not happen when initial conditions are small.

In this analysis the Fourier transform of the solution happens to be purely imaginary, the boundary conditions are absent (the fluid domain is R^3), and there is no driving force. Blow up means that both $\int |u|^2$ per unit volume (the energy) and $\int |curl\ u|^2$ per unit volume (called the 'enstrophy') become infinite at finite time. Particulary noteworthy is the form of the solution at the blow up time, at the blowup location (taken at 0 for simplicity): $u = i\frac{1}{|x|^\gamma}\phi(\frac{x}{|x|})$,with $\gamma = 6$ and ϕ a real vectorfield with complicated spherical angle dependance. The singularities of these solutions occur at points, rather than along curves.

The program to produce such solutions has these summary stages:

1): Consider the equations in the Fourier domain. Let $v(k,t)$ be the Fourier Transform of $u(x,t)$ (and do *not* assume that u is real).

$$v(k,t) = e^{-t|k|^2}v(k,0) + i\int_0^t e^{-(t-s)|k|^2} \int_{R^3} < k, v(k-k',s) > P_k v(k',s)\, dk'\, ds \quad ,$$

where the standard inner product on $\mathbb{C}^3$ is written as $< \cdot >$ and P_k is the orthogonal projection (in $\mathbb{C}^3$) to the orthogonal complement of k. The incompressibility condition takes the form $< v(k,t), k >= 0$. This evolution equation is infinite dimensional: $v(k,t)$ is determined by all other degrees of freedom k integrated over all previous times.

2): Renormalization Group approach: Solutions that blow up in the spatial domain have Fourier transforms that diverge at infinity. Determine conditions under which solutions exist and blow up at infinity (in Fourier space) at a prescribed time, along a specific spatial direction, say $\mathbf{e}$, in a specific manner, both along this direction and transverse to this direction. The (essential) domains of support (i.e. of blowup) of the solutions is bounded by a parabola of revolution and importantly, the dependence of $v(k,t)$ at one magnitude of $k \cdot \mathbf{e}$ depends only on data *closer* to 0 and this is key to have a convergent (i.e not just formal) computation.

3): Rigorously execute this, to prove existence and convergence.

The result of this program so far are the following. There exists, in the infinite dimensional space referred to, a 10-dimensional manifold $\mathcal{M}$, of purely imaginary flow fields with prescribed singularities. A certain set of 'small' initial conditions converges to this stable manifold (see the references for details). The geometric analysis of the flowfields in this manifold has begun and yields point source singularities of the above form. The understanding of the time evolution for an initial condition starting on this manifold has been completed: 4 temporally unstable dimensions, and 6 neutral dimensions.

Note: References: Dong Li, Yakov Sinai, *Blow Ups of Complex Solutions of the 3D Navier-Stokes System* (Submitted to Xarchive 13 Oct 2006) http://arxiv.org/abs/physics/0610101.
Dong Li, Yakov Sinai, *Complex Singularities of the Burgers System and Renormalization Group Methods*, in: Current Developments in Mathematics (International Press), 181-210, 2006.

Stability Analysis in Control Theory

Murat Arcak, Rensselaer Polytechnic.

email: arcakm@rpi.edu

May 15, 2 pm

Consider the system $u \to H \to y$ by which is meant an input u and an Ordinary Differential Equation with output y. For example,

$$\dot{y} = -\alpha y + \beta u \quad .$$

If we require the numbers α and β to be positive, this system is *passive*. This means that there is a smooth positive function $S : \mathbb{R} \to \mathbb{R}$ such that there is a positive γ for which $\dot{S} \leq \gamma uy$. In this case $S = \frac{1}{2}y^2$ works. The system is called Output Strictly Passive if there is $r > 0$ such that $\dot{S} \leq -ry^2 + \gamma uy$ (one of r or γ may be scaled away).

a feedback cycle of order n consists of n copies of the above system in series together with the negative feedback $y_n = -u_1$. For example, the order 2 feedback cycle of passive systems is given by:

$$\begin{aligned} \dot{y}_1 &= -\alpha_1 y_1 - \beta_1 y_2 \\ \dot{y}_2 &= -\alpha_2 y_2 + \beta_2 y_1 \end{aligned}$$

the passivity theorem states that this system is stable (if $||y(0)||$ is small then $||y(t)||$ is small for $t > 0$). the proof consists in choosing a Lyapunov function $V = \beta_2 S_1 + \beta_1 S_2$ where, as before, $S_i = \frac{1}{2}y_i^2$. The cancelation that occurs here does not happen in dimension 3 or higher, so the result does not generalize.

For a general feedback cycle C (of order n) the socalled secant criterion states that the derivative matrix of the system is Hurwitz (or: all its eigenvalues have negative real part) if $\prod_{i=1}^{n} \frac{\beta_i}{\alpha_i} < (\sec(\frac{\pi}{n})^n$. This of course implies that the system is asymptotically stable (all orbits tend to zero as $t \to \infty$).

Arcak and Sontag have proved the following extension of this result. Let C' be a nonlinear feedback cycle of order n. Suppose that the subsystems $u_i \to H_i \to y_i$ are OSP and that the linearization of C' satisfies the secant criterion. Then their results states that there is a sequence $\{d_i\}_1^n$ of positive reals so that $V = \sum d_i S_i$ is a Lyapunov function (with S_i as before). Thus such a system is stable.

Certain biochemical reactions form such a feedback cycle. Many examples come in the form:

$$\dot{x}_i = f_i(x_i) + g_i(x_i)h_{i-1}(x_{i-1}) \quad ,$$

where $i \in \{1, \cdots n\}$ and in the indices we identify $0 = n$. Suppose there is a fixed point x^*. We further assume that h_i is increasing except for $i = n$: h_n is decreasing. The subsystems given by the above equations are OSP with

$$S_i = \int_{x_i^*}^{x_i} \frac{h_i(s) - h_i(x_i^*)}{g_i(x_i^*)}\, ds \quad .$$

The secant criterion becomes $\prod \gamma_i < (\sec(\frac{\pi}{n}))^n$, where

$$\gamma_i \equiv \sup_x \frac{|\partial_x h_i|}{|\partial_x (f_i/g_i)|} \quad .$$

This last result can in turn be generalized to similar equations but now coupled along a directed graph with nodes $\{1, \cdots n\}$ and a directed edge set $\mathcal{E}$ consisting of arrows ij from i to j.

$$\dot{x}_i = f_i(x_i) + g_i(x_i) \sum_{ij \in \mathcal{E}} h_{ij}(x_j) \quad .$$

The functions h_{ij} must be either increasing or decreasing. Consider the dissipativity matrix E whose elements $E_{\ell k}$, $\ell = ij$ and $k = jm$ in $\mathcal{E}$, are non-zero only if ijm is a path in the directed graph. Set $E_{\ell\ell}$ equal to $-1/\gamma_\ell$ and non-zero off-diagonal elements $E_{\ell k}$ equal to $\mathrm{sgn}(h_\ell)$ (ℓ and k in $\mathcal{E}$). If there is a diagonal matrix D such that $E^T D + DE$ is definite negative, then the diagonal entries $\{d_i\}_1^n$ of D again define a Lyapunov function $V = \sum d_i S_i$ where S_i as before. There is no simple analytical criterion available to ascertain whether such a matrix D exists, but its numerical determination is a convex problem and can thus be done fairly conveniently.

Geometry and Rigidity

Yair Minsky, Yale University.

email: yair.minsky@yale.edu

May 22, 2 pm

Geometry and group theory meet when the geometry of a space S is studied through the group G of its isometries. Examples are: S^n whose isometries form the orthogonal group $O(n)$; $\mathbb{R}^N$ whose isometries form the Euclidean group $E(n)$, a semi-direct product of the translations $\mathbb{R}(n)$ and $O(n)$; and hyperbolic space $\mathbb{H}^n$ with the group $O(n,1)$.

A lattice Λ is discrete subgroup of G such that G/Λ has finite volume. For example in $\mathbb{R}^2$ consider the lattices $\Lambda_1 = \mathbb{Z}^2$, $\Lambda_2 = \left(\frac{\mathbb{Z}}{2}\right)^2$, and $\Lambda_3 = A\mathbb{Z}^2$ where $A = \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix}$. The fundamental domain of Λ_2 has a different area than that of Λ_1 and Λ_3 so they cannot be isometric. $\mathbb{R}^2/\Lambda_1$ has two loops of minimal length (corresponding to $[1,0]$ and $[0,1]$) whereas $\mathbb{R}^2/\Lambda_3$ has only one (namely: $[0,1]$, the next shortest loop has length $\sqrt{5/4}$). In $\mathbb{H}^2$ the family of geometrically different lattices isomorphic to a given one is studied by Teichmüller theory.

The orientation preserving isometry group of $\mathbb{H}^2$ is $PSL(2,\mathbb{R})$, and so the 'modular group' $PSL(2,\mathbb{Z})$ is a lattice in there. They are the fractional linear transformations $z \to \frac{az+b}{cz+d}$ where a, b, c, and d are integers, and $ad - bc = 1$. The group operation is given by the composition of functions. This group of transformations is isomorphic to the projective special linear group of all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a, b, c, and d are integers, and $ad - bc = 1$, and pairs of matrices A and $-A$ are considered to be identical. The group operation is the usual multiplication of matrices. The fundamental domain $\mathbb{H}^2/PSL(2,\mathbb{Z})$ is a geodesic triangle (a triangle whose sides are geodesic segments) with one corner at infinity. This domain is not compact, but it nonetheless has finite volume.

The hyperbolic case of the Mostow-Prasad Rigidity Theorem states that if Λ_1 and Λ_2 are 2 isomorphic (as groups) lattices in $O(n,1)$ and $n \geq 3$, then $\mathbb{H}^n/\Lambda_1$ is isometric to $\mathbb{H}^n/\Lambda_1$. This is *false* for lattices in $\mathbb{R}^n$ (see above examples) or lattices in $O(2,1)$.

The presentation of a group gives its generators and all relations on them that are necessary to describe the group. So $\mathbb{Z}^2 =< x, y \,|\, xyx^{-1}y^{-1} = 1 >$. The modular group is generated by the transformations $T : z \to z + 1$ and $S : z \to -1/z$. The generators S and T obey the relations $S^2 = 1$ and $(ST)^3 = 1$. It can be shown that these are a complete set of relations, so the modular group has the presentation: $< x, y \,|\, S^2 = 1, (ST)^3 = 1 >$.

A finitely generated group (that is: which has finitely many generators) can always be made to act on a geometric space by the following construction. Pick a set A of generators of Γ. Construct the Cayley graph $X(\Gamma, A)$ by assigning a vertex to each element of G. There is a directed edge from the vertex i to the vertex j if there is an a in A so that $j = ia$. We give the graph a metric by letting each edge have length one. Γ acts isometrically on $X(\Gamma, A)$ by left multiplication. This construction depends on the generators one starts with. For instance $\mathbb{Z}^2$ can be generated using $(1, 0)$ and $(0, 1)$, but also using $(1, 0)$ and $(2, 1)$.

A quasi-isometry $f : X \to Y$ between metric spaces is a map for which there are constants K and δ such that $\frac{1}{K} d(x, y) - \delta \leq d(f(x), f(y)) \leq K d(x, y) + \delta$ and Y is contained in the ϵ neighborhood of $f(X)$. (f is not required to be continuous, so if X and Y are compact every map between them is a quasi-isometry). If we take two sets A and A' of generators of a group Γ we get two graph that are quasi-isometrically related, that is: $f : X(\Gamma, A) \to X(\Gamma, A')$ and $f' : X(\Gamma, A) \to X(\Gamma, A)$, and f and f' are quasi-isometries. Thus we can look at finitely generated groups as metric spaces defined up to quasi-isometry.

One quasi-isometric invariant (or "coarse property") is the growth rate $N(r)$ of a group: the number of elements in the Cayley graph that are within a distance r of the identity. Finitely generated Abelian groups are direct sums of $\mathbb{Z}^n$ and finite cyclic groups and so their growth rate is that of $\mathbb{Z}^n$ which is polynomial. Suppose the commutator subgroup $[\Gamma, \Gamma]$ of Γ is called Γ_1. Set $\Gamma_{i+1} \equiv [\Gamma, \Gamma_i]$. Γ is called nilpotent if for some n, $\Gamma_n = \{0\}$. Nilpotent groups also have polynomial growth. Gromov has proved a reverse: If a group Γ has polynomial growth then it is virtually (up to finite index and quotient by finite subgroup) nilpotent.

Properties of the group are then reflected in the coarse properties of this metric space, such as growth rates of balls, embedded Euclidean subspaces, etc. From this point of view, which has its origins with Milnor, Svarc, and Gromov, a basic question is: what algebraic properties of the group can be detected from the coarse properties of the geometry? A group G is called quasi-isometrically rigid if whenever $X(G)$ and $X(\Gamma)$ are quasi-isometric, then G and Γ are virtually (up to finite index) isomorphic. These are the groups whose coarse geometry tells us about their algebraic properties. Theorems by Schwartz, Kleiner-Leeb and Eskin-Farb, and others assert that all lattices in semi-simple Lie groups of dimension ≥ 3 whose fundamental domain is not compact are quasi-isometrically rigid. Hamenstadt, and independently Behrstock, Kleiner, Minsky, and Mosher have proved that the Mapping Class Group (see Handel's lecture of April 4th) of a surface is quasi-isometrically rigid. Even though it is not in general a lattice it shares many features of lattices.

Slow Relaxation and Aging

Gérard Ben Arous, Courant Inst.

email: gba1@nyu.edu

May 29, 2 pm

In this talk X_1, X_2, $\cdots$, X_N are iid (see Touchette, april 24) variables, $S_N \equiv \sum_{i=1}^N X_i$, and $M_N \equiv \max\{X_i\}$. The expression $S_i \xrightarrow{\mathcal{D}} S$ denotes convergence in distribution (or "weak convergence"). This means that the probability densities μ_i associated with S_i converge to the probability density μ associated to S in the sense that for every positive continuous test function f: $\int f \, d\mu_i \to \int f \, d\mu$ (as real numbers).

Around the 1940's work by Lévy, Khinchin, and Kolmogorov had led to an essentially complete understanding of the statistics of S_N and M_N. Both quantities when appropriately rescaled so that their mean is 0 and their standard deviation 1, converge to universal distributions. These universal distributions depend on only two parameters, a shape parameter α, and a skewness parameter β. More precisely,

$$\frac{S_N - a_N}{b_N} \xrightarrow{\mathcal{D}} L(\alpha, \beta) \quad \text{and} \quad \frac{M_N - a'_N}{b'_N} \xrightarrow{\mathcal{D}} W(\alpha, \beta) \quad .$$

Reference to β will be suppressed in the remainder (the discussion will be restricted to symmetric distributions). The former are called the Lévy skew α stable (or SaS) distributions, and the latter can be one of three distributions that

are known as the Weibull, Fréchet, and Gumbel distributions. Suppose f is a distribution such that linear sums of iid variables with that distribution but different mean and variance, is a new random variable with the *same distribution* (but with possibly different mean and variance). Then f must be an SaS distribution (this is the meaning of the word 'stable' here). The normal distribution corresponds to $\alpha = 2$.

Suppose now that we have random variables X_i such that the probability $P(X_i \geq x)$ is proportional to $\frac{K}{x^\alpha}$ where $\alpha \in (0, 1)$. In particular, this distribution does not have a mean (the integral does not converge). Even in that case the following holds:

$$\frac{S_N}{N^{1/\alpha}} \xrightarrow{\mathcal{D}} L(\alpha) \quad \text{and} \quad \frac{M_N}{N^{1/\alpha}} \xrightarrow{\mathcal{D}} W(\alpha) \quad .$$

Consider the function that maps $\frac{i}{N}$ to $\frac{S_i}{N^{1/\alpha}}$ for $i \in \{1, \cdots N\}$. The limit as N tends to infinity of this function corresponds to a random variable Y_t where t ranges from 0 to 1 which is called the α stable subordinator. It has the curious property that the *range* of that variable is fractal. To be precise: the probability that

$$\{\cup_{t\in[0,1]} \{Y_t\}\} \cap [a, b] = \emptyset$$

equals $F_\alpha(\frac{a}{b})$ where F_α is a smooth function. For $\alpha = 1/2$ this leads to Lévy's arcsine law.

In spin glass models a configuration $\sigma = (\sigma_1, \cdots \sigma_N)$ represents the value of each spin $\sigma_i \in \{-1, +1\}$. The phase space of these configurations form an N dimensional cube. The Hamiltonian is an energy function ('landscape') defined on this cube:

$$H(\sigma) = \frac{1}{\sqrt{N}} \Sigma_{ij} J_{ij} \sigma_i \sigma_j \quad ,$$

where for simplicity J_{ij} are iid normally distributed random variables with mean 0 and variance 1. This system evolves according to Glauber dynamics: at each clock tick a random spin is flipped if it lowers the energy, and flipped with probability p if it doesn't. Thus σ is a function of the number t of clock ticks. Then under some conditions, $e^{\beta\sqrt{N}H(\sigma(t))}$ converges in distribution to the α stable subordinate. The main condition is that t is large, but still very small compared to the time needed to approximate equilibrium. Under the same condition for t and s the probability that $\sigma(t) \approx \sigma(t + s)$ is approximately $F_\alpha(\frac{t}{t+s})$. Thus the system *decorrelates* more slowly as it ages.

In this (and other) complex high dimensional systems where the time to reach equilibrium is very long (slow relaxation), these result mean that the way in which equilibrium is reached is universal, even if the form of the equilibrium distribution itself is not. In the above example the equilibrium distribution is given by the Gibbs measure $e^{-H(\sigma)}/Z$ where Z is a normalizer, and so depends on the details of H.

Accuracy of One-Bit Quantization, Fair Duels, Tilings

Sinan Güntürk, Courant Inst.

email: gunturk@cims.nyu.edu

June 5, 2 pm

Consider the problem of finding a function (or "digits") $q : \mathbb{N} \to \{-1, +1\}$ so that the *bias*

$$B_q(\epsilon) \equiv \epsilon \sum_{i=0}^{\infty} (1 - \epsilon)^n q(n)$$

vanishes as rapidly as possible as ϵ approaches 0. The bit sequence q forms a one bit quantization of the signal 0. The theorem here is that there are q such that $B_q(\epsilon) = \mathcal{O}(e^{-c/\epsilon})$ (ϵ varies, q is fixed) and that this is optimal. (The best value for the constant c is unknown.) This result can be used to prove that for any power series $f = \sum a_i z^i$ with coefficients $a_i \in [-1, 1]$ can be approximated by a one bit quantization $Q(z) = \sum_i q_i z^i$ so that $|f(z) - Q(z)| = \mathcal{O}(e^{-c/|1-z|})$ on a sub-domain of the unit disk containing the real segment $[0, 1]$ (a "Stolz angle" to be precise).

The name 'fair duel' arises from the following reformulation. Suppose two random shooters face off in a duel where each shot has probability ϵ of killing the opponent (and is independent of the previous shots). The sequence q determines the shooting order: '+1' stands for one shooter and '-1' for the other. After shots 0 through $n-1$ the probability that both shooters are alive equals $(1-\epsilon)^n$. If that is the case, if $q(n)=+1$ one shooter has a chance ϵ to kill the other, and if $q(n)=-1$ the other does. Thus $(1-\epsilon)^n\epsilon q(n$ equals the probability that the "+"-shooter kills the other minus the probability that the "-"-shooter kills his opponent. The bias function equals the probability that one shooter survives minus the probability that the other does.

Suppose that $q(n)$ are coin tosses, then $< B_q(\epsilon) > \asymp \sqrt{\epsilon}$, where $\asymp$ means that the quotient of the two function is uniformly bounded away from both 0 and ∞. Not a good result. To look at biases for other sequences define the generating function $F_q = \sum_i q(i)z^i$ associated to q. If for a given a closed sum for F_q can be calculated then of course $B_q(\epsilon) = \epsilon F_q(1-\epsilon)$ can be computed. Here are a few cases.

Digits q	Generating function F_q	Bias B_q
$(+1)^\infty$	$\frac{1}{1-z}$	1
$(+1-1)^\infty$	$\frac{1-z}{1-z^2} = \frac{1}{1+z}$	$\asymp \epsilon$
$(+1-1-1+1)^\infty$	$\frac{(1-z)(1-z^2)}{1-z^4} = \frac{1-z}{1+z^2}$	$\asymp \epsilon^2$
$(+1-1-1+1-1+1+1-1)^\infty$	$\frac{(1-z)(1-z^2)(1-z^4)}{1-z^8} = \frac{(1-z)(1-z^2)}{1+z^4}$	$\asymp \epsilon^3$

The multiplicity of the root at $z=1$ is decisive. So for the limiting sequence, the Thue-Morse sequence t, one obtains: $F_t(z) = \prod(1-z^{2^n})$ and *its* bias will be better than polynomial.

Denote by Sq the sequence of partial sums of q, so that $q_n = (Sq)(n) - (Sq)(n-1)$. The last relation implies that $F_q(z) = (1-z)F_{Sq}(z)$. Iterating this leads to $B_q(\epsilon) = \epsilon^{k+1}F_{S^kq}(1-\epsilon)$ and thus

$$|B_q(\epsilon)| \leq \frac{\epsilon^{k+1}}{1-|1-\epsilon|}\|F_{S^kq}\|_\infty = \epsilon^k\|F_{S^kq}\|_\infty \quad ,$$

where $\|F_w\|_\infty$ means $\max_n |w_n|$. A combinatorial argument yields that for the Thue-Morse sequence t: $\|S^kt\|_\infty = 2^{\binom{k-1}{2}} \approx 2^{0.5k^2}$. Hence

$$|B_q(\epsilon)| \leq \inf_k \epsilon^k\|S^kt\|_\infty = \inf_k \epsilon^k 2^{k^2/2} = e^{k\ln\epsilon + k^2\frac{\ln 2}{2}} \quad .$$

Minimizing the exponent gives that the bias is $\mathcal{O}(-c(\ln\epsilon)^2)$. This is better than polynomial, although not optimal. The optimal estimate follows from a recursive equation that is inspired by sigma-delta modulation described next.

Consider a sampled signal $\{x(n)\}_n$ where $x(n) \in [0,1]$. Define $q(n) \equiv \lfloor (Sx)(n)\rfloor - \lfloor (Sx)(n-1)\rfloor$. Then $(Sq)(n) = \lfloor (Sx)(n)\rfloor$. The accumulated error is $u(n) \equiv (Sx)(n) - (Sq)(n)$. Clearly,

$$\Delta u(n) \equiv u(n) - u(n-1) = x(n) - q(n) \quad . \tag{0.2}$$

The error estimate is of order N^{-1}: $\left|\frac{1}{N}\sum_{n=k+1}^{k+N}(x(n)-q(n))\right| \leq \frac{1}{N}$. However, as shown above, it is advantageous to choose $q(n)$ such that Equation (0.2) can be replaced by

$$\Delta^m u(n) \equiv u(n) - u(n-1) = x(n) - q(n) \quad , \tag{0.3}$$

for some positive integer m as long as the sequence $u(n)$ is bounded. Thus we redefine the digits $q(n)$ as follows.

First define $\hat{u}(n) \equiv (u_1(n), \cdots u_m(n)^T$ and reinterpret Equation (0.3) as:

$$\begin{aligned} u_1(n) &= u_1(n-1) + x(n) - q(n) \\ u_2(n) &= u_1(n-1) + u_2(n-1) + x(n) - q(n) \\ &\vdots \\ u_m(n) &= \sum_{i=1}^m u_i(n-1) + x(n) - q(n) \end{aligned}$$

For each digit $q(n)$ this gives an affine transformation $T : \mathbb{R}^m \to \mathbb{R}^m$, given by

$$T : \hat{u}(n) = L\hat{u}(n-1) + (x(n) - q(n))1 \quad ,$$

where L is the lower triangular of only ones (including the diagonal) and 1 is the vector of only ones.

Consider again the case where $x(n) = x$ is constant. The plan is to construct digits q so that the sequence $\|\hat{u}(n)\|$ is bounded. One method is to partition $\mathbb{R}^m$ into two domains Ω_- and Ω_+ so that $\hat{u}(n-1) \in \Omega_\pm$ then $q(n) = \pm 1$. It is an open problem which partition lead to bounded sequences $\|\hat{u}(n)\|$. However, if one chooses such a partition (and if some other conditions are satisfied), then there is an open invariant set $\Gamma \subset \mathbb{R}^m$ of T. On this set the map T commutes with its projection to the standard torus and from this one concludes (using the Ergodic Theorem) that the invariant set Γ "tiles" $\mathbb{R}^m$. Another open problem in this case is to know the dimension of its boundary, since larger dimension leads larger computational errors.

Note: This talk lasted 2 hours.

Knot Theory and the Lorenz Attractor

Ilya Kofman, College of Staten Island (CUNY).

email: kofman@math.csi.cuny.edu

June 12, 2 pm

For certain parameter values the Lorenz Equation are known to give rise to a strange attractor.

$$\dot{x} = 10(y - x) \qquad \dot{y} = 28x - y - xz \qquad \dot{z} = xy - \frac{8}{3}z$$

is the most famous case. All orbits are asymptotic to the attractor which itself has fractal dimension. On the attractor itself the behavior is sensitive to initial conditions.

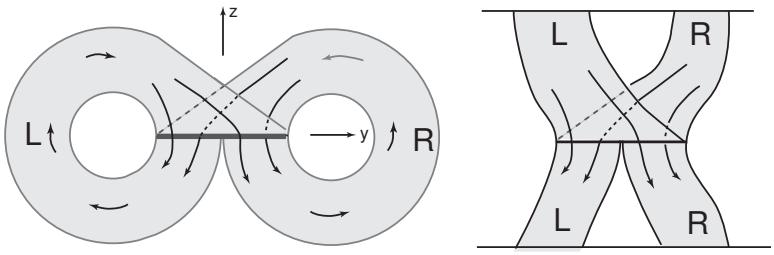

Figure 0.1: *In the first figure, the Lorenz template with the flow indicated by arrows; in the second, the construction of the braid-diagram.*

Guckenheimer and Williams constructed the so-called Lorenz template in 1979. This is a branched surface embedded in $\mathbb{R}^3$ together with a smooth flow and serves as a geometrical model for the flow on the attractor. All forward orbits must intersect the branch cut ℓ after which they go round either the Left "ear" or the Right "ear" before returning to ℓ (see Figure 0.1). Parametrizing ℓ by the interval $[0, 1]$, the Poincaré map $P : \ell \longrightarrow \ell$ defined by the flow (in this geometric model) is $x \longrightarrow 2x \mod 1$. Birman and Williams (1983) gave a topological description of the orbits on the templates as knots (see below) satisfying certain restrictions which makes them rather rare. Out of the approximately 1.7 million simplest knots, only 20 occur as knots on the Lorenz template. Ghrist (1995) constructed a flow on S^3 that contains *all* knots. Only recently (Tucker, 2002) has proved (using rigorous interval arithmetic) that this geometric model indeed gives an accurate description of the dynamics.

A link is a 1-dimensional subspace of the 3-sphere (or S^3) whose connected components are homeomorphic to circles. A knot is a link with one component. Two links are equivalent if they are isotopic (roughly speaking, one equals the other up to continuous deformation). Given a link L, a Seifert surface is an oriented surface $S \subset S^3$ whose

boundary equals L. A link $L \subset S^3$ is called *fibered* if there is a one parameter set of homeomorphic Seifert surfaces S_θ whose boundary equals L, where θ runs through the unit circle, and such that for distinct θ the corresponding Seifert surfaces intersect precisely in their boundary L, and so that $\cup S_\theta = S^3$. Remove a very small segment from each of two knots A and B. Then reconnect the 4 strands so that 1 new knot is obtained. Such a knot is called a composite knot. A prime knot is one that is not composite.

The restrictions Birman and Williams proved (1983) were that every knot occurring in the template must be fibered and prime. They observed that it follows immediately from the template that every knot in it has the property that all crossings are *positive*. To explain the last statement we need to look at braids.

A link L can be deformed in such a way that it consists of of two rows of horizontally aligned strands numbered from 1 through k. (Imagine looking at the link from 'above'.) Connect the strands in such a way that every crossing appears between the two rows. This is the *braid diagram* associated to the link. The link itself is thought of as the 'closure' of this diagram: connect the two loose ends of strands i for each i in 1 through k (this can be done without introducing new crossings). Of course (see Figure 0.1) any periodic orbit on the template automatically gives rise to such a diagram. Orient the orbits in the template so that they *descend* as they cross the branch-cut. A crucial **observation** is the following: *orbits descending from the Left never cross each other and neither do orbits descending from the Right. Orbits descending from the Right always* overcross *orbits descending from the Right.* This explains the 'positivity' of the crossings observed by Birman and Williams. (At the crossing, let e_- be the tangent vector in $\mathbb{R}^2$ in the direction of the flow along the strand that undercrosses and e_+ along the strand that overcrosses. The parity or handed-ness of the basis (e_-, e_+) of $\mathbb{R}^2$ is called the index of the crossing. The sum of these indices along a knot is called the *writhe*.) Knots with the property that all crossings positive are called *positive*.

A positive knot (or link) is uniquely determined by the permutation on the two rows of strands in the braid diagram. Thus they can be represented as compositions of σ_i, the exchange of strands i and $i+1$. For instance the (p,q)-torus link can be made from a closed braid with p strands. The appropriate 'braid word' is $(\sigma_1 \cdots \sigma_{p-1})^q$. A "T-link" is defined as a certain concatenation of torus links: namely $(\sigma_1 \cdots \sigma_{p_1-1})^{q_1} (\sigma_1 \cdots \sigma_{p_2-1})^{q_2} \cdots (\sigma_1 \cdots \sigma_{p_k-1})^{q_k}$, where p_i is strictly increasing. On the other hand, above observation implies that any braid of the template is determined by the non-decreasing sequence of positive integers $< j_1, j_2, \cdots j_r >$ that informs that strand i jumps to $i + j_i$. Recently (2008) Birman and Kofman have proved that every knot on the Lorenz template is a T-link, that all T-links occur as links on the template, and that $(\sigma_1 \cdots \sigma_{p_1-1})^{q_1} (\sigma_1 \cdots \sigma_{p_2-1})^{q_2} \cdots (\sigma_1 \cdots \sigma_{p_k-1})^{q_k} =< p_1^{q_1} \cdots p_k^{q_k} >$.

Ghys established in 2006 that there is a one-to-one correspondence between modular knots and the knots that can occur on the Lorenz template. This fascinating story is beyond the scope of this summary but can be found in http://www.ams.org/featurecolumn/archive/lorenz.html.

Yang-Lee Zeros for Diamond Lattices and 2D Rational Dynamics.

Misha Lyubich, University of Toronto and Suny at Stony Brook.

email: mlyubich@math.sunysb.edu

June 19, 2 pm

Consider a finite graph Γ_n, with a set v of n vertices. In the Ising model a spin configuration is a function $\sigma : V \to \{-1, +1\}$. Define the energy or Hamiltonian $H(\sigma)$ as:

$$H(\sigma) = -\frac{J}{2} \sum_{v,neigbrs\, w} \sigma(v)\sigma(w) - \frac{h}{2} \sum_{v,neigbrs\, w} (\sigma(v) + \sigma(w)) \quad ,$$

where J (the magnetic interaction) and H (the external field) are fixed parameters. Given a temperature T, define the weight of the configuration σ, $W(\sigma) = e^{-H(\sigma)/T}$, and define the partition function of the graph Γ_n as $Z_n = \sum_\sigma W(\sigma)$. Define further $z = e^{-h/T}$ and $t = e^{-J/T}$ (z and $t \in (0,1)$ when $T \in (0, \infty)$), so that now $Z = Z_n$ is a Laurent polynomial $Z = P(z, \sqrt{t})$. Distribution of the zeros of these polynomials is physically important as they control phase transitions in the model.

Theorem: (Lee and Yang, 1950's) *Assume that $J > 0$ (ferromagnetic case), and given any value $0 < t < 1$, then, irrespective of the underlying graph, the zeros z of $P(z, \sqrt{t})$ lie on the unit circle.*

For any fixed polynomial P, the locus of the zeros of P is contained in the cylinder with the unit circle z-direction horizontal and the t direction vertical. These zeros can be represented as a set of curves parametrized by $t \in [0, 1]$, connecting the top to the bottom: Each of these curves is also a graph over t, and the number of these curves equals **d**. These curves can also be thought of a 'pre-foliation', with finitely many, namely **d**, leaves, and these leaves can therefore be thought of as a transversal measure μ. Thus a graph Γ_n produces a partition function Z and a transverse measure μ. This measure is singular for a finite graph because it is concentrated on the leaves. However in the limit as $n \to \infty$ one can think of this measure as having a density ρ parametrized by the angle in the z direction θ: $\mu' = \rho(\theta) d\theta$. One can think of a **sequence** of graphs Γ_n and therefore **sequences** of partition functions Z_n and transversal measures μ_n, the latter converging to the measure μ.

The authors consider the Migdal-Kadanoff hierarchical model: Γ_1 is a horizontal graph with two vertices connected by an edge. Γ_2, a diamond shaped graph, with 4 vertices. In general Γ_n is obtained by Γ_{n-1} by replacing every edge in Γ_{n-1} by a copy of Γ_2. The spin configurations on Γ_1 are $++$, $--$, $-+$, and $+-$. Let $U_1 := W(++) = \dfrac{1}{z_1\sqrt{t_1}}$, $V_1 := W(-+) = \sqrt{t_1}$, $W_1 := W(--) = \dfrac{z_1}{\sqrt{t_1}}$. Since, by symmetry, $V_1 = W(+-)$, therefore $Z_1 = Z = U_1 + 2V_1 + W_1$. For Γ_n, define the variables U_n, V_n, Z_n in the corresponding manner. Precisely, let $+*+$ denote the set of all spin configurations that have a $+$ at the left most point, and a $+$ on the right most point. Define $U_n := Z_n(+*+)$. Define similarly $V_n := Z_n(-*+)$ and $W_n := Z_n(-*-)$. Then again, the partition function Z_n equals $U_n + 2V_n + W_n$.

The Migdal-Kadanoff renormalization group equation, or RGE, expresses U_{n+1}, V_{n+1}, and W_{n+1} in terms of U_n, V_n, and W_n. The RGE can also be formulated as a transformation from (z_n, t_n) to the (z_{n+1}, t_{n+1}) variables. Here (z_n, t_n) is defined by: $U_n := \dfrac{1}{z_n\sqrt{t_n}}$, $V_n = \sqrt{t_n}$. This relation has the following form:

$$(z_{n+1}, t_{n+1}) \equiv R(z_n, t_n) = \left(\frac{z_n^2 + t_n^2}{z_n^{-2} + t_n^2}, \frac{z_n^2 + z_n^{-2} + 2}{z_n^2 + z_n^{-2} + t_n^2 + t_n^{-2}} \right)$$

Since the transformation R preserves the cylinder described before, its restriction to the cylinder can be studied. Now $R(e^{i\theta}, 0) = e^{4i\theta}$, which has degree 4, while $R(e^{i\theta}, 1) = e^{2i\theta}$ has degree 2.

Consider now the locus γ_n in the cylinder of the zeroes of Z_n and construct them as a sequence of pullbacks. Thus γ_1 are the zeroes of $z_1^2 + 2t_1 z_1 + 1 = 0$. Representing $z_1 = e^{i\theta}$, this curve is given by $t_1 = -2cos(\theta)$ and has two monotone 'branches' connecting $t = 0$ with $t = 1$, one going 'up' from $t =$ to $t = 1$, the other down from $t = 1$ to $t = 0$. γ_2, the is the pullback $\gamma_2 = R^*(\gamma_1)$, defined as the zeroes of $z_2^2 + 2t_2 z_2 + 1 = 0$, where z_2 and t_2 are functions of z_1 t_1 given by RGE. γ_2 consists of two groupings of 4 curves. The 4 curves in each grouping are monotone up, but meet at two points on $t = 1$. One then keeps iterating: $\gamma_n = R^*(\gamma_{n-1})$.

Theorem: (Bleher-Lyubich-Roeder). *The set γ_n consists of 2.4^{n-1} vertical branches. They converge to a vertical foliation $F = \lim[\gamma_n]$. The sequence of measures μ_n also has a limit μ. There exists a $0 < t_{crit} < 1$ so that*
i): The foliation F is real analytic for $t < t_{crit}$, and the measure μ in the same range of t is absolutely continuous in the θ variable.
ii): For $t_{crit} < t < 1$ the leaves of F are only dense; the measure μ is still absolutely continuous in the θ variable but its density vanishes on a Cantor set of positive length (and is real analytic on the complementary open dense set).
iii): When $t = 1$, the measure μ on this level is singular with support on the inverse images under R of $(z, t) = (-1, 1)$, the dyadic rationals.