# Analysis Seminar

Jared Marx-Kuo

Oct. 4th - Dec. 6th, 2019

## 10/4: Sanchit's Talk on "Boltzmann's H-theorem & Why kintetic theory is so hard"

1. We have
$$\partial_t f + v_i \partial_{x_i} f = Q(f,f)$$
where
$$Q(f,f) = \int_{\mathbb{R}^3} dv_* \int_{S^2} B(v - v_*, \sigma)(f|_{v'} f|_{v'_*} - f|_v f|_{v_*})$$

2. Here the set up is we have two particles with initial velocities $\vec{v}$ and $\vec{v}_*$. They then collide and leave with velocities $\vec{v}'$ and $\vec{v_*}'$. In paritcular
$$v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2}\sigma, \quad v'_* = \frac{v + v_*}{2} - \frac{|v - v_*|}{2}\sigma$$
via some laws of conservation of momentum and energy.

3. We further have
$$\sin\theta B(v - v_*, \sigma) \cong \langle v - v_* \rangle^\gamma \theta^{2-2s} = \sqrt{1 + (v - v_*)_i^2}$$

4. **Theorem:** (H-theorem): (negative) Entropy is decreasing
$$\int_x \int_v \log t[\partial_t f + v_i \partial_{x_i} f] = \int_x \int_v Q(f,f)$$
this gives us
$$\frac{d}{dt} H(f(t,\cdot,\cdot)) = -\int D(f,c,\cdot)dx \leq 0$$
where
$$D = \int dv dv_* d\sigma B(f' f'_* - f f_*) \log\left(\frac{f' f'_*}{f f_*}\right)$$
and
$$H = \int f \log f$$
and $f'$ denotes $f\Big|_{v'}$ and the like....

5. Consider the map
$$(x,y) \mapsto (x - y)(\log x - \log y)$$
which is increasing, so that the integrand in the expression for $D$ is positive

6. Somehow this tells us that
$$f' f'_* = f f_* \qquad f \cong e^{-|v - u(x)|^2}$$
Boltzmann proved that
$$\lim_{t \to \infty} f(x) = M$$
where $x$ depends on $t$, and $M$ represents a "travelling maxwellian"

7. This comes from the transport equation

$$\partial_t f + v_i \partial_{x_i} f = 0, \quad f\Big|_{t=0} = f_0(x,v)$$
$$f(t,x,v) = f_0(x - tv, v)$$

8. We have the following properties:

$$||f(t,x,v)||_{L_x^\infty L_v^\infty} = ||f_0||_{L_x^\infty L_v^\infty}$$
$$\int f dV = \int f_0(x - tv, v) dV \leq f_0(1+t)^{-3}$$

9. Assume our space is $\pi_x^3$, i.e. a 3-torus. We expect $f\Big|_{t=\infty} = M$, but we cannot prove this. However, if we begin close enough to this travelling Maxwell distribution, $M$, i.e. $f = M + \epsilon_{jm}$, then there is a convergence of $f \to M$ at $\exp^{-\lambda t}$ rate

10. Spatially Homogenous case:

$$\partial_t f = \overline{a}_{ij} \partial^2_{v_i v_j} f - \overline{c} f$$

Here, one uses entropic dissipation. We have that

$$\overline{a}_{ij} = a_{ij} * f$$
$$a_{ij}(z) = \left( S_{ij} - \frac{z_i z_j}{|z|^2} \right) |z|^{2+\gamma}, \quad \forall \gamma \in [-3, 1]$$
$$c = \partial^2_{v_i v_j} a$$

11. The general idea is to get a bound of this form

$$-\frac{d}{dt} H(t) \geq \theta(H(t))$$

for $\theta$ a positive function. Then we'll get that

$$H(f) \leq e^{-\theta t} H(f_0)$$

12. **Theorem:** (Fokker Planck)

$$\partial_t f = \nabla_v \cdot (\nabla_v f + f v)$$

we want to consider

$$H(f\Big|M) = \int f \log(f/M) \qquad I(f\Big|M) = \int f |\nabla_v \log(f/M)|^2$$

these are important to get exponential convergence.

13. Log-Sobolev tells us that

$$I(f\Big|M) \geq 2H(f\Big|M) \qquad \frac{d}{dt} H = -I(f\Big|M)$$

and so

$$-\frac{d}{dt} H \geq 2H(f) \implies H(f) \leq e^{-2t} H(f_0)$$

by gronwall's inequality

14. Open questions:

    (a) Local existence for spatially inhomogenous Boltzmann for some physical regimes

    (b) Global existence and regularity questions for large data

    (c) Does $f$ converge to Maxwellian for bounded domain?

## 10/11: Felipe Hernandez's talk on "Quantum Information Theory for the Analyst"

1. We have $H = \mathbb{C}^2 = \text{span}\{|0>, |1>\}$ hilbert
2. Ex:
$$v = \frac{1}{\sqrt{2}}|0> + \frac{1}{\sqrt{2}}|1>, \qquad w = \frac{1}{\sqrt{2}}|0> - \frac{1}{\sqrt{2}}|1>$$
then probability that $|0>$ and $|1>$ is each $1/2$ in both states
3. Evolution: Occurs by applying a unitary matrix. As an example
$$U = 1/\sqrt{2}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix} \implies Uv = |1> \qquad Uw = |0> w$$
4. Observables
    (a) Self-adjoint operators $A$, $v^*AV$ is some expectation
$$e.g. \qquad A = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \implies v^*Av = \frac{1}{2}$$
    for $v$ as before
5. What is "$|0>$ with $p = 1/2$ and $|1>$ with $p = 1/2$? I.e. can get something that yields these probabilities, but in a way that its invariant under unitary matrices $U$?
6. Solution: Expand our space from the hilbert space to a collection of **density matrices**
Instead of $v$, think about $vv^*$
7. Time evolution: $v \mapsto Uv$, and $vv^* \mapsto Uvv^*U$
8. Observable: $v \mapsto v^*Av$ and $vv^* \mapsto \text{tr}(Avv^*)$
9. Note: $vv^*$ is
    (a) positive semi-definite
    (b) Trace equal to 1
    (c) Self-adjoint
10. Both of the above properties are preserved under taking convex combinations
11. Ex: "$|0>$ with $p = 1/2$ and $|1>$ with $p = 1/2$ is given by the density matrix
$$\frac{1}{2}|0><0| + \frac{1}{2}|1><1| = \begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix}$$
the above is a mixed state, i.e. it satisfies 1-3. Note that this is not a pure state, i.e. not equal to $vv^*$ for some wave state $v$. This is because all such matrices $vv^*$ are rank 1, while the above is rank 2. In fact, any such matrix which is a mixed state and is rank 1 is a pure state, i.e. arises as $vv^*$ for some $v$ a wave state
12. As an aside, note that for
$$v = \frac{1}{\sqrt{2}}|0> + \frac{1}{\sqrt{2}}|1> \implies vv^* = \begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$$
13. If $\rho$ is a density matrix, then spectral theorem tell us that
$$\rho = \sum_i \lambda_i v_i v_i^* \qquad 0 \le \lambda_i \le 1, \qquad \sum_i \lambda_i = 1$$
and $\{v_i\}$ the eigenvectors of the system. It makes sense to defien
$$S(\rho) = \sum_i -\lambda_i \log \lambda_i = \text{tr}(\rho \log \rho)$$

14. Quantum Channels

   (a) Ex: $\mathcal{U}[\rho] = U\rho U^*$

   (b) In general, we want

$$\mathcal{O} : B(H_1) \to B(H_2)$$

   such that the above operator

      i. preserves positive semi-definiteness

      ii. Preserves trace

   (c) $\mathcal{O}$ is called a completely positive trace-preserving operator

   (d) Ex: Measurement

$$M[\rho] = |0><0| <0|\rho|0> + |1><1| <1|\rho|1> \implies \text{tr}(M[\rho]) = \text{tr}(\rho)$$

   (e) Example:

$$v = \frac{1}{\sqrt{2}}|0> + \frac{1}{\sqrt{2}}|1>$$

$$m = \frac{1}{\sqrt{2}}|0> \otimes |\text{device says "0"}> + \frac{1}{\sqrt{2}}|1> \otimes |\text{device says "1"}>$$

   Then

$$\rho_m = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} = mm^*$$

   where the rows and columns are labelled by 00, 01, 10, and 11.

   (f) The above still has 0 entropy because all of its eigenvalues are 0 or 1

   (g) To get non-zero entropy, we look at **Partial Trace**

      i. The follow set up

$$H = H_1 \otimes H_2 \qquad A \in B(H_1), \quad A_2 \in B(H_2)$$
$$\text{tr}_{H_2}(A_1 \otimes A_2) = A_1 \text{tr}(H_2)$$

      ii. In our previous example, if we have

$$\text{tr}_{H_2}(\rho_m) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

      From the above, we can think of measurement as an operation that preserves the system and then a partial trace

      iii. Note: Partial traces are monotone w.r.t. the entropy, and so for any quantum channel, entropy is always increasing!!!!!

   (h) Part 2: For Analysts

      i. Choose $r > 0$ ($r$ is large and $\to \infty$), and a function

$$\chi \in C_c^\infty(\mathbb{R}^d) \text{ s.t. } ||\chi||_{L^2} = 1 \quad supp(\chi) \subseteq B_1, \quad \text{"supp } \hat{\chi} \subseteq B_1\text{"}$$

      The last statement means that

$$|\hat{\chi}(p)| \leq C(1 + |p|)^{-100^1 00}$$

      i.e. $\hat{\chi}$ gets real fucking small outside of the unit ball

      ii. Now for $(x_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$, then

$$\phi_{x_0,p_0}(x) = r^{-d/2} e^{ip_0 \cdot x} \chi\left(\frac{x - x_0}{r}\right) = |x_0, p_0> = |\xi_0> \qquad \text{s.t. } \xi_0 = (x_0, p_0)$$

      The Hilbert space we're working in is $H = L^2(\mathbb{R}^d)$ but the above collection of functions is not a basis

iii. Cool thing

$$Id = \int_{\mathbb{R}^d} |\xi >< \xi| d\xi$$

the left hand side can be thought of as an integral of operators mapping from $L^2(\mathbb{R}^d) \otimes L^2(\mathbb{R}^d) \to \mathbb{R}$. The above holds because

$$\langle f, g \rangle = \int \langle f, \phi_\xi \rangle \langle \phi_\xi, g \rangle d\xi$$

And when you do the quadruple integral out, it all works out

iv. Fact: $\langle \xi | \xi \rangle = 1$

v. Fact: $||f||^2 = \int |\langle f | \xi \rangle|^2 d\xi$

vi. Fact: $\langle \xi | \eta \rangle \cong 0$ unless $|x-y| \le Cr$ and $|p-q| \le Kr^{-1}$, where $\xi = (x,p)$ and $\eta = (y,q)$

vii. If $\rho \in B(L^2(\mathbb{R}^d))$ (bounded maps from $H$ to itself?), then

$$M[\rho] = \int |\xi >< \xi| < \xi | \rho | \xi > d\xi$$

viii. **Question**: When is $M[\rho] \cong \rho$?
**Answer**: If $\rho = \int F(\eta) |\eta > d\eta$ and $F$ is smooth.
**Proof:**

$$\int F(\xi) | < \eta | \xi > |^2 |\xi >< \xi| d\xi \approx \rho$$

some how, maybe because the $| < \eta | \xi > |^2$ is about 1 somewhere of interest

(i) Why we care:

i. Want to study the wave equation

$$i\partial_t \psi + \frac{1}{2}\Delta\psi = \epsilon V \psi$$

for $V$ a random weak potential

ii. We have that $H = -\frac{1}{2}\Delta + \epsilon V$, then $e^{itH}\psi_0$ is a random linear combination of wavepackets

iii. Idea: $\rho_0$ is a mixture of wavepackets. Need 2 facts

A. $\mathcal{U}_t[\rho_0] = e^{itH}\rho_0 e^{-itH}$

B. Then we look at

$$\mathcal{U}_t \circ \cdots \circ \mathcal{U}_t[\rho_0]$$

and $\mathcal{U}_t[M[\rho]]$ is easy to work with.

We want to replace

$$\mathcal{U}_t \circ \cdots \circ \mathcal{U}_t[\rho_0]$$

with

$$\mathcal{U}_t \circ M \circ \mathcal{U}_t \cdots \circ M \circ \mathcal{U}_t[\rho_0]$$

## 10/18 Kevin Yang's Talk on "Logarithmic Sobolev Inequalities"

1. In this talk $\mu_H(dx) = Z_H^{-1} e^{-H(x)} dx$ where

$$Z_H = \int_{\mathbb{R}^N} e^{-H(x)} dx$$

2. In this talk, we assume $H \in C^2$ and $D^2 H \ge k > 0$ (this is what "convexity" means) where $D^2 H$ is the Hessian of $H$

3. Ex: $H(x) = \frac{ax^2}{2}$ and $\mu_H$ is a Gaussian with variance $1/a$.

4. Note: can be extended to closed manifolds.

5. Here: convexity = "compactness"

6. **Definition:** (Relative entropy/ KL divergence)

$$H(f) = \int_{\mathbb{R}^n} f \log f d\mu \qquad f \in L^1_{d\mu}, \quad f \geq 0, \quad \int f d\mu = 1$$

7. **Lemma:** under these assumptions, $H(f) \geq 0$ and $H(f) = 0 \iff f \equiv 1$.
   **Proof:** Use convexity of $x \mapsto x \log x$ and apply Jensen's inequality

8. **Lemma:** The following two equations holds

   (a) $2||f-1||_{L^1} \leq \sqrt{H(f)}$ (Pinsker inequality)

   (b) For $p > 1$, we have $H(f) \leq 2||f-1||_{L^p} + \frac{2}{p-1}||f-1||^p_{L^p}$

   (c) $H(f) = \sup_{\varphi \in L^\infty}\{\int f\varphi - \log \int e^\varphi d\mu\}$

   The third thing tells us

   $$\int f\varphi d\mu \leq H(f) + \log \int e^\varphi d\mu$$

   Note that $||f-1||_{L^1}$ is a measure of the variation of $f$ from the equilibrium measure.

9. **Definition:** The dirichlet energy/form of $f$ is

   $$D(f) = \int_{\mathbb{R}^n} |\nabla f|^2 d\mu$$

   Here $d\mu$ is not just the lebesgue measure, because it must be normalized. Note: if $\mu(dx) = dx$, then

   $$D(f) = -\int_{\mathbb{R}^n} f\Delta f dx$$

10. **Definition:** (Fisher Information) For $f \geq 0$

    $$\overline{D}(f) := D(\sqrt{f}) = \int_{\mathbb{R}^n} |\nabla \sqrt{f}|^2 d\mu = \int \frac{|\nabla f|^2}{f} d\mu$$

11. Operator

    $$L_H(f) = \Delta f - \nabla H \cdot \nabla f$$

    where $H$ is our hamiltonian.

12. Check:

    $$D(f) = -\int f L_H f d\mu$$

13. Note: $L_H$ is symmetric w.r.t. $\mu$ so

    $$\int f(L_H g) d\mu = \int (L_H f) g$$

14. Dynamic equation

    $$\partial_T u(T,x) = L_H u(T,x)$$

    Ex: If $H(x) = ax^2/2$, then $L_H$ is a generator of an Orstein-Ulumbeck process

15. The relative entropy is some measure between $f d\mu$ and $d\mu$

16. Entropy Production: $u(T,x)$ solves $(\partial_T - L_H)u(T,x) = 0$

17. **Lemma:** If $u(0,x) \geq 0$, $\int_{\mathbb{R}^n} u(0,x)d\mu = 1$ then a) $u(T,x) \geq 0$, b) $\int_{\mathbb{R}^N} u(T,x)d\mu = 1$.
    **Proof:** a) follows by the maximum principle, b) follows by differntiating under the integral

    $$\partial_T \int u(T,x)d\mu = \int L_H u(T,x)d\mu = 0$$

    but now note that $L_H$ is symmetric, so we could write

    $$\int L_H u(T,x)d\mu = \int L_H u(T,x) \cdot 1 d\mu = \int u(T,x) \cdot L_H(1) d\mu = 0$$

    because $L_H(1) = 0$

18. Dynamical quantities:

$$H(T) = \int_{\mathbb{R}^N} u(T,x)\log u(T,x)d\mu, \qquad D(T) = \int_{\mathbb{R}^N} |\nabla \sqrt{u(T,x)}|^2 d\mu$$

19. **Lemma:**

$$\frac{d}{dT}H(T) = -4D(T)$$

**Proof:** Just do it and use the parabolic equation

20. Note there's an analogous thing that happens when instead of a differential operator, we have a jump process (with no Leibniz rule). Then we have

$$\dot{H}(T) \le -4D(T)$$

21. Further note that if $\dot{H}(T) \le -CH(T)$, then $H(T) \le e^{-CT}H(0)$

22. But:

$$\frac{1}{T}\int_0^T D(s)ds \le \frac{H(0)}{T}$$

23. Ex: Take an interval $[0, N]$, and $\mu(dx) = \frac{1}{N}dx$, with $u(0, x)$ supported in some finite interval of width 1 and height $N$, then

$$H(0,x) \le \log N, \quad \frac{1}{N^2T}\int D(s)ds \le \frac{c\log(N)}{N^2T}$$

24. **Theorem:** (Bakry-Emery Theorem) $H(f) \le \frac{2}{k}\overline{D}(f)$, where $D^2H \ge k > 0$

25. Application: For $\partial_T u(T,x) = \Delta u(T,x)$ (i.e. a brownian motion/solution to classical heat equation), then we get a **magical fact**

$$||u(T,x)||_{L^\infty} \le (4\pi T)^{-N/2}||u(0,x)||_{L^1(dx)}$$

We get this by looking at

$$\frac{d}{dt}\log||u(T,x)||_{p(T)} = \frac{\dot{p}(T)}{p(T)^2}\left[-\frac{4(p(T)-1)}{\dot{p}(T)}\int|\nabla F(T)|^2dx + \int|F(T,X)|^2\log|F(T,X)|^2dx\right]$$

where $p(T)$ is some function such that $p(0) = 1$ and $p(T) = \infty$ and

$$F(T,X) = \frac{u(T,X)^{p(T)/2}}{||u(T,X)||_{p(T)}^{p(T)/2}}$$

If we choose $p(T)$ so that

$$\frac{a^2}{\pi} = \frac{4(p(T)-1)}{\dot{p}(T)}$$

then the LSI (logarithmic sobolev inequality) yields

$$\frac{d}{dT}\log||u(T,x)||_{p(T)} \le -\frac{N\dot{p}(T)}{p(T)^2}\left(1 + \frac{1}{2}\log\left(\frac{4\pi(p(T)-1)}{\dot{p}(T)}\right)\right)$$

having used gronwall and $p(s) = \frac{T}{T-S}$

26. **Theorem:** (Carlen-Loss), for $\partial_T - L = 0$, $L = \nabla\cdot(D(T,x)\nabla) + b\nabla$

27. Proof of Bakery-Emery: Recall

$$\dot{H}(T) = -4D(T)$$

We proceed as follows

$$\dot{D}(t) = \frac{d}{dT}\int|\nabla h_T|^2 d\mu$$

where $h_T = \sqrt{u(T,x)}$ and so

$$\partial_T h_T = \frac{1}{2h_T}Lh_T^2 = Lh_T + \frac{1}{h_T}(\nabla h_T)^2$$

and so

$$\dot{D}(T) = \int 2\nabla h_T \cdot \nabla \partial_T h_T d\mu = \int 2\nabla h_T \cdot \nabla L h_T d\mu + 2\int \nabla h_T \cdot \nabla \left( \frac{|\nabla h_T|^2}{h_T} \right) d\mu$$

more stuff follows, pushing through we finally get

$$\frac{d}{dT}\dot{D}(T) \leq -2\int \nabla h_T \cdot D^2 H \nabla h_T dx \leq -2kD(T)$$

so integrating gives $D(T) \leq e^{-2kT}D(0)$ so as $T \to \infty$ we have $u(T,x) \to 1$.

28. From here, we get LSI and exponential relaxation

## 10/25: Shintaro Fushida-Hardy's talk on "Using (a little bit of) Entropy to Classify Surface Geometries"

1. Uniformization theorem
   (a) For $\Sigma$, a complex structure, simply connected, then such a surface is biholomorphic to one of the following: A Riemann sphere, the unit disk, or the entire complex plane
   (b) In fact this gives a conformal equivalence between any such surface and one of the three above
2. The guide for this talk is as follows:
   (a) Geometry
   (b) Ricci Flow
   (c) A priori estimates
   (d) Convergence of solutions
3. Set up is: $(M, g)$ a Riemannian manifold with a metric. We write

$$h = e^u g \text{ s.t. } u : M \to \mathbb{R} \text{ smooth}$$

   Note that for this definition of $h$, we have $h$ is conformally equivalent to $g$
4. We consider the equivalence classes of such conformally equivalent metrics, yielding $(M, [g])$
5. Connection: $\nabla$
6. **Theorem:** (Fundamental theorem of Riemannian geometry): A levi-civita connection exists
7. Riemann curvature:

$$R(X,Y)Z = \nabla_X \nabla_Y Z + \dots$$

   where the coefficients are given by $\{R^i_{jkl}\}$. We also have

$$\mathrm{Ric}_{jl} = R^i_{jil}, \qquad Sc = \mathrm{Ric}^j_j$$

8. TFAT (The following are true):
   (a) $\mathrm{Ric}[cg] = \mathrm{Ric}[g]$ if $c > 0$ is constnat
   (b) For a 2-manifold: $\mathrm{Ric} = \frac{1}{2}Sc \cdot g$. Note that on a surface, the gaussian curvature is a constant multiple of the scalar curvature
   (c) If $g = e^u h$, then

$$Sc[g] = e^{-u}\left(Sc[h] - \Delta_h u\right)$$

   (d) For a two manifold:

$$[\Delta, \nabla] = \frac{1}{2}Sc \cdot \nabla$$

   where $\Delta = \nabla \cdot \nabla = \nabla^i \nabla_i$

9. Ricci Flow:
$$\frac{\partial}{\partial t} g = -2\text{Ric}, \qquad g(0) = g_0$$
the solution to this equation is the Ricci flow

10. Canonical example is $(S^n, g_0)$ where $g_0$ is the canonical metric on $S^n$ via embedding into $\mathbb{R}^{n+1}$.

    Note: Ric $= (n-1)g_0$ for the sphere. We guess $g(t) = r^2(t)g_0$, and then
$$2r\frac{\partial r}{\partial t} g_0 = \frac{\partial}{\partial t} g = -2\text{Ric}[g] = -2\text{Ric}[g_0] = -2(n-1)g_0$$
which yields
$$r(t)^2 = r_0^2 - 2(n-1)t$$
having used scalar invariance of Ricci curvature.

11. The above is bad because the Ricci flow dies in finite time! And we can't really classify surfaces via ricci flow if they vanish in finite time

12. Volume: define as $\int_M d\mu$ where $d\mu = \sqrt{\det g} dx^1 \wedge \cdots \wedge dx^n$.

13. Claim: If $g$ is a solution to Ricci flow, then
$$\frac{\partial}{\partial t} d\mu = -Sc d\mu$$

14. With the above claim, we normalize ricci flow by adding back in this scalar term to allow for volume preservation. In particular
$$\frac{\partial}{\partial t} g = -2\text{Ric} + \frac{2}{n} rg \text{ s.t. } g(0) = g_0$$
where $r = Avg(Sc) = \left(\int_M d\mu\right)^{-1} \left(\int_M Sc d\mu\right)$

15. Note: we're assuming that our manifolds are closed and oriented.

16. Hamilton: There is a one to one correspondence between solutions to Ricci flow and solutions to normalized Ricci flow by reparameterizing time and space

17. Surfaces: (closed and oriented)

18. Uniformization: $(\Sigma, g)$ a surface, $g$ is conformally equivalent to a metric of constant curvature

19. Uniformization II: $(\Sigma, g_0)$ a surface, then there exists a solution to normalized ricci flow:
$$\frac{\partial}{\partial t} g = (r - Sc)g \text{ s.t. } g(0) = g_0$$
and the solution exists for all time $t$. Moreover $\lim_{t\to\infty} g(t)$ converges to a constant curvature metric in all $C^k$ norms

    In particular, in the above, the solution is just an exponential times the initial metric, so we get a conformal equivalence.

20. Gauss-Bonnet: $Area(\Sigma) \cdot r = \pi \cdot \chi(\Sigma)$ so $r$ is a constant!

21. Going back to the solution for normalized Ricci flow on a surface:
$$g(t) = e^u g_0, \qquad Sc_{g(t)} = e^{-u}(Sc_{g_0} - \Delta_{g_0} u)$$
$$\frac{\partial}{\partial t} Sc = \Delta Sc + Sc(Sc - r)$$
the last equation is nice because there's a diffusion term, $\Delta Sc$, and then a reaction term, $Sc(Sc - r)$

22. Maximum principle:

    (a) $M$ closed, $F$ locally lipschitz

(b) Suppose $u$ satisfies

$$\frac{\partial u}{\partial t} = \Delta_{g(t)} u + F(u)$$

Suppose $\exists C \in \mathbb{R}$ such that

$$u(x,0) \leq C \quad \forall x \in M$$

Let $\varphi$ solve

$$\frac{d\varphi}{dt} = F(\varphi), \qquad \varphi(0) = C$$

then $u(x,t) \leq \varphi(t)$ for all $t$.

23. Other maximum principles exist: e.g. just reverse the direction of the inequalities and this also solves

24. Using both of the maximum principles: If $g$ is a solution to the normalized Ricci Flow on $(\Sigma, g_0)$, then there exists a $C$ such that:

$$\begin{aligned} r < 0 &\implies r - Ce^{rt} \leq Sc \leq r + Ce^{rt} \\ r = 0 &\implies \text{some polynomial bounds on both sides} \\ r > 0 &\implies -Ce^{-rt} \leq Sc \leq r + Ce^{rt} \end{aligned}$$

25. **Proposition:** For $(\Sigma, g_0)$ our closed and oriented surface, there exists a unique solution, $g(t)$, to the Normalized Ricci Flow for all time $t$

26. Is Uniformization theorem true?

    If $r < 0$, then things work out. Similar for other two cases. In particular for the $r > 0$ case, we define

$$N(g) = \int_\Sigma Sc \log(Sc) d\mu$$

then

$$\frac{d}{dt} N = -\int \frac{|\nabla Sc + Sc\nabla f|^2}{Sc} d\mu = -2\int_\Sigma |M|^2 d\mu$$

where

$$\Delta f = Sc - r, \qquad M := \text{trace free Hessian of f}$$

so in particular, the entropy never increases. Moreover, we can bound scalar curvature with something of the form

$$\log(Sc\Big|_t) < N(t) < N_1$$

which will help us finish the $r > 0$ case to show that ricci flow converges to constant curvature metric.

## 11/1: Yuval's Talk on "Claude Shannon, Master of Uncertainty"

1. Yuval will talk about the history and origin of Entropy and how it relates to information

2. In this talk,

   (a) $\chi$ is a finite set

   (b) $X$ is a random variable on $\chi$

   (c) $p(x)$, for $x \in X$, is defined $p(x) = Pr[X = x]$

   (d) The entropy is

$$H(X) = \sum_{x\in\chi} p(x) \log((1/p(x))) = -\sum_{x\in\chi} p(x) \log(p(x)) = E_X[-\log p(x)]$$

3. Historically, Hartley in the '20s defined information in $X$ is equal to $\log|\chi|$. This is additive if we consider $\chi_1$ and $\chi_2$ and take the product space.

4. Problem: This definition doesn't care about the distribution of $X$

5. How much information do we get by observing an outcome $x \in \chi$?
6. Shannon says $-\log p(x)$ is a good measure
7. Ex: we're supposed to guess a number between 1 and 8 by yes/no questions. How do? Binary search because at each term, the probability of getting the answer yes is 1/2 and no is 1/2

   At each step, you get $-\log(1/2) = 1$ bits of information. In total 3 bits of information
8. In the "happy birthday" strategy (i.e. consecutively asking is it 1? 2? all the way until 8 until you get a "yes"), at turn 1 we have $Pr[Yes] = \frac{1}{8}$ and $Pr[no] = \frac{7}{8}$. If you get no, you've earned $-\log(7/8)$ bits of information

   At turn 2, $Pr[Yes] = 1/7$ and $Pr[no] = 6/7$. Suppose you finally get it right at step $n$, total info:

   $$-\log(7/8) - \log(6/7) - \cdots - \log(n/(n+1)) - \log(1/n) = -\log\left(\frac{7}{8} \cdot \frac{6}{7} \cdots \frac{n}{n+1} \cdot \frac{1}{n}\right) = -\log\left(\frac{1}{8}\right) = 3$$

   where we get a bunch of "no"s until the last one which is a yes occurring with probability $1/n$. Convince yourself that any other guessing strategy will always result in 3 bits of information!
9. Another way to see that this is a good definition of information is there exists an axiomatic formulation such that the information function exists and is unique up to a constant. This shows that the information must be $K\sum_{x\in\chi} p(x)\log(p(x))$.
10. note that

    $$0 \leq H(x) \leq \log|\chi|$$

    where the left hand bound happens when we have a totally deterministic situation, i.e. $p(x_0) = 1$. The right hand bound occurs when we're distributed uniformily
11. Aside: sometimes $-\log p(x)$ is called the "surprise" so the entropy, $E_X[-\log p(x)]$, is called the expected surprise!!!
12. Claude Shannon is a boss and apparently established a lot of this entropy/information theory machinery in one paper
13. **Theorem:** (Noiseless coding theorem) Suppose you want to compress $X$ s.t. if can always be recovered with 100 percent accuracy (to be defined). Then

    $$E[length] \geq H(X)$$

    no matter how this is done. Moreover it can be done with $E[length] \leq H(X) + 1$.

    Formally, we think of encoding a function $f : X \to \{0,1\}^N$ for $N = \log_2|X|$ or comparable and then we're taking $E_X[|f(x)|]$
14. Apparently, the above is how files are zipped and unzipped
15. **Theorem:** (Noisy/source coding theorem) If $X_1, \ldots, X_n$ are iid copies of $X$, then the vector $(X_1, \ldots, X_n)$ can be compressed into $\{0,1\}^k$, if $k > nH(x)$ with negligible probability. However, no matter how you map to $\{0,1\}^k$, if $k < nH(x)$, then $Pr[error] \to 1$.
16. Apparently, the above is how .jpeg's are formmed
17. **Theorem:** (Channel Coding theorem) Let $W$ be a noisy channel "W is a random function from $\chi \to \mathcal{Y}$" then $\exists C = C(W)$ some number. We define

    $$\text{rate} = \frac{\text{Number of information bits communicated divided by}}{\text{Number of bits sent}}$$

    If rate $\ll C(w)$, then there exists a coding scheme that fails with negligible probability. If rate $> C(W)$, then all schemes fail with probability tending to 1. If the noise is gotten by "adding $X$", then this capacity is $1 - H(X)$ (maybe there should be some renormalization here).
18. The above is how phones work! When transmitting a signal, there is actually a lot of noise coming from cosmic rays, trees, birds, etc. but there are built in error correcting codes which helps us transmit a coherent signal

19. Noiseless coding proof

    (a) **Definition:** A map $f : \chi \to \{0,1\}^*$ (the codomain is the set of all finite binary strings) is called a prefix code if no $f(x)$ is a prefix of any $f(x')$, e.g.

    $$\{a, b, c\} \to \{0, 10, 11\}$$

    is a good prefix code, but

    $$\{a, b, c\} \to \{0, 01, 11\}$$

    is bad, because 0 is a prefix of 01.

    (b) We convert as follows

    $$01011110010011 \to 0\ 10\ 11\ 11\ 0\ 0\ 10\ 0\ 11 \to a\ b\ c\ c\ a\ a\ b\ a\ c$$

    this procedure works because there's no ambiguity given the prefix code

    (c) **Theorem:** (Kraft's inequality) Let $f : \chi \to \{0,1\}^*$ be a prefix code. Let $\ell(x) =$ length of $f(x)$. Then

    $$\sum_{x \in \chi} 2^{-\ell(x)} \leq 1$$

    **Proof:** Pick a uniformily random infinite binary string $U$. Then

    $$1 \geq Pr[\exists x \ :\ f(x) \text{ is a prefix of U}] = \sum_x Pr[f(x) \text{ is a prefix of } U] = \sum_{x \in \chi} 2^{-\ell(x)}$$

    (d) **Theorem:** Let $f : \chi \to \{0,1\}^*$ be a prefix code. Then $E_x[\ell(X)] \geq H(X)$
    **Proof:** We compute

    $$H(x) - E[\ell(X)] = E_X[\log(1/p(x)) - \ell(x)] = E_X[\log\left(\frac{1}{p(x)2^{\ell(x)}}\right)] \overset{Jensen}{\leq} \log E_x[\log \frac{1}{p(x)2^{\ell(x)}}] = \log \sum_x 2^{-\ell(x)} \leq 0$$

    finishing the proof

20. Noisy coding proof

    (a) We want an encoder $E : \chi^n \to \{0,1\}^k$ and a decoder $D : \{0,1\}^k \to \chi^n$. Then

    $$Pe = \text{Probability of Error} = Pr_{X_1,\dots,X_n}[D(E(X_1, \dots, X_n)) \neq (X_1, \dots, X_n)]$$

    (b) **Theorem:** If $k > n(H(X) + 2)$, there exists a $D,\ E$ such that $Pe \to 0$ as $n \to \infty$. If $k < n(H(X) - \epsilon)$, then for all $D$ and $E$, we have $Pe \to 1$
    **Proof:** Define $Y_n = -\log p(X_1, \dots, X_n) = -\sum_{i=1}^n \log p(X_i)$. We can use the weak law of large numbers on $Y_n$, to get

    $$\forall \delta > 0, \qquad \lim_{n \to \infty} Pr[\frac{1}{n}\left|Y_n - EY_n\right| > \delta] = 0$$

    where

    $$EY_n = E[-\log p(X_1, \dots, X_n)]$$

    We can define the typical set

    $$T_{n,\delta} = \{(x_1, \dots, x_n) \in \chi^n \mid 2^{-n(H(X)+\delta)} \leq p(X_1, \dots, x_n) \leq 2^{-n(H(X)-\delta)}\}$$

    Then we have

    $$Pr[(X_1, \dots, X_n) \notin T_{n,\delta}] \to 0$$

    this somehow finishes the proof.

# 11/15: Nikolas Kuhn's talk on "Perelman Entropy and Ricci Flow with Surgery"

1. Talk sketch
   (a) Hamilton's program (1982)
   (b) Basic properties; long + short time existence
   (c) Analysis of singularities
   (d) Noncollapsing + Perelman entropy
2. In dimension 2, Ricci flow gives us the uniformization theorem.
3. Hamilton's program told us that
   $$\dot{g} = -2\text{Ric}$$
   so if Ric is positive then $M$ (our manifold with the metric $g$) shrinks. If Ric is negative, then $\mu$ expands
4. On $S^n$, $g(t) = 2(n-1)(T-t)g_0$
5. **Theorem:** (Hamilton) For $(M, g)$ a 3-fold such that Ric is everywhere positive, then blowup occurs as $t \to T$ at some time $T$ with $0 < T < \infty$. Moreover, $g(t) \to g_\infty$ positive with $g_\infty$ a metric of constant positive, sectional curvature.
6. Don't expect uniformization. Even worse, consider $S^3$ with a metric such that it looks like two spheres connected by a little tube diffeomorphic to $S^2 \times I$. We hope that Ricci flow uniformizes this deformation of $S^3$ back to a single $S^3$, but really it makes this little tube thinner and longer, stretching it towards two copies of $S^3$ connected by a thin tube
7. Hamilton's idea: Just before the singularity occurs, do surgery to break the tube and get two nice copies of $S^3$ under Ricci flow
8. We now get some equations: If $g(t)$ satisfies Ricci-flow then $\lambda^2 g(t/\lambda^2)$ does too.
9. We'll get an equation like
   $$\dot{R} = \Delta R + 2(\text{Ric})^2$$
   which we can bound below by $\Delta R + \frac{2}{d}R$, which is a heat transport esque equation and is in a nice class of PDEs
10. We get some more equations that are too burdensome to copy down
11. In parabolic theorem, we have the weak maximum principle: Given $u$ such that
    $$\frac{\partial u}{\partial t} \leq \Delta u + F(u,t), \qquad u(x,0) \leq \alpha$$
    for some function $F$. Suppose we also have a solution of $\frac{d\phi}{dt} = F(\phi, t)$ such that $\phi(0) = \alpha$, then we can bound $u$ in terms of $\phi$. In particular
    $$u(x,t) \leq \phi(t) \qquad \forall x, t$$
    now apply this to the absolute value of the Riemann tensor, squared, $|Rm|^2$ and $F = Cr^{3/2}$ so that
    $$|Rm| \leq \frac{M}{1 - \frac{1}{2}CMt}, \qquad \text{if} \quad |Rm|_{t=0} \leq M$$
12. Short time existence:
    $$\dot{g} = -2\text{Ric}$$
    and we linearize
    $$\frac{\partial h}{\partial t} = \Delta_L h + \mathcal{L}_{\delta G(h)^\#} g =: L$$
    we want to compute the symbol of $L$ to show that it is parabolic, then
    $$\sigma(L)(x,\xi)(h) = |\xi|^2 h - \xi \times h(\xi^\#, \cdot) - h(\xi^\#, \cdot)\xi + (\xi \times \xi)\text{tr}h$$
    where $\xi^\#$ denotes the metric induced musical isomorphism. This is unfortunately not parabolic ⌢̈

13. Via explicit computation, if we have $h = \xi \times \xi$ then the above symbol vanishes!

14. Through some reasoning, let
$$g(t) = \psi_t^* \tilde{g}(t)$$
with $\tilde{g}$ parabolic. Then
$$\dot{\tilde{g}} = -2\mathrm{Ric}(\tilde{g}) + \mathcal{L}_{X(t,\tilde{g})}\tilde{g}$$
In fact, $x = (T^{-1}\delta(s(T)))^{\#}$ for $T$ any positive symmmetric 2-cotensor.

15. Now the symbol of the linearization is $|\xi|^2 h$, which implies existence of a solution to our equation. With more work we can get uniqueness

16. Looooooooong-time existence: Suppose $(M, g(t))$ has maximally extended Ricci flow on $[0, T)$

17. **Theorem:** Then $\sup_{x \in M} |Rm|(x,t) \to \infty$ as $t \uparrow T$

18. **Lemma:** If $|\mathrm{Ric}| \leq M$, then
$$e^{-2Mt}g(0) \leq g(t) \leq e^{2Mt}g(0)$$
We also need bounds for $|\nabla^k Rm|$ but this follows by estimates and an induction argument. This implies that $g(T)$ exists and is smooth, so we can write $[0, T]$

19. Analysis at singularities: $M, g(t)$ a maximally extended Ricci-flow on $[0, T]$. Let $\{t_i\}$ be a sequence $t_i \uparrow T$, and let $x_i$ such that $|Rm(t_i, x_i)|$ is maximal. Then define
$$g_i(t) := |Rm(x_i, t_i)| g\left(t_i + \frac{t}{|Rm(p_i, t_i)|}\right)$$

20. now think of this as a flow with a new time origin at $t_i$. Then $|Rm_{g_i}(0, x)| \leq 1$, and it exists on $[-t_i \cdot |Rm(t_i, x_i)|, (T-t)|Rm(p_i, t_i)]$

21. We want to get a limiting flow $g_\infty$

22. We also want that as $t_i \uparrow T$, that $r_{inj, r_i}$ doesn't decay faster than $|Rm|$ grows, where $r_{inj}$ is the injectivity radius

23. In fact: $r_{inj}$ can be bounded below by the volume of small balls $B(p, r)$ for certain $r \leq C|Rm|$

24. Thus we want the volume of these balls to be bounded

25. This was achieved by Perelman using some entropy functional
$$W(g, f, \tau) = \int_M [\tau(R + |\nabla f|^2) + f - d] \frac{e^{-f}}{(4\pi\tau)^{d/2}} dV = W(\frac{g}{\tau}, f, 1)$$
where the last equality is some scale invariance by doing it out

26. Let $u = \frac{e^{-t}}{(4\pi\tau)^{d/2}}$ then we get
$$\frac{d}{dt} W(g, f, t) = 2\tau \int_M |\mathrm{Ric} + Hess(f) - \frac{g}{2\tau}|^2 \frac{1}{(4\pi\tau)^{d/2}} e^{-f} d\mu$$

27. Suppose $g$ is Ricci flow, $0 < \tau$ and $\dot{\tau} = -1$, then $\dot{u} = -\Delta u = Ru$

## 11/22: Jared Marx-Kuo's talk on "An Entropic View of the Central Limit Theorem"

## 12/6: Andrea Ottolini's Talk on "Equivalence of Ensembles"

1. We have a bunch of particles (ensemble) in some larger ambient collection, e.g. water. We have another ensemble further away, but the latter ensemble is at temperature equilibrium
2. Let $\nu_n^C$ = microcanonical distribution of n particles whose energy lies in $C$ where $C$ is some interval (the particles have maximum entropy given energy in $C$—)
3. Also define $\gamma_n^\beta$ = canonical distribution of $n$ particles whose average energy (temp) is $1/\beta$
4. Ex: Ideal gas - $n$ particles $v = (v_1, \ldots, v_n)$ and $\sum_{i=1}^n v_i^2 = nE$. In this case the microcanonical distribution is $\nu_n^{nE}$ is uniform on $\sqrt{nE}S^{n-1}$. Also $\gamma_n^\beta$ is the product of Gaussians with variance $E$ for $\beta = 1/(2E)$
5. The theorem we're trying to prove is: **Theorem:**

$$||\nu_{n,1}^{nE} - \nu_{n,1}^{1/(2E)}||_{TV} \leq \frac{1}{n}$$

   where $\nu_{n,1}^{nE}$ denotes the one-dimensional marginal distribution (which is the context to which we apply the total variation measure)
6. Lattice Systems: define $\{\wedge_n\}_{n\in\mathbb{Z}} \subseteq \mathbb{Z}^d$, $V_n = |\wedge_n|$, $\Omega_n = \{-1,1\}^{V_n}$ where $\omega \in \Omega_n$ then $\xi_s(\omega) = \omega(s)$ is the spin at site $s$
7. Also define $M_n = \sum_{s\in\wedge_n} \xi_s$ = total magnetization and $\rho_n$ is a reference measure on $\Omega_n$. We have a map

$$T_n : \Omega_n \to X \subseteq \mathbb{R}^k, \qquad T \text{ is a "sufficient statistic"}$$

8. **Definition:** The microcanonical ensemble with energy in $C \subseteq X$ is given by

$$\nu_n^C(\cdot) = \alpha_n^C(\cdot)\rho_n^C(\cdot)$$
$$\alpha_n^C(\cdot) = \frac{\mathbb{1}_{T_n^{-1}(C)}(\cdot)}{\rho_n(T_n^{-1}(C))}$$

9. **Definition:** The canonical distribution with inverse temperature $\beta \in \mathbb{R}^k$ is

$$\gamma_n^\beta(\cdot) = \alpha_n^\beta(\cdot)\rho_n(\cdot)$$
$$\alpha_n^\beta = \frac{\exp(V_n\langle\beta, T_n(\cdot)\rangle)}{\exp(V_n p_n(\beta))}$$

10. Ex: Take $k = 1$, $\rho_n$ the counting measure, $X = [-1,1]$ and $T_n = \frac{M_n}{V_n}$ then this is called the paramagnet in physic. In math, the microcanonical ensemble fixes the number of heads and then calculates a distribution, whereas the canonical distribution is just normal coin flipping
11. Ex: $k = 1$, $\rho_n^{(\cdot)} = \exp(a\frac{M_n^2(\cdot)}{2V_n})$, $X = [-1,1]$, $T_n = \frac{M_n}{V_n}$ (Curie-Weiss Model)
12. Ex: $k = 2$, $U_n = -\sum_{i\sim s} \xi_i\xi_s$, $X = [-d,d] \times [-1,1]$ and $T_n = \left(\frac{U_n}{V_n}, \frac{M_n}{V_n}\right)$, $\rho_n$ the counting measure (Ising model)
13. **Definition:** We say equivalence of ensembles holds if $\forall C \subseteq X$, (for $C$ convex and open), there exists a $\beta \in \mathbb{R}^k$ such that $\forall\Delta$ finite subset of a lattice:

$$||\nu_{n,\Delta}^C - \gamma_{n,\Delta}^\beta||_{TV} \to 0$$

14. **Definition:** Given two measures $\lambda_1, \lambda_2$ on some probability space $\Omega$, the relative entropy of $\lambda_2$ w.r.t. $\lambda_1$ is given by

$$H(\lambda_1 \mid \lambda_2) = \int_\Omega \ln\left(\frac{d\lambda_1}{d\lambda_2}\right) d\lambda_1$$

   if the Radon-Nikodym derivative, $\frac{d\lambda_1}{d\lambda_2}$ exists. Else $H(\lambda_1 \mid \lambda_2) := +\infty$

15. Suppose $\gamma_n^\beta$ is a product measure. Take $\Delta_1, \Delta_2$ two disjoint copies of a fixed $\Delta$. Under this assumption, if we look at

$$H(\nu^C_{n,\Delta_1\cup\Delta_2} \mid \gamma^\beta_{n,\Delta_1\cup\Delta_2}) \geq H(\nu^C_{n,\Delta_1} \mid \gamma^\beta_{n,\Delta_1}) + H(\nu^C_{n,\Delta_2} \mid \gamma^\beta_{n,\Delta_2})$$

16. We also have an eaiser bound

$$H(\nu_n^C \mid \gamma_n^\beta) \geq \frac{V_n}{|\Delta|} H(\nu^c_{n,\Delta} \mid \gamma^\beta_{n,\Delta}) \overset{pinsker}{\geq} \frac{V_n}{|\Delta|} ||\nu^C_{n,\Delta} - \gamma^\beta_{n,\Delta}||_{TV}$$

where we assume that $\Delta$ and the $\Delta_i$'s are nice shapes.

17. If $\frac{H(\nu_n^C \mid \gamma_n^\beta)}{V_n} \to 0$, then equivalence of ensembles holds

18. Note that $\gamma_n^\beta$ being a product measure is a big assumption, but by reductions using subadditivity, we can remove this constraint for weaker ones

19. Somehow, we want to change the integral over $\Omega$ in the definition of relative entropy to just an integral over $X$ by using symmetry and the fact that our distributions only depend on the energy level

20. Idea: Use a change of variables from $\Omega_n \to X$ and the fact that $T_n$ is sufficient to reduce the problem to

$$M_n(\cdot) = \rho_n(T_n^{-1}(\cdot)), \qquad M_n[\cdot \mid C] = \frac{M_n[\cdot \,\cap\, C]}{M_n[C]}$$

$$M_n^\beta[dx] = \frac{\exp(v_n\langle\beta, x\rangle)M_n[dx]}{\exp(v_n\rho(\beta))} M_n[dx]$$

$$\frac{H(\nu_n^V \mid \gamma_n^\beta)}{V_n} = -\int_C \langle\beta, x\rangle M_n[dx \mid \mathbb{C}] + p_n(B) - \frac{1}{V_n}\ln M_n[C]$$

note that $M_n$ is a measure on $X$

21. A short crash course on large deviations:

 (a) **Definition:** We have a sequence of measures and scaling functions $\{(M_n, V_n)\}$. We say that this sequence satisfies "L.D.P." (large deviation principles) with rate function $-\mu$ if

$$\exists M : X \to \mathbb{R} \qquad \text{s.t.} \qquad \forall C, \quad \text{open, convex}$$

 that

$$\lim_{n\to\infty} \frac{1}{V_n} \ln M_n[C] = \sup_{x\in\overline{C}} \mu(x)$$

 (b) From now on, an "*L.D.P.*" is assumed to hold. What is

$$\lim_{n\to\infty}\inf \int_C \langle\beta, x\rangle M_n[dx \mid C] \geq \inf_{x\in X_{\overline{C}}} \langle\beta, x\rangle$$

 where $X_{\overline{C}} = \{x \in X \;\text{ s.t. }\; \mu(x) = \sup_{y\in\overline{C}} \mu(y)\}$

 (c) Moreover, we have

$$p_n(\beta) \to p(\beta) = \sup_{x\in X}\{\langle\beta, x\rangle + \mu(x)\}$$

$$\exp(v_n p(\beta)) \approx \sum_x \exp(v_n\langle\beta, x\rangle)\exp(v_n\mu(x))$$

 and we define

$$X_B := \{x \in X \;\text{ s.t. }\; \langle\beta, x\rangle + \mu(x) = p(\beta)\}$$

22. **Theorem:** if an $\widetilde{L.D.P}$ holds, and if $X_C \subseteq X_\beta$, then equivalence of ensembles holds.

 **Proof:**

$$\limsup_{n\to\infty} \frac{H(\nu_n^C \mid \gamma_n^\beta)}{V_n} \leq \inf_{X_{\overline{C}}} \langle\beta, x\rangle + p(\beta) - \sup_{x\in C}\mu(x)$$
$$\leq \sup_{X_{\overline{C}}}\{p(\beta) - \langle\beta, \lambda\rangle - \mu(x)\} = 0$$

23. We consider two examples

24. Ex: (Paramagnet) $\mu(x) = (1+x)\ln(1+x) + (1-x)\ln(1-x)$. Given $C = (c_1, c_2)$ then

$$X_{\overline{C}} = \begin{cases} c_1 & c_1 > 0 \\ c_2 & c_2 < 0 \\ 0 & 0 \in (c_1, c_2) \end{cases}$$

For all $C$, there exists a $\beta$ such that $X_C \subseteq X_\beta$, so we're good

25. Ex: (Curie-Weiss), we have $\mu(x) = \mu_0(x) + ax^2/2 + C$ for some constant $C$