THE LOW ENTROPY METHOD AND MEASURE RIGIDITY FOR $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$.

ALEX WRIGHT

Abstract. These are notes for the final talk in a reading group on the Pisa notes of Einsiedler and Lindenstrauss. In the first section we state the high and low entropy methods, and show how they are combined to give the measure rigidity theorem of Einsiedler, Katok, and Lindenstrauss. In the second section we discuss the low entropy method in this context.

Reader beware: these notes may contain errors or misleading statements. Corrections are welcome.

1. Measure rigidity for $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$.

Statement of result. Let A denote the subgroup of $SL(3,\mathbb{R})$ consisting of diagonal matrices with positive diagonal entries. The goal of this talk is to explain

Theorem. *Let μ be an A invariant and ergodic probability measure on $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$. Assume that there is a one parameter subgroup of A which acts with positive entropy. Then μ is Haar measure.*

A version of this theorem is true when $SL(3,\mathbb{R})$ is replaced with $SL(n,\mathbb{R})$, but the conclusion is that μ is algebraic. The theorem is not true when $SL(n,\mathbb{Z})$ is replaced with an arbitrary lattice. Removing the entropy assumption is a major open problem.

Notation. The following subgroups of $G = SL(3,\mathbb{R})$ play a central role:

$$U_{12} = \begin{pmatrix} 1 & * & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad U_{13} = \begin{pmatrix} 1 & 0 & * \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$U_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix}, \quad G^- = \begin{pmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that $[U_{12}, U_{23}] = U_{13}$, and that both U_{12} and $U_{2,3}$ commute with U_{12}. Denote by μ_x^{ij} the leaf-wise measure on U_{ij} at a point $x \in X = SL(3,\mathbb{R})/SL(3,\mathbb{Z})$.

The high entropy method. Recall from the last talk

Theorem (The high entropy method.)**.** *If μ^{12} and μ^{23} are non-trivial, then μ is invariant under U_{13}.*

A leaf-wise measure is trivial if it is the dirac mass at the identity.

We will use without proof the following easy corollary of Ratner's Theorem.

Fact. *If μ is a measure on $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$ that is invariant under A as well as a one parameter unipotent subgroup, then μ is Haar.*

Therefor, the high entropy method in fact yields that if μ^{12} and μ^{23} are non-trivial, then μ is Haar.

The low entropy method. Suppose that $a \in A$ acts with positive entropy. Thus, some μ_x^{ij} is almost surely non-trivial. Without loss of generality, we will assume that μ_x^{13} is almost surely non-trivial. For concreteness, we will assume

$$a = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

This assumption is not justified, most of the ideas of the proof are visible already in this case. Note that with this choice of a, the G^- above is indeed the subgroup of G contracted by $\mathrm{Ad}(a)$.

Let $C_G(U_{13})$ denote the centralizer of U_{13} in $G = SL(3,\mathbb{R})$. We will be interested in $C_G(U_{13}) \cap G^-$, which is equal to all of G^- in our case. This would not be true if $SL(3,\mathbb{R})$ were replaced with $SL(n,\mathbb{R})$, so we will continue to write $C_G(U_{13}) \cap G^-$ to emphasize that in general we would only be interested in the part of G^- which commutes with U_{13}.

Theorem (The low entropy method.)**.** *There are μ-generic x such that $\mu_x^{13} = \mu_{hx}^{13}$ for some $h \in C_G(U_{13}) \cap G^-$.*

Before discussing the low entropy method in the second section, we explain how it is applied.

Combining the high and low entropy methods. We start out by assuming only that μ_{13} is non-trivial. There are now two possibilities.

1. (1) The support of $\mu_x^{G^-}$ is almost surely contained in U_{13}. (This is equivalent to: there is a subset of full measure $X' \subset X$ such that any two points $x, y \in X'$ which lie on the same G^- orbit in fact lie on the same U_{13} orbit.)
2. (2) The support of $\mu_x^{G^-}$ is almost surely not contained in U_{13}.

Because μ is A ergodic (and U_{13} is normalized by A), no intermediate cases are possible.

Lemma. *If $h \in C_G(U_{13}) \cap G^-$, and x is generic, and $\mu_x^{13} = \mu_{hx}^{13}$, then hx is contained in the support of $\mu_x^{G^-}$.*

By "in the support of" or "generic" we mean in a chosen full μ-measure set of points of X which are Birkhoff generic and have other basic properties. All statements such as the lemma implicitly hold only almost surely.

Proof. Since x is in the support of μ, the identity is in the support of μ_x^{13} and hence also μ_{hx}^{13}. Hence hx is in the support of μ and h is in the support of $\mu_x^{G^-}$. □

Lemma. *If the support of $\mu_x^{G^-}$ is almost surely contained in U_{13}, then μ is U_{13} invariant, and thus by the fact above μ is Haar.*

This step is essentially (9) in Section 2.1 of the EKL paper. It is not explained in the Pisa notes. The idea is that the low entropy method provides some translation invariance for many μ_x^{13}, and this can be upgraded to showing μ_x^{13} is Haar by using the a dynamics.

Proof. Let $B \subset X$ by the set where μ_{13} is invariant by a non-trivial translation in U_{13}. First suppose, in order to find a contradiction, that B has measure zero. The low entropy is compatible with avoiding a given set of measure zero, and hence can produce $x \in X \setminus B$ and $h \in C_G(U_{13}) \cap G^-$ such that $\mu_x^{13} = \mu_{hx}^{13}$. By the previous lemma and our assumption we see that h is in fact in U_{13}, giving the desired contradiction.

Hence B has positive measure. By ergodicity, it has full measure.

Now suppose in order to find a contradiction that a positive measure subset of B is not invariant by all of U_{13}. Since the stabilizer of a measure is a closed subgroup, this gives that a positive measure subset of B has that the stabilizer of μ_x^{13} is discrete and hence contains a smallest non-zero element. This element is contracted by the a action, and thus Poincaré Recurrence gives a contradiction.

Hence almost every x has that μ_x^{13} is invariant under U_{13}. Hence μ_x^{13} is almost surely Haar and μ is U_{13} invariant. □

Thus we are reduced to case (2): The support of $\mu_x^{G^-}$ is almost surely not contained in U_{13}. Recall from two weeks ago,

Proposition (The product structure)**.** *$\mu_x^{G^-}$ is proportional to $\iota(\mu_x^{12} \times \mu_x^{13} \times \mu_x^{23}$, where $\iota : U_{12} \times U_{13} \times U_{23} \to G^-$ is the product map.*

The support of $\mu_x^{G^-}$ is thus the product of the supports of the μ_x^{ij}. Thus by assumption at least one of μ_x^{12} and μ_x^{23} is non-trivial. Suppose without loss of generality that μ_x^{23} is non-trivial.

By the time symmetry of entropy for a, one of μ_x^{21}, μ_x^{31}, μ_x^{32} must also be non-trivial. All of U_{21}, U_{31}, U_{32} fail to commute with at least one of U_{13}, U_{23}. This gives a pair U, U' of non-commuting one parameter unipotent subgroups whose leaf-wise measures are almost surely non-trivial. The high entropy method applies equally well to this case, giving that μ is invariant under the one parameter unipotent group $[U, U']$. By the fact above, we conclude that μ is Haar.

2. The low entropy method for $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$.

Setup. μ is an A invariant ergodic probability measure on $X = SL(3,\mathbb{R})/SL(3,\mathbb{Z})$, and $a \in A$ acts with positive entropy. We are assuming that μ_x^{13} is almost surely non-trivial.

The low entropy method hinges on the dynamics of the action of the unipotent subgroup U_{13}. This is remarkable, because the measure μ is not assumed to be U_{13} invariant. Write

$$u(s) = \begin{pmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Proposition (Polynomial shearing)**.** *Let $x, y \in X$ be nearby points, which are not in the same orbit of $C_G(U_{13})$. Then $u(s)x$ and $u(s)y$ diverge at polynomial speed, and the direction of first divergence is in the centralizer $C_G(U_{13})$.*

Of course, if x and y differ in the direction of the centralizer, their orbits do not separate at all. The proposition is based on the simple matrix calculation

$$\begin{aligned} & u(s)\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} u(-s) \\ = & \begin{pmatrix} m_{11}+sm_{31} & m_{12}+sm_{32} & m_{13}+sm_{33} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} u(-s) \\ = & \begin{pmatrix} m_{11}+sm_{31} & m_{12}+sm_{32} & m_{13}+sm_{33}-s(m_{11}+sm_{31}) \\ m_{21} & m_{22} & m_{23}-sm_{21} \\ m_{31} & m_{32} & m_{33}-sm_{31} \end{pmatrix}. \end{aligned}$$

If the m_{ij} are very close to δ_{ij}, then when this expression first becomes distance 1 from the identity matrix it is approximately in

$$\left\{ \begin{pmatrix} m & m_{12} & m_{13} \\ 0 & m_{22} & m_{23} \\ 0 & 0 & m \end{pmatrix} \right\} = C_G(U_{13}).$$

(It may seem like the $\pm s m_{31}$ in the $(1,1)$ and $(3,3)$ entries is a problem, but the $s^2 m_{31}$ in position $(1,3)$ will attain size one before this is significant.)

Idea of the proof of the low entropy method. By Lusin's Theorem, there is a compact subset $K \subset X$ of measure at least $1 - \varepsilon$ where the function $x \mapsto \mu_x^{13}$ is continuous. (Technically, this may require either a stronger than usual form of Lusin, or careful normalization of leaf-wise measures so that they lie in a compact space. I don't really understand the technicalities of Lusin.)

Suppose there are two points $x, y \in K$ in the support of μ, so that $y = gx$ with g very small and not in $C_G(U_{13})$, and with $\mu_x^{13} = \mu_y^{13}$.

Recall that

$$u(s)y = (u(s)gu(-s))(u(s)x),$$

so the orbits diverge in the direction of $u(s)gu(-s)$, which is approximately in $C_G(U_{13})$ when it is first of macroscopic size. Suppose that s can be chosen so that $h = u(s)gu(-s)$ has size about 1 and $u(s)x, u(s)y \in K$. Set $x' = u(s)$ and $y' = u(s)y \approx hx'$. If we can find a sequence of y converging to x, we can assume that the x' and y' converge, giving in the limit $x', y' = hx' \in K$ with $\mu_{x'}^{13} = \mu_{y'}^{13}$ and $h \in C_G(U_{13})$. (The fact that x', y' are in K is used to infer that the equality $\mu_{x'}^{13} = \mu_{y'}^{13}$ persists in the limit.) This is the desired conclusion.

There are two issues with making this rigorous. First, starting pairs x, y must be found with the desired property. We will treat this issue next. After this we will say something about the far bigger issue of choosing s. This is the technical heart of the argument, and involves first moving x and y by the diagonal action.

Initial condition. Let

$$A' = C_G(U_{13}) \cap A = \left\{ \begin{pmatrix} s & 0 & 0 \\ 0 & s^{-2} & 0 \\ 0 & 0 & s \end{pmatrix} \right\}.$$

By the arguments that give the product structure (specifically, Corollary 8.13 in the Pisa notes), $\mu_x^{13} = \mu_{a'x}^{13}$ for any $a' \in A'$.

Now, by Poncaré recurrence, $a'(s)x$ is very close to x for many s. For such s, $a'(s)x = gx$ with g very close to the identity. To run the argument above, we need to arrange that g is not in $C_G(U_{12})$. This is an interesting point in the argument, since if $SL(n,\mathbb{Z})$ is replaced with some other lattice, then examples are known where this is impossible. However, for $SL(n,\mathbb{Z})$ (more specifically we are taking $n = 3$), this can be arranged with some care.

One is forced to consider the decomposition of μ into measures which are ergodic for $a'(s)$. The danger is that each such ergodic component is locally supported on an orbit of $C_G(U_{12})$. In fact one can show that the real danger is that each ergodic component is supported on

$$C_G(\langle U_{13}, U_{13}\rangle).$$

Note that $\langle U_{13}, U_{31}\rangle$ is isomorphic to $SL(2,R)$, and its centralizer is (in our case $n = 3$) A'. By Poincaré recurrence, almost every x has many large s for which $a'(s)x$ is very close to x. If the measure were supported locally on an A' orbit, then $a'(s)x = a'(t)x$ for t very small. Thus $a'(s-t) \in A' \cap SL(3,\mathbb{Z})$ which is a contradiction. (The case $n > 3$ is a bit harder since the centralizer is larger, but not that much harder. See Theorem 5.1 in the EKL paper.)

Here is why I think it is okay to consider $C_G(\langle U_{13}, U_{13}\rangle)$ instead of $C_G(U_{13})$. If the support of μ_x^G is concentrated on a single constant subgroup, then this subgroup will be A normalized. Furthermore, if some element of A contracts part of this subgroup, then this same element should also expand part of this subgroup, because the measure is A invariant. The largest subgroup of $C_G(U_{13})$ that satisfies these conditions is $C_G(\langle U_{13}, U_{13}\rangle)$.

Most choices of s give the right drift. Having explained why suitable x and y exist, we now move to picking the s. Because $u(s)x$ and $u(s)y$ drift apart polynomially, in the range $[-S, S]$ where there are distance at most 1 apart, at least (say) 90% of the s have that $u(s)x$ and $u(s)y$ are between $\frac{1}{2}$ and 1 apart. It is important to note that this is with respect to Lebesgue measure. It will presently become clear that is is μ_x^{13} which is what counts, not Lebesgue.

In fact we can be a bit more precise: since the distance between $u(s)x$ and $u(s)y$ is quadratic, except for two small intervals near the roots of the quadratic, the distance will be at least $\frac{1}{2}$.

Fubini, or an ergodic theorem? We will need to arrange that $u(s)x$ and $u(s)y$ are in K. Since K has large measure, for μ_x^{13} most $s \in [-S, S]$ and most x, we have that $u(s)x \in K$. This much is seems to be just Fubini. However, since actually the S is specified after the x, a purpose built maximal ergodic theorem is used. The end result is that we may assume that for μ_x^{13} most $s \in [-S, S]$, we have that $u(s)x$ and $u(s)y$ are in K.

Showing that of μ_x^{13} isn't concentrated in the two small intervals. We need to pick $s \in [-S, S]$ which is not in the two small intervals, but which is in the set of large μ_x^{13} measure $\{s : u(s)x, u(s)y \in K\}$. This is possible unless most of the μ_x^{13} measure is concentrated in the two intervals, i.e. unless the complement of these two intervals has very small μ_x^{13} measure.

For example, it would be a problem if μ_x^{13} has a dirac mass in one of the intervals. Since μ_x^{13} contributes positive entropy, we know that this doesn't happen, as the identity (i.e. $s = 0$) is not isolated in the support.

It would be good if there leaf-wise measures were very regular. For example imagine there was a doubling condition: for some $\rho \in (0, 1)$ and all $T > 0$,

$$\mu_x^{13}((-\rho T, \rho T)) \leq \frac{1}{2}\mu_x^{13}((-T, T)).$$

With T about $S/(10\rho)$, this would say that the two intervals, each of size about say $S/10$, could not use up all the μ_x^{13} in $[-S, S]$.

The doubling condition above is probably too much to ask for. However, two things are in our favor. First, we have some freedom in picking x, and in particular we could replace it with $a(r)x$, where

$$a(r) = \begin{pmatrix} r^{-1} & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & r^3 \end{pmatrix}.$$

Second, we just saw that we only need the regularity at a certain scale (because we want to say something about intervals of length $S/10$ only).

A very small amount of regularity is obvious. There is some $\rho > 0$ so that

$$\mu_x^{13}((-\rho, \rho)) < \frac{1}{2}\mu_x^{13}((-1, 1))$$

except possible on a set Z of small measure. One can apply the standard maximal ergodic theorem to the action of $a(r)$ to conclude that for most x, most $r \in [0, R]$ satisfy that

$$\mu_x^{13}((-\rho e^{2r}, \rho e^{2r})) < \mu_x^{13}((-e^{2r}, e^{2r})),$$

which is equivalent to $a(r)x \in Z^c$. (The exponents here are probably slightly off.)

It is nonetheless possible that the desired r (computed from S and ρ) does not have $a(r)x \in Z^c$. At this point there is some magic. Suppose for example that S did not change as we replace x with $a(t)x$.

In this case the desired r remains constant, and so we in fact need $a(r)(a(t)x) \in Z^c$, which is to say $a(r+t)x \in Z^c$. Since most $a(r)x \in Z^c$ this can be arranged for some t.

Once can show that even if S varies in t as x is replaced with $a(t)x$, it does so in such a way to allow this argument to be applied. To do this, one must compute S as a function of t.

This last step seems delicate, and I would be interested to know if there is a moral reason that it works out. I think you are worried about S decreasing with t, and we rely on this not happening to quickly.