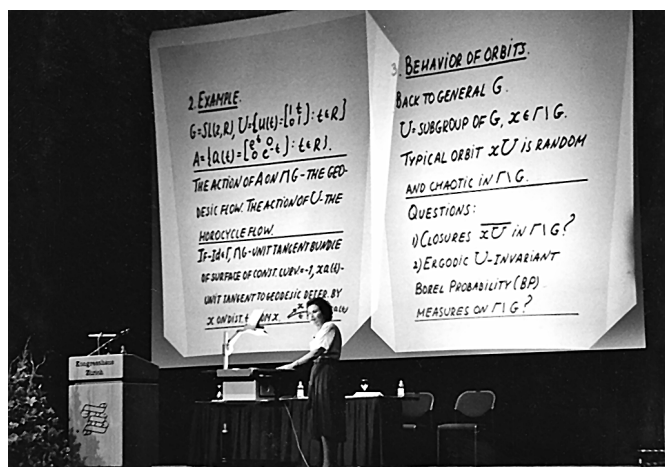


Ratner's Work on Unipotent Flows and Its Impact

Elon Lindenstrauss, Peter Sarnak, and Amie Wilkinson



Ratner presenting her rigidity theorems in a plenary address to the 1994 ICM, Zurich.

In this note we delve a bit more into Ratner's rigidity theorems for unipotent flows and highlight some of their striking applications, expanding on the outline presented by

Elon Lindenstrauss is Alice Kusiell and Kurt Vorreuter professor of mathematics at The Hebrew University of Jerusalem. His email address is elon@math.huji.ac.il.

Peter Sarnak is Eugene Higgins professor of mathematics at Princeton University and also a professor at the Institute for Advanced Study. His email address is sarnak@math.princeton.edu.

Amie Wilkinson is a professor of mathematics at the University of Chicago. Her email address is wilkinso@math.uchicago.edu.

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: <https://doi.org/10.1090/noti/1829>

Dani above. As the name suggests, these theorems assert that the closures, as well as related features, of the orbits of such flows are very restricted (rigid). As such they provide a fundamental and powerful tool for problems connected with these flows. The brilliant techniques that Ratner introduced and developed in establishing this rigidity have been the blueprint for similar rigidity theorems that have been proved more recently in other contexts.

We begin by describing the setup for the group of $d \times d$ matrices with real entries and determinant equal to 1 — that is, $SL(d, \mathbb{R})$. An element $g \in SL(d, \mathbb{R})$ is *unipotent* if $g - 1$ is a nilpotent matrix (we use 1 to denote the identity element in G), and we will say a group $U < G$ is unipotent if every element of U is unipotent. Connected unipotent subgroups of $SL(d, \mathbb{R})$, in particular one-parameter unipotent subgroups, are basic objects in Ratner's work. A unipotent group is said to be a *one-parameter unipotent group* if there is a surjective homomorphism defined by polynomials from the additive group of real numbers onto the group; for instance

$$u(t) = \begin{pmatrix} 1 & t \\ & 1 \end{pmatrix} \quad \text{and} \quad u(t) = \begin{pmatrix} 1 & t & t^2/2 \\ & 1 & t \\ & & 1 \end{pmatrix}.$$

In both cases it is easy to verify directly that these polynomials do indeed define a homomorphism: i.e., for any $s, t \in \mathbb{R}$ it holds that $u(t + s) = u(t) \cdot u(s)$. While there is essentially no loss of generality in discussing only the case of $SL(d, \mathbb{R})$, a more natural context is that of linear algebraic groups — subvarieties of $SL(d, \mathbb{R})$ defined

by polynomial equations that are closed under multiplications and taking inverses (this notion actually makes sense for more general fields than the real numbers; if we want to emphasize that we are working with the field of real numbers we will call such groups linear algebraic groups *over* $\mathbb{R}$). Connected unipotent subgroups of $SL(d, \mathbb{R})$ are always linear algebraic groups. Another nice class of examples are the orthogonal groups. Given a quadratic form $Q(\mathbf{x})$ over $\mathbb{R}$ (positive definite or not) in d variables, one can consider the group $SO(Q)$ of all matrices in $SL(d, \mathbb{R})$ that preserve this form, i.e. $d \times d$ -matrices M so that $Q(M\mathbf{x}) = Q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$. This group will be compact if and only if Q is a positive definite or a negative definite form.

Ratner's theorems on rigidity of unipotent group actions deal with the action of a unipotent group U on a quotient space of G by a discrete subgroup. An important example of such a quotient space is when $G = SL(d, \mathbb{R})$ and $\Gamma = SL(d, \mathbb{Z})$, in which case G/Γ can be identified with the space lattices in $\mathbb{R}^d$ that have unit covolume. A lattice in $\mathbb{R}^d$ can be specified by giving d linearly independent vectors that generate it — i.e. vectors $v_1, \dots, v_d \in \mathbb{R}^d$ (that we prefer to think of as column vectors) so that $\Lambda = \mathbb{Z}v_1 + \dots + \mathbb{Z}v_d$, and the condition that the lattice has unit covolume amounts to requiring that $\det(v_1, \dots, v_d) = 1$, or in other words that the matrix $g = (v_1, \dots, v_d)$ obtained by joining together these d vectors be in $SL(d, \mathbb{R})$. The generators of the lattice Λ are not uniquely determined: $v'_1, \dots, v'_d$ generate the same lattice as $v_1, \dots, v_d$ if and only if $(v'_1, \dots, v'_d) = (v_1, \dots, v_d)\gamma$ for $\gamma \in SL(d, \mathbb{Z})$, in other words, lattices of unit covolume in $\mathbb{R}^d$ are in one-to-one correspondence with elements of $SL(d, \mathbb{R})/SL(d, \mathbb{Z})$. Any matrix $h \in SL(d, \mathbb{R})$ acts on this space by left multiplication; in terms of lattices this amounts to the map from the space of unit covolume lattices to itself taking a lattice $\Lambda < \mathbb{R}^d$ to the lattice $\{h.v : v \in \Lambda\}$.

This quotient space has the important property of having finite volume, or more precisely an $SL(d, \mathbb{R})$ -invariant probability measure. A subgroup Γ of a topological group G which is discrete and such that G/Γ has finite volume is called a lattice (admittedly, this can be a bit confusing at first since our basic example of such G/Γ is the space of lattices in $\mathbb{R}^d$, though this terminology is consistent). Hermann Minkowski seems to have been the first to realize the importance of such quotients, and in particular the space of lattices in $\mathbb{R}^d$, to number theory at the turn of the 19th century. In the introduction to his book *Geometrie der Zahlen*, Minkowski writes¹

This book contains a new kind of applications of analysis of the infinite to the theory of numbers or, better, creates a new bond between these two areas. . . Geometry

¹Translated from the original German to English.

of Numbers is how I have called this book, since I arrived at the methods, which deliver in it proofs of arithmetic theorems, through spatial considerations.

Ratner's work is a remarkable contribution in the general theme of applying "analysis of the infinite" and "spatial considerations" to number theory.

So what did Ratner prove in these remarkable papers? Perhaps the easiest to explain is her Orbit Closure Classification Theorem, confirming an important conjecture of M. S. Raghunathan:

Theorem 1 (Ratner's Orbit Closure Theorem [M3]). *Let G be a real linear algebraic group as above, Γ a lattice in G and $U < G$ a connected unipotent group. Then for any point $x \in G/\Gamma$ the closure of its U -orbit is a very nice object: a single orbit of some closed connected group L that is sandwiched between U and G (and may coincide with either). Moreover, this single orbit of L has finite volume.*

Recall that the U -orbit of a point x is simply the set $\{u.x : u \in U\}$. Note that in particular this shows that any U -orbit closure has a natural U -invariant probability measure attached to it. We also remark that one can loosen the requirement that U be unipotent to U being generated by one-parameter unipotent groups — the passage from Theorem 1 to this more general statement is not very difficult. Unlike previous work towards Raghunathan's Conjecture, in particular Margulis' proof in the mid 1980s of the (then) fifty year old Oppenheim Conjecture using a special case of Raghunathan's Conjecture, Ratner's route to classifying orbit closures was not direct but by via a measure classification result:

Theorem 2 (Ratner's Measure Classification Theorem [M2, M1]). *Let G , Γ and U be as in Theorem 1. Then the only (Borel) probability measures on G/Γ that are invariant and ergodic under U are the natural measures on the orbit closures described in Theorem 1.*

This requires a bit of explanation: We equip $X = G/\Gamma$ with the Borel σ -algebra $\mathcal{B}$, and consider probability measures on the measurable space $(X, \mathcal{B})$. Such a measure μ is U -invariant if the push forward of it under left multiplication by every $u \in U$ remains the same; μ is U -ergodic if every U -invariant Borel subset of X is either null or conull. Every U -invariant probability measure can be presented as an average of ergodic ones, hence classifying the U -ergodic measures gives a description of all U -invariant probability measures on X . Dani conjectured this measure classification result in the same paper where Raghunathan's Conjecture first appeared.

It is possible to reduce both Theorem 1 and Theorem 2 to the case where U is a one-parameter unipotent group. The following theorem implies both of the theorems quoted

above in the one-parameter case, but is used by Ratner as a bridge allowing her to pass from the measure classification theorem (which, as we said in the outset, is the heart of her work on unipotent flows) to the orbit closure theorem:

Theorem 3 (Ratner's Distribution Rigidity Theorem [M3]). *Let G , Γ , U be as above, and let $x \in G/\Gamma$. Then there is a U -ergodic probability measure m_x of the form given above (i.e. the uniform measure on a finite volume orbit of a connected group sandwiched between U and G) so that x is in the support of m_x and for any bounded continuous function f on G/Γ we have that the ergodic averages*

$$\frac{1}{T} \int_0^T f(u(t).x) dt \rightarrow \int f dm_x \quad \text{as } T \rightarrow \infty. \quad (0.1)$$

The reader with some basic knowledge of ergodic theory might be fooled to think that (0.1) is an application of the Birkhoff Pointwise Ergodic Theorem. Not so! The Birkhoff Pointwise Ergodic Theorem only gives information about *almost every* point (with respect to a given ergodic measure). The whole point of Ratner's Distribution Rigidity Theorem is that it is true for each and every $x \in G/\Gamma$. Almost everywhere results are almost always much easier to prove,² but in a mathematical manifestation of Murphy's Law, such results might say something about virtually all points but if you are given a specific point and want to study its behaviour under a given action they tell you absolutely nothing. To give a simple analogy, it is trivial to prove that for a.e. $x \in [0, 1]$ the asymptotic density of occurrence of each of the digits $0, 1, 2, \dots, 9$ in the decimal expansion of x is $1/10$, but asking whether this holds for particular numbers of interest such as $2^{1/3}$ or π seems at present to be a hopelessly difficult question!

As it turns out, for some of the most juicy applications of these rigidity results a more general setup is required. To begin with, one may consider linear algebraic groups over other fields; and since the topological structure is very much in play here, the natural class of fields to look at are local fields, i.e. topological fields whose topology is locally compact, such as $\mathbb{R}$ or the p -adic numbers $\mathbb{Q}_p$. Both Ratner [M4] and independently Margulis and Tomanov [GAGM] extended the above results to this setting, and more generally to quotients G/Γ where $G = \prod_{i=1}^k G_i$ with each G_i a linear algebraic group over a local field of characteristic zero.³ We shall refer to such quotient spaces G/Γ as *S-arithmetic* quotients, a terminology that probably needs some explanation which we omit to avoid too much of a

digression. It would have been interesting to have such rigidity results also for local fields of positive characteristic such as $\mathbb{F}_q((t))$ — the field of formal Laurent series with coefficients in the finite field $\mathbb{F}_q$ with q elements — but there seem to be serious technical obstacles to doing so and only partial results in this direction are known.

The rigidity theorems of Ratner have had numerous applications in many areas of mathematics. A highly non-trivial special case of her general measure classification result, namely the classification of measures on a reducible product $(\mathrm{SL}(2, \mathbb{R})/\Gamma_1) \times (\mathrm{SL}(2, \mathbb{R})/\Gamma_1)$ invariant under a one-parameter unipotent group (the interesting case is classifying measures that project to the uniform measure on each $(\mathrm{SL}(2, \mathbb{R})/\Gamma_i)$ factor, or in the ergodic theoretic terminology, *joinings*) was proved by Ratner already in the early 1980s. The original motivation of Ratner in studying these flows was to understand better (and give natural examples for) a property of measure preserving systems called Loosely Bernoulli — we can view this somewhat anachronistically as an application of unipotent flows to the abstract theory of dynamical systems. Since then her work has had several other applications to abstract ergodic theory and descriptive set theory. There are very striking applications of her work to mathematical physics, for instance in the work of Marklof and Strömbergsson on the Lorentz gas, and to geometry. In this note we have chosen to highlight a couple of the many applications of her theorems (as well as the extension to products of linear groups over local fields as above) to number theory.

In making his famous conjecture, Raghunathan was motivated by the connection to the Oppenheim Conjecture, a connection that allowed Margulis to resolve this long-standing open problem by establishing a special case of the conjecture posed by Raghunathan [G]. Oppenheim conjectured in the 1930s that for any indefinite quadratic form Q in $d \geq 3$ variables that is not proportional to a quadratic form with integer coefficients, the set of values attained by Q at integer vectors, that is to say $Q(\mathbb{Z}^d)$, contains zero as a *non-isolated* point. Using Ratner's Measure Classification Theorem, and relying upon prior work by Dani and Margulis, Eskin, Margulis, and Mozes [AGS] were able not only to show that there are integer vectors $\mathbf{n} \in \mathbb{Z}^d$ for which $Q(\mathbf{n})$ is close to a given value (say 0), but to *count* the number of such vectors. More precisely, for indefinite quadratic forms as above, not of signature (1,2) or (2,2), Eskin, Margulis, and Mozes show that for any $a < b$, the number of integer vectors $\mathbf{n} \in \mathbb{Z}^d$ inside a ball of radius R for which $a < Q(\mathbf{n}) < b$ is asymptotically given by the volume of the corresponding shape cut by the two hyper-surfaces $Q(\mathbf{x}) = a$ and $Q(\mathbf{x}) = b$ in this ball. Perhaps an illustration of the delicacy of the question is that this natural statement is false(!) for quadratic forms of signature

²This is a slight pun—"almost everywhere" is used in the above sentence in its precise mathematical sense, whereas "almost always" is used in the ordinary, non-mathematical sense of the phrase...

³Note that our definitions of unipotent groups and one-parameter unipotent groups make sense over any field, and can be easily extended to the product case, e.g. a subgroup $U < \prod_{i=1}^k G_i$ (with each G_i defined over a different local field) is a one-parameter unipotent group if there is an i so that U is a one-parameter unipotent subgroup of G_i .

(1,2) or (2,2), though in a follow-up paper Eskin, Mozes, and Margulis were able to prove this estimate for quadratic forms of signature (2,2) under a suitable Diophantine condition, a result which is of interest in the context of the study of the statistics of energy levels of quantization of integrable dynamical systems.

The reason unipotent dynamics is relevant to the Oppenheim Conjecture is that the symmetry group of an indefinite (real) quadratic form with ≥ 3 variables contains (indeed, is generated by) one-parameter unipotent groups. Surprisingly, there is a relatively recent application of the S -arithmetic analogue of Ratner's results to *positive definite, integral* forms.

Legendre's Three Squares Theorem says that a positive integer n can be presented as a sum of three squares if and only if it is not of the form $4^a(8b+7)$, with a, b integers. This is an example of a local-to-global principle: the quadratic form $Q(x, y, z) = x^2 + y^2 + z^2$ represents an integer n if and only if the congruences $Q(x, y, z) \equiv n \pmod{p^a}$ are solvable for any prime p and any $a \in \mathbb{N}$ (for a given p , consistency of this infinite set of congruences is equivalent to $Q(x, y, z) = n$ being solvable by p -adic integers). In this particular case, only the prime 2 can be an obstacle though there is another restriction on n implicit in the way that we set up the problem — that n is positive — which can be said to come from the “place at infinity,” in other words from the necessity that $Q(x, y, z) = n$ be solvable over $\mathbb{R}$.

Legendre's Three Squares Theorem can be viewed as a special case of the following problem: Given a fixed positive definite integral quadratic form Q in many (say k) variables, which quadratic forms Q' in $\ell < k$ variables can be represented by Q ? That is to say, when can we find a $k \times \ell$ integer matrix M so that as quadratic forms $Q' = Q \circ M$? For $\ell = 1$ and $Q = x^2 + y^2 + z^2$ this reduces to the question addressed by Legendre: the form $Q' = nx^2$ can be represented by Q iff n can be written as a sum of three squares. Local solvability — the existence of such matrix M with entries in $\mathbb{Z}_p$ for every p — is an obvious necessary condition that can be verified with a finite calculation.

Hsia, Kitaoka, and Kneser in 1978 established the validity of such a local-to-global principle for representing any form Q' in ℓ variables with sufficiently large square free discriminant by a given form Q in k variables once $k \geq 2\ell + 3$ by using more traditional number theoretic methods. This remained the best result on this very classical problem (essentially dating back to the work of Gauss) until Ellenberg and Venkatesh [JA] were able to use the S -arithmetic extensions to Ratner's Orbit Closure Theorem to very significantly reduce the restriction on k and ℓ to be $k \geq \ell + 5$. While we cannot get into the details of the

argument, we note that even if a quadratic form Q is positive definite, hence its symmetry group over $\mathbb{R}$ is compact, over the p -adic numbers in general for $k \geq 3$ variables it would be a non-compact group with plenty of unipotents. In truth, the relevant symmetry group for this case is not the symmetry group of Q but the subgroup of this symmetry group fixing a given quadratic form in ℓ variables, but this is precisely why in this problem one needs to employ p -adics.

An even more surprising application of Ratner's work to number theory was given by Vatsal and Cornut–Vatsal (e.g. [V]). We do not give details here, but in these works families of elliptic curve L -functions, and in particular their central values (or derivatives thereof when their functional equation is odd rather than even), are considered. Using Ratner's Orbit Closure Theorem as a basic ingredient Vatsal (and in the more general cases Cornut and Vatsal) showed that all but finitely many of these values are not zero. When combined with well-known results towards the Birch and Swinnerton–Dyer Conjecture, this proves a conjecture of Mazur: essentially all the points on an elliptic curve whose coordinates lie in ring class fields with restricted ramification are generated by explicit special points first constructed by Heegner.

The impact of Ratner's work cannot be measured only by direct application of her seminal results. Techniques introduced by Ratner to study ergodic theoretic joinings in her early works on unipotent flows in the 1980s were a main inspiration in the work of the first named author on diagonalizable flows and its applications to Arithmetic Quantum Unique Ergodicity and equidistribution. Benoist and Quint were similarly inspired by Ratner's work in their breakthrough work understanding stationary measures and orbit closures for actions of thin groups on homogeneous spaces, and Eskin and Mirzakhani transformed the study of moduli spaces of abelian and quadratic differentials on Riemann surfaces by proving an analogue of Ratner's work in this setting.

The prevalence of deep and surprising applications of Ratner's Rigidity Theorems on unipotent flows is remarkable, and shows the richness of the subject of homogeneous dynamics and how interconnected it is with many other subjects. It is also a tribute to a wonderful mathematician who has left a legacy to future mathematicians for many years to come.

References

- [AGS] Eskin A, Margulis G, Mozes S. Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture, *Ann. of Math.* (2), no. 1 (147):93–141, 1998, DOI [10.2307/120984](https://doi.org/10.2307/120984), [MR1609447](https://doi.org/10.2307/1609447)

- [JA] Ellenberg J, Venkatesh A. Local-global principles for representations of quadratic forms, *Invent. Math.*, no. 2 (171):257–279, 2008, DOI [10.1007/s00222-007-0077-7](https://doi.org/10.1007/s00222-007-0077-7). [MR2367020](#)
- [G] Margulis G. Discrete subgroups and ergodic theory, *Number theory, trace formulas and discrete groups* (Oslo, 1987); 1989:377–398. [MR993328](#)
- [GAGM] Margulis G A, Tomanov G M. Invariant measures for actions of unipotent groups over local fields on homogeneous spaces, *Invent. Math.*, no. 1-3 (116):347–392, 1994, DOI [10.1007/BF01231565](https://doi.org/10.1007/BF01231565). [MR1253197](#)
- [M1] Ratner M. On measure rigidity of unipotent subgroups of semisimple groups, *Acta Math.*, no. 3-4 (165):229–309, 1990, DOI [10.1007/BF02391906](https://doi.org/10.1007/BF02391906). [MR1075042](#)
- [M2] Ratner M. On Raghunathan’s measure conjecture, *Ann. of Math. (2)*, no. 3 (134):545–607, 1991. [MR1135878](#)
- [M3] Ratner M. Raghunathan’s topological conjecture and distributions of unipotent flows, *Duke Math. J.*, no. 1 (63):235–280, 1991, DOI [10.1215/S0012-7094-91-06311-8](https://doi.org/10.1215/S0012-7094-91-06311-8). [MR1106945](#)
- [M4] Ratner M. Raghunathan’s conjectures for Cartesian products of real and p -adic Lie groups, *Duke Math. J.*, no. 2 (77):275–382, 1995, DOI [10.1215/S0012-7094-95-07710-2](https://doi.org/10.1215/S0012-7094-95-07710-2). [MR1321062](#)
- [V] Vatsal V. Uniform distribution of Heegner points, *Invent. Math.*, no. 1 (148):1–46, 2002, DOI [10.1007/s002220100183](https://doi.org/10.1007/s002220100183). [MR1892842](#)

Credits

Photo of Marina Ratner is courtesy of Anna Ratner.