L. Dever and D. Milićević (2021) "Ambient Prime Geodesic Theorems on Hyperbolic 3-Manifolds," International Mathematics Research Notices, Vol. 00, No. 0, pp. 1–48
https://doi.org/10.1093/imrn/rnab048

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

Ambient Prime Geodesic Theorems on Hyperbolic 3-Manifolds

Lindsay Dever[1] and Djordje Milićević[1,2,*]

[1]Bryn Mawr College, Department of Mathematics, 101 North Merion Avenue, Bryn Mawr, PA 19010, USA and [2]Max-Planck-Institut für Mathematik, Vivatsgasse 7, D-53111 Bonn, Germany

[*]*Correspondence to be sent to: e-mail: dmilicevic@brynmawr.edu*

We prove prime geodesic theorems counting primitive closed geodesics on a compact hyperbolic 3-manifold with length and holonomy in prescribed intervals, which are allowed to shrink. Our results imply effective equidistribution of holonomy and have both the rate of shrinking and the strength of the error term fully symmetric in length and holonomy.

1 Introduction

Closed geodesics on a smooth and connected Riemannian manifold M act as important geometric and dynamical invariants. Closed geodesics support periodic orbits of the geodesic flow and in turn its invariant measures, whereas the length of the shortest closed geodesic of M (its systole) acts as the first threshold of global geometry and dynamics. On locally symmetric spaces, the trace formula connects closed geodesics to the spectrum of the Laplacian (which quantizes the geodesic flow), just as elliptic elements of Fuchsian groups enter dimensions of spaces of cusp forms. In arithmetic cases, lengths and multiplicities of geodesics can often be explicitly related to invariants such as class numbers and regulators of indefinite binary quadratic forms. Hence, for many

Received September 22, 2020; Revised February 1, 2021; Accepted February 16, 2021

© The Author(s) 2020. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

reasons, one seeks to understand the set of closed geodesics on M and, in particular, as in this paper, its size and structure.

1.1 Prime Geodesic Theorems and Holonomy

We will be concerned with compact hyperbolic 3-manifolds M arising as the quotient $M \simeq \Gamma\backslash\mathbb{H}^3$ of the hyperbolic upper half-3-space by a uniform, torsion-free lattice $\Gamma \subseteq G = \mathrm{PSL}_2\mathbb{C}$. Closed geodesics C_γ on M arise from non-identity conjugacy classes $[\gamma]$ in Γ, with primitive classes corresponding to infinitely many *prime* geodesics of increasing lengths $\ell(\gamma) \to \infty$. The celebrated Prime Geodesic Theorem, in the form with an explicit error term, is due to Sarnak [23, Theorem 5.1] and may be stated as

$$\pi_\Gamma(x) := \left|\left\{[\gamma] \text{ primitive in } \Gamma : \ell(\gamma) \leqslant x\right\}\right| = \mathrm{Ei}_\Gamma(x) + \mathrm{O}_{\Gamma,\epsilon}\left(e^{(\frac{5}{3}+\epsilon)x}\right), \tag{1}$$

where the main term $\mathrm{Ei}_\Gamma(x)$ may be defined in terms of an absolutely continuous density measure $\mathrm{d}\varpi_\Gamma$ as

$$\mathrm{Ei}_\Gamma(x) = \int_2^x \mathrm{d}\varpi_\Gamma(t), \quad \mathrm{d}\varpi_\Gamma(t) = \left(\frac{e^{2t}}{t} + \sum_{j=1}^{k} \frac{e^{(1+\nu_j)t}}{t}\right)\mathrm{d}t, \tag{2}$$

and $\{0 < \nu_k \leqslant \cdots \leqslant \nu_1 < 1\}$ correspond to the eigenvalues $1 - \nu_j^2$ of the hyperbolic Laplacian Δ on M in $(0,1)$ as described in Section 2.1. The set of $\{\nu_j\}$ depends on Γ only and is predicted by Selberg's conjecture to be empty for arithmetic Γ. In any case, $\mathrm{Ei}_\Gamma(x) \sim_\Gamma \mathrm{Ei}(2x) = \int_2^{2x} e^t/t\,\mathrm{d}t \sim e^{2x}/2x$, recovering the general asymptotics of [19, Theorems 3, 4] and [10, Proposition 5.4]. Moreover, (1) gives a power-saving asymptotic for $\pi_\Gamma(x, x+h) := \pi_\Gamma(x+h) - \pi_\Gamma(x)$ as long as $h \gg e^{-(1/3-\delta)x}$.

Attached to each closed geodesic C_γ is the geometric action of the associated class $[\gamma]$. For $\gamma \sim \begin{pmatrix} e^{(\ell+i\theta)/2} & 0 \\ 0 & e^{-(\ell+i\theta)/2} \end{pmatrix}$ with $\ell > 0$ (recall that Γ is uniform), this is given by a shift by $\ell = \ell(\gamma)$ along C_γ and a rotation around C_γ by the angle θ (which corresponds to parallel transport along C_γ). Thus, each C_γ carries two geometric invariants: the length $\ell(\gamma)$ and the *holonomy* $\mathrm{hol}(\gamma) := \theta \in \mathbb{R}/2\pi\mathbb{Z}$; in this setting of 3-manifolds one often also talks about the *complex length* $\mathbb{C}\ell(\gamma) := \ell(\gamma) + i\,\mathrm{hol}(\gamma)$. It becomes a natural counting question to refine the count (1) according to holonomy. The equidistribution theorem of Sarnak–Wakayama [25, Theorem 1, Corollary 1] in the present case of a compact 3-manifold $M \simeq \Gamma\backslash\mathbb{H}^3$ states that, for every interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$,

$$\pi_\Gamma(x; J) = \left|\left\{[\gamma] \text{ primitive in } \Gamma : \ell(\gamma) \in [0, x],\, \mathrm{hol}(\gamma) \in J\right\}\right| \sim_\Gamma \frac{|J|}{2\pi}\pi_\Gamma(x) \quad (x \to \infty). \tag{3}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

We will be interested in asymptotics with precise control on $\mathrm{hol}(\gamma)$, the "compact part of the complex length," in the same way as the sharp cutoff in $\ell(\gamma)$ in the Prime Geodesic Theorem (1). In particular, our Theorem 6.5 proves the following effective version of the equidistribution result (3):

$$\pi_\Gamma(x;J) = \frac{|J|}{2\pi}\pi_\Gamma(x) + \mathrm{O}_\Gamma\big(e^{5x/3}\big), \tag{4}$$

with the implied constant *independent of* J. We also remark that the error term in (4) is a pure exponential. Using dynamical methods, Margulis, Mohammadi, and Oh [20, 21, Theorem 1.3] proved an asymptotic for $\pi_\Gamma(x;J)$ with a small, unspecified power savings in (3), for a broad class of geometrically finite, Zariski dense $\Gamma < \mathrm{PSL}_2(\mathbb{C})$, including all lattices.

The asymptotic (4) provides power savings for the refined count with holonomy in an interval of length $|J| \gg e^{(-\frac{1}{3}+\delta)x}$ and is reminiscent of a prime number theorem in arithmetic progressions with explicit level dependence. Indeed, we deduce it from the following uniform estimate on "holonomy character sums" (see Proposition 5.1):

$$K_n(x) := \sum_{[\gamma] \text{ primitive in } \Gamma : \ell(\gamma) \leqslant x} e^{in\,\mathrm{hol}(\gamma)} \ll_\Gamma x^{-1}e^{5x/3} + n^2 e^x \quad (n \neq 0), \tag{5}$$

where the implied constant depends only on Γ. The decomposition into pure harmonics of holonomy in $\mathbb{R}/2\pi\mathbb{Z}$ implicit in the passage between (4) and (5) (which was also the key implement in [25]) gives analytic access to automorphic constituents of $L^2(\Gamma\backslash G)$ via the trace formula; see Section 1.2. Substantial cancellation in holonomy character sums in (5), which quickly leads to (4), is in this sense analogous to classical bounds on sums of Dirichlet characters with explicit conductor dependence.

Theorem 6.5 requires fine control in both $\ell(\gamma)$ and $\mathrm{hol}(\gamma)$. More broadly, we argue that, for many purposes including counting in short ranges, the two geometric parameters $\ell(\gamma)$ and $\mathrm{hol}(\gamma)$ have the same standing, and that it is most natural to talk about the *joint distribution* of the pair $(\ell(\gamma), \mathrm{hol}(\gamma))$. Such a result might be called an *ambient prime geodesic theorem*. In Theorem 6.12, we obtain our main result, an asymptotic count for primitive closed geodesics on M according to the pair $(\ell(\gamma), \mathrm{hol}(\gamma))$.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Theorem 1.1. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for any intervals $I \subseteq [0, x]$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$,

$$\begin{aligned}\pi_\Gamma(I,J) &:= \left|\left\{[\gamma] \text{ primitive in } \Gamma : (\ell(\gamma), \mathrm{hol}(\gamma)) \in I \times J\right\}\right| \\ &= \iint\limits_{I\times J} \mathrm{d}\varpi_\Gamma(t)\,\frac{\mathrm{d}\theta}{2\pi} + \mathrm{O}_\Gamma\Big((|I|+|J|)^{2/3}\frac{e^{5x/3}}{x^{2/3}} + \frac{e^{3x/2}}{x^{1/2}}\Big).\end{aligned}$$

In particular, Theorem 1.1 proves the uniform asymptotic (see Theorem 6.5 and its Corollary 6.7)

$$\pi_\Gamma(I,J) = \iint\limits_{I\times J} \mathrm{d}\varpi_\Gamma(t)\,\frac{\mathrm{d}\theta}{2\pi} + \mathrm{O}_\Gamma\big(e^{5x/3}\big), \tag{6}$$

which provides a power savings as long as $|I \times J| \gg e^{(-\frac{1}{3}+\delta)x}$ with $\delta > 0$, where each interval may be short independently of each other; see Remark 6.8. Theorem 1.1 further extends this range when both intervals I and J are short, down to as short as $|I \times J| \gg e^{(-\frac{1}{2}+\delta)x}$ when $|I| \asymp |J|$; see Remark 6.13 for details.

Joint distribution results such as our Theorem 1.1 may be seen as instances of spectral geometry on the group quotient $\Gamma\backslash G$, as we explain in Section 1.2. We emphasize that our results apply regardless of whether the subgroup Γ is arithmetic or not; for asymptotics on arithmetic quotients in the length aspect, we refer to [25] as well as to recent advances on arithmetic hyperbolic 3-manifolds [1–3]. While it would certainly be of interest to obtain stronger error terms in ambient prime geodesic theorems on arithmetic hyperbolic 3-manifolds, our goal here is to establish universal, baseline results.

1.2 Spectral Geometry of $\Gamma\backslash G$, Trace Formulas, and Ambient Counting

Closed geodesics on M are often considered along with the spectrum of the Laplace–Beltrami operator Δ_M, which we recall involves averages over infinitesimal balls in M and thus naturally quantizes the dynamics of the geodesic flow on M. On a rank one compact locally symmetric space $M = \Gamma\backslash S$, the classical spherical trace formula relates the eigenvalues of Δ_M (that is, frequencies of Δ_S appearing in $L^2(\Gamma\backslash S)$) with lengths of geodesics corresponding to non-identity conjugacy classes in Γ. This may also be seen as the correspondence principle of quantum mechanics, relating long-term dynamics on M with the semi-classical (high-energy) limit of the quantized system on $L^2(M)$, or as a noncommutative version of Fourier duality.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

A key question of spectral geometry is whether isospectral manifolds (having the same Laplacian spectrum) are also isometric. For hyperbolic 3-manifolds, the answer is "No": in 1980, Vignéras [28] found a pair of hyperbolic 3-manifolds that are isospectral but not isometric; see also [11, 27]. However, Gangolli [9] showed that for a compact hyperbolic 3-manifold, the Laplacian spectrum determines the set of lengths of closed geodesics. While for hyperbolic surfaces the Laplacian spectrum also determines the multiplicities of closed geodesics, in higher dimensions this is an open problem [14, Remark 0.3]. In the converse direction, Kelmer [14] showed that the length spectrum (including multiplicities) determines the Laplacian spectrum for compact hyperbolic manifolds. For arithmetic hyperbolic 3-manifolds, the complex length spectrum (and in fact, the rational length set) determines the commensurability class [5, 22].

In any case, *all structure* encoded in the trace formula should be reflected in the spectral-to-geometric correspondence. In the present case of $M \simeq \Gamma\backslash \mathrm{PSL}_2\mathbb{C}/\mathrm{PSU}_2$, regular conjugacy classes $[\gamma]$ are parametrized by $t_\gamma \in T/S_2 = \{t = \left(\begin{smallmatrix} z & 0 \\ 0 & z^{-1} \end{smallmatrix}\right) : z \in \mathbb{C}^\times\}/(t \sim t^{-1})$, $L^2_0(\Gamma\backslash \mathrm{PSL}_2\mathbb{C})$ is spanned by principal series representations $\pi = \mathrm{ind}_T^G(\chi_{\nu,p})$ indexed by unitary characters $\chi_{\nu,p} : T \to S^1$, and the trace formulas on $\Gamma\backslash \mathrm{PSL}_2\mathbb{C}$ (Theorems 2.1, 2.2, and 2.4, below) relate roughly, up to fixed smooth weights,

$$\sum_{\pi_{\nu,p} \subseteq L^2(\Gamma\backslash G)} \hat{F}(\chi_{\nu,p}^{-1}) + \dots \quad \leftrightsquigarrow \quad \sum_{[\gamma] \subset \Gamma} F(t_\gamma) + \dots, \tag{7}$$

for a compactly supported smooth function $F : T/S_2 \to \mathbb{C}$ and its Abel transform $\hat{F} : T^*/S_2 \to \mathbb{C}$; see Section 2.2 for details. Our guiding principle, then, is to view the spectral geometry of $\Gamma\backslash\mathbb{H}^3$ as duality between the classical spherical Maass forms and counts such as the classical Prime Geodesic Theorem (1), essentially specializing (7) to $(\mathrm{PSU}_2 \cap T)$-invariant test functions, and the spectral geometry of $\Gamma\backslash G$ as the full duality, as encoded by (7), between the entire spectrum of all PSU_2-types and ambient prime geodesic theorems such as (4) and our Theorem 1.1.

As an imperfect but convincing analogy, consider a higher rank real symmetric space, say a compact quotient $\Gamma\backslash G/K = \Gamma\backslash \mathrm{SL}_n\mathbb{R}/\mathrm{SO}_n$, with rank $r = n - 1$, where $L^2_0(\Gamma\backslash G/K)$ decomposes into a direct sum of principal series representations induced from a character ν in the dual of the Lie algebra of the maximal torus $\mathfrak{a}^* \simeq \mathbb{R}^r$. Weyl's law is a classical spectral count, central for quantum mechanics or thermodynamics, of eigenforms in $L^2(\Gamma\backslash G/K)$ with Laplace eigenvalue up to a given bound, or equivalently with $\|\nu\|_2 \leqslant X$. From the point of view of equidistribution in families of automorphic forms [4, 16, 24], one is more broadly interested in counting representations $\pi \subseteq L^2(\Gamma\backslash G/K)$ in a prescribed region $\Omega \subseteq \widehat{G}$ within the *ambient* space $\widehat{G} \simeq \mathfrak{a}^*_{\mathbb{C}}$ of all such representations; a count with r parameters rather than just one (see [4, Proposition 7.2]).

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

The natural dual question to this is to count primitive conjugacy classes $[t_\gamma]$ occurring in Γ not just according to their length $\|\log|t_\gamma|\|_2$ but simply within prescribed regions in the r-dimensional space $[G] \simeq T/S_n$ of all regular conjugacy classes of G (see [6]).

An ambient prime geodesic theorem in $\mathrm{PSL}_2\mathbb{C}$ such as our Theorem 1.1 should similarly count primitive conjugacy classes in Γ according to their full parameter, the complex length $\mathbb{C}\ell(\gamma) \in \mathbb{C}/\{\pm 1\}$. In a different context, Kelmer [13, Corollary 3.1] proved effective equidistribution of holonomy for closed geodesics on products of n hyperbolic planes (corresponding to conjugacy classes in an irreducible co-compact lattice $\Gamma < (\mathrm{PSL}_2\mathbb{R})^n$ that are hyperbolic at one place and elliptic at all others), including the asymptotic joint length–holonomy count.

This analogy brings about the natural question of using the non-spherical trace formula of Section 2.2 to count geodesics in more general regions $\Omega \subseteq [G]$, the geometric counterpart to the spectral count of $\pi \subseteq L^2(\Gamma\backslash G)$ in regions of $\widehat{G}$ as in [4, Proposition 7.2]. Passage to the geometric sharp cutoff count and estimates of boundary terms as in (13) involve estimating contributions on the dual (spectral) side, extending deep into the tempered spectrum $\widehat{G}^{\mathrm{temp}}$ (in all directions), which is of more moderate growth. This in turn involves the rate of decay of the Fourier transform $\widehat{\chi_\Omega}$ and thus the *shape* of Ω and its boundary, as in classical lattice-point counting. We leave this intriguing lead for future work.

1.3 Overview

For the sake of the reader, we now present an overview of the proof, omitting details. Throughout, $\Gamma < \mathrm{PSL}_2\mathbb{C}$ is a discrete, co-compact, torsion-free subgroup. We refer to Section 2.1 for background on the geometry and representation theory of $G = \mathrm{PSL}_2\mathbb{C}$ and its quotients.

One of the key tools we use is the non-spherical trace formula. We start with a version of Selberg's trace formula, explicated by Lin and Lipnowski [17, Corollary 2], which captures both the length and holonomy of geodesics. We specialize this, on the spectral side, to representations of a particular type and, on the geometric side, to a particular frequency of holonomy, and obtain Theorem 2.2, which states that for every $n \in \mathbb{Z}$ and every smooth, even, compactly supported $g : \mathbb{R} \to \mathbb{C}$,

$$\begin{aligned}\frac{1}{2}\sum_{\nu}(m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n}))\int_{-\infty}^{\infty} g(u)e^{u\nu}\,\mathrm{d}u + \delta_0(n)\int_{-\infty}^{\infty} g(u)e^{u}\,\mathrm{d}u - \frac{1}{2}\delta_{\pm 1}(n)\hat{g}(0)\\ = \frac{1}{2\pi}\,\mathrm{vol}(\Gamma\backslash G)(n^2 g(0) - g''(0)) + \sum_{[\gamma]}\ell(\gamma_0)w(\gamma)g(\ell(\gamma))\cos(n\,\mathrm{hol}(\gamma)),\end{aligned} \tag{8}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

where the first sum is over the unitary principal and complementary series representations $\pi_{\nu,\pm n}$ (where $\nu \in i\mathbb{R}$ for principal series) occurring with multiplicities $m_\Gamma(\pi_{\nu,\pm n})$ in $L^2(\Gamma\backslash G)$, $w(\gamma) \asymp_\Gamma e^{-\ell(\gamma)}$, and the last sum is over the nontrivial hyperbolic and loxodromic conjugacy classes of Γ. For purposes of counting the length spectrum, where one is typically interested in (8) with a test function g of varying and extended support, it is natural to combine the complementary series and identity terms into a single integral of g against a measure $\mathrm{d}\varpi_\Gamma^*$ related to $\mathrm{d}\varpi_\Gamma$; see (20)–(21). We also explicate a complementary "odd" trace formula, sampling geodesics with a weight $h(\ell(\gamma))\sin(n\,\mathrm{hol}(\gamma))$ for an odd $h \in C_c^\infty(\mathbb{R})$; see Theorem 2.4.

In particular, using the trace formula (8) with a specific $g \in C_c^\infty(\mathbb{R})$ that emphasizes the spectral terms with $R-1 \leqslant |\nu| \leqslant R+1$ with weight $\hat{g}(i\nu/2\pi) \gg 1$ and keeps all spectral terms non-negative, we prove in Proposition 2.5 a bound on multiplicities of representations in an interval of fixed length:

$$\sum_{R-1\leqslant|\nu|\leqslant R+1} m_\Gamma(\pi_{\nu,n}) \ll \mathrm{vol}(\Gamma\backslash G)\cdot(R^2+n^2)+\mathrm{O}_\Gamma(1). \tag{9}$$

The spectral bound (9) is a bound on local spectral densities, whose leading term agrees with the Plancherel measure. Such a bound is a standard tool in the passage from a smooth to sharp count of the spectrum in the proof of Weyl's law (cf. [4, Proposition 10.1]).

As the first step toward a sharp geodesic count, in Lemma 3.1 we use the trace formula (8) to get a handle on the geometric sum

$$T_n^{\cos}[g_{Y,\eta}] = \sum_{[\gamma]} \ell(\gamma_0)w(\gamma)g_{Y,\eta}(\ell(\gamma))\cos(n\,\mathrm{hol}(\gamma)) \tag{10}$$

with a smooth, even function $g_{Y,\eta}$, which approximates $\chi_{[-Y,Y]}$ and is supported on $[-Y-\eta, Y+\eta]$. The principal series terms in (8) are then weighted with its Fourier transform $\hat{g}_{Y,\eta}(i\nu/2\pi)$, which is essentially supported up to roughly $|\nu| \ll 1/\eta$ (exhibiting Schwartz decay past this range); we estimate these terms using (9). In Lemma 3.3, we majorize the contributions to (10) of classes with $\ell(\gamma) \in [Y-\eta, Y+\eta]$ by a suitable non-negative smooth bump function on $[Y-2\eta, Y+2\eta]$ (whose Fourier transform again extends to roughly $\asymp 1/\eta$), and show with another application of (8) and (9) that these boundary terms contribute $\mathrm{O}_\Gamma(e^Y\eta + 1/\eta^2)$ to (10). Taking the boundary length $\eta = e^{-Y/3}$, we obtain in Proposition 3.5

$$T_n^{\cos}(Y) := \sum_{\ell(\gamma)\leqslant Y} \ell(\gamma_0)w(\gamma)\cos(n\,\mathrm{hol}(\gamma)) = \delta_0(n)\int_{-Y}^{Y} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\left(e^{2Y/3}+n^2Y\right) \tag{11}$$

and a corresponding asymptotic for $T_n^{\sin}(Y)$ (and thus $T_n(Y)$ in (34)).

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Counts such as (11) arising from the trace formula naturally involve the weights $w(\gamma)$ (from the Weyl discriminant) and all conjugacy classes $[\gamma]$, including imprimitive ones. In Section 4 we address these two technical aspects and show they are essentially harmless, namely, using that $w(\gamma) = e^{-\ell(\gamma)} + O_\Gamma(e^{-2\ell(\gamma)})$ and that imprimitive classes contribute comparatively very few (namely $O_\Gamma(e^Y)$) terms to (11), we prove in Lemma 4.1 that a simpler sum $S_n^P(Y)$ over primitive geodesics satisfies

$$S_n^P(Y) := \sum_{\ell(\gamma)\leqslant Y}^{P} \ell(\gamma) e^{-\ell(\gamma)+in\operatorname{hol}\gamma} = T_n(Y) + O_\Gamma(Y). \tag{12}$$

Combining (11) and (12) and using integration by parts, we obtain in Proposition 5.1, as cited above, the estimate (5) on "holonomy character sums" $K_n(Y)$. In particular, Proposition 5.1 for $n \neq 0$ quantifies cancellation among the holonomies of primitive geodesics with $\ell(\gamma) \leqslant Y$ and shows that these are equidistributed throughout $\mathbb{R}/2\pi\mathbb{Z}$.

While we have so far in (11), (12), and (5) emphasized the traditional sharp cutoff $\ell(\gamma) \leqslant Y$, we in fact throughout also prove estimates for analogous smooth cutoff quantities like $S_n^P[g_{Y,\eta}]$ and $K_n[g_{Y,\eta}]$, with explicit dependence on η. Moreover, we observe structural analogies in the length and holonomy aspects, such as in the comparison of our Lemmata 3.3 and 3.6, which state roughly that

$$\sum_{Y-\eta\leqslant\ell(\gamma)\leqslant Y+\eta} \ell(\gamma_0) w(\gamma) \ll_\Gamma \eta e^Y + \frac{1}{\eta^2}, \qquad \sum_{\substack{\ell(\gamma)\leqslant Y \\ \theta_0-\eta'\leqslant\operatorname{hol}(\gamma)\leqslant\theta_0+\eta'}} \ell(\gamma_0) w(\gamma) \ll_\Gamma \eta' e^Y + \frac{Y}{\eta'^2}. \tag{13}$$

This analogy is fundamentally due to the fact that, in each case, the dual (spectral) sum over $\pi_{\nu,n}$ extends up to roughly $1/\eta$ or $1/\eta'$ (in the ν- and n-direction, respectively) and that the Plancherel measure shown in (9) and (23) is symmetric; see Remark 3.7.

With this in mind, in Sections 6.1 and 6.2 we prove $4 = 2\times 2$ asymptotic formulas for "ambient" prime geodesic counts, beginning with the smooth count (Proposition 6.1) of the form

$$\begin{aligned} \pi_\Gamma(g_{Y,\eta}, f) &:= \sum_{[\gamma]}^{P} f(\operatorname{hol}(\gamma)) g_{Y,\eta}(\ell(\gamma)) \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \int_2^\infty g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + O_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y\eta^2}\|\hat f\|_1 + e^Y\|\hat f\|_{2,1}\Big) \end{aligned} \tag{14}$$

for a smooth $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, $\|\hat f\|_{2,1} := \|\hat f\|_1 + \|\widehat{f''}\|_1$, and $0 < \eta \leqslant \eta_0$, and then for the related counts $\pi_\Gamma(Y, f)$, $\pi_\Gamma(g_{Y,\eta}, J)$, and $\pi_\Gamma(Y, J)$ (for $Y > 0$ and any interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$),

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

which have sharp cutoffs in the length, holonomy, and in both aspects, respectively. In particular, by spectrally expanding f into a Fourier series and estimating $K_n(y)$ using (5), in Theorem 6.2 we prove that

$$\pi_\Gamma(y,f) := \sum_{\ell(\gamma)\leqslant y}^{P} f(\mathrm{hol}(\gamma)) = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \int_2^y \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_\Gamma\Big(\|\hat{f}\|_1 \frac{e^{5y/3}}{y} + \|\widehat{f''}\|_1 e^y\Big),$$

which recovers [25, Theorem 1] for a fixed f. However, our explicit dependence on f allows us to choose a smooth $f_{J,\eta'}$ approximating a sharp holonomy cutoff, while maintaining explicit dependence on η', and then estimate the terms with holonomy within η' of the boundary of J using the second bound in (13). In exact analogy with the passage to the sharp count in (11), we choose $\eta' = e^{-y/3}$, which leads in Theorem 6.5 to the effective count (4) for a sharp length and holonomy count, and then to its consequence (6).

In Section 6.3, we prove asymptotic formulas that provide counts for the number of geodesics in intervals I and J of length and holonomy, respectively. These counts feature a combination of sharp and smooth cutoffs in the length and holonomy, in complete parallel to Section 6.2, but with improvements in the error term when the lengths of I and/or J are shrinking. After explicating in Lemma 6.9 an asymptotic analogous to (14) for the count $\pi_\Gamma(g_{I,\eta}, f_{J,\eta'})$ with suitable smooth length and holonomy cutoffs, we derive as Corollary 6.10 the upper bound

$$\sum_{\substack{\ell(\gamma)\in I\\ \mathrm{hol}(\gamma)\in J}} 1 \ll_{\Gamma,\eta_0} (|I|+\eta)(|J|+\eta')\frac{e^{2y}}{y} + \frac{e^y}{y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^y}{\eta'^2}, \tag{15}$$

for every $I \subseteq [0,y]$, $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$ and $0 < \eta \leqslant \eta_0$, $0 < \eta' \leqslant 2\pi$. This is a (normalized; recall that $w(\gamma) \sim_\Gamma e^{-\ell(\gamma)}$) ambiental analogue of (13), where we additionally profit in the first term when lengths and holonomies are sampled from short intervals. Then, we estimate the smooth count using Lemma 6.9 and bound the ambiguous regions with (15), which leads to Proposition 6.11 and our main result, Theorem 1.1.

1.4 Notation

We write $f = \mathrm{O}(g)$ or $f \ll g$ to mean that $|f| \leqslant Cg$ for some constant $C > 0$, which may be different from line to line and is absolute unless explicitly indicated with a subscript. We also write $f \asymp g$ to denote that $f \ll g \ll f$, and $f \sim g$ to denote that $\lim f/g = 1$, where the direction of the limit is clear from the context, again with dependencies of

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

implied constants and rate of convergence only as indicated. (No confusion should arise with the usage of $\sim$ to also denote conjugate elements in a matrix group.)

We use the convention $\hat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi ix\xi}\,\mathrm{d}\xi$ to denote the Fourier transform of a Schwartz function f, and we choose the normalization $\hat{f}(n) = \int_{\mathbb{R}/2\pi\mathbb{Z}} f(x)e^{-inx}\,\mathrm{d}x$ for the Fourier coefficient of a periodic function f on $\mathbb{R}/2\pi\mathbb{Z}$. We also use the shorthand notation $\log^*(x) := \log(x+2)$ for $x > 0$. Finally, we write $\delta_a(n)$ for the Kronecker delta-function, that is, $\delta_a(n) = 1$ if $n = a$ and 0 otherwise (including writing $\delta_{\pm 1}(n)$ as shorthand for whether $n = \pm 1$ or not).

2 Non-spherical Trace Formulas and Weyl's Law

2.1 Background on Groups and Representations

Let $G = \mathrm{PSL}_2\mathbb{C}$, and let Γ be a discrete, torsion-free, co-compact subgroup of G. The group $G = \mathrm{PSL}_2\mathbb{C}$ is in one-to-one-correspondence with the group of orientation-preserving isometries of the 3-dimensional, hyperbolic upper half space $\mathbb{H}^3$, which we describe shortly. We are primarily concerned with the geometry of the fundamental domain $M = \Gamma\backslash\mathbb{H}^3$, which is a compact hyperbolic 3-manifold, and its covering $\Gamma\backslash G$. In this section, we collect some background material about the group G, its geometric action on $\mathbb{H}^3$, the geometry of M, and the representation theory of $L^2(G)$ and $L^2(\Gamma\backslash G)$.

The group $\mathrm{PSL}_2\mathbb{C}$ has the Iwasawa decomposition $G = UAK$, where

$$U = \left\{ \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} : z = x + iy \in \mathbb{C},\, x, y \in \mathbb{R} \right\}, \quad A = \left\{ \begin{pmatrix} e^{u/2} & 0 \\ 0 & e^{-u/2} \end{pmatrix} : u \in \mathbb{R} \right\},$$

$$K = \mathrm{PSU}_2 = \left\{ \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C},\, |\alpha|^2 + |\beta|^2 = 1 \right\} / \{\pm 1\}.$$

Here, UA is a Borel subgroup of G with the unipotent subgroup U and with A a maximal torus in G, and K is a maximal compact subgroup of G. The Haar measure on each of these subgroups is unique up to a constant multiple. We choose the Euclidean measure $\mathrm{d}x\,\mathrm{d}y$ on U, $\mathrm{d}u$ on A and the volume 1 Haar measure $\mathrm{d}k$ on K. This induces a Haar measure on G.

The quotient G/K may be identified with the upper half space $\mathbb{H}^3 = \{z + ir : z \in \mathbb{C}, r > 0\}$, a 3-dimensional hyperbolic space with the hyperbolic metric $\mathrm{d}s = (|\mathrm{d}z|^2 + (\mathrm{d}r)^2)^{1/2}/r$. The action of G by left multiplication induces an action of G on $\mathbb{H}^3$ by orientation-preserving isometries (preserving the hyperbolic metric), which may also

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

be described in terms of (z, r)-coordinates; see [8, Section 1.1]. In fact, the group G is in one-to-one-correspondence with the group of orientation-preserving isometries of $\mathbb{H}^3$.

Elements of G can be classified into identity, parabolic, elliptic, hyperbolic, and loxodromic, each with a distinct type of geometric action on $\mathbb{H}^3$. Parabolic elements are conjugate to an element of the unipotent group U described above. All other elements are diagonalizable. Every diagonalizable element $\gamma \in G$ is conjugate to some

$$t_\gamma \in T := \left\{ \begin{pmatrix} e^{(u+i\theta)/2} & 0 \\ 0 & e^{-(u+i\theta)/2} \end{pmatrix} : u \in \mathbb{R},\, \theta \in \mathbb{R}/2\pi\mathbb{Z} \right\}.$$

We will also refer to a matrix of this form as $t_{u,\theta}$. If $u = 0$, then γ is elliptic; if $\theta = 0$, then γ is hyperbolic. All other elements are said to be loxodromic, and this term is sometimes also applied to hyperbolic elements. Note that we can conjugate t_γ by $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ to swap the diagonal elements; therefore, we can and do choose the length/holonomy pair to have non-negative length.

If γ is hyperbolic or loxodromic, it has two fixed points on the boundary $\partial\mathbb{H}^3 \cup \{\infty\} = \hat{\mathbb{C}}$ and acts on $\mathbb{H}^3$ by a shift along the geodesic connecting these two fixed points by the *length* $\ell(\gamma) = u$, followed by a rotation around the same axis by the *holonomy* $\mathrm{hol}(\gamma) = \theta$. We also talk about the *complex length* $\mathbb{C}\ell(\gamma) = u + i\theta$. A hyperbolic or loxodromic $\gamma \in \Gamma$ corresponds to a closed geodesic in $\Gamma\backslash\mathbb{H}^3$. Since Γ is discrete, $\Gamma\backslash\mathbb{H}^3$ contains a geodesic of minimum length, which we refer to throughout the paper as $\eta_0(\Gamma)$.

The group T of diagonal elements of G has a Haar measure $\mathrm{d}u\,\mathrm{d}\theta/2\pi$. For $\nu \in \mathbb{C}$ and $p \in \mathbb{Z}$, define the character $\chi_{\nu,p}$ on T by

$$\chi_{\nu,p}\left(\begin{pmatrix} e^{(u+i\theta)/2} & 0 \\ 0 & e^{-(u+i\theta)/2} \end{pmatrix}\right) = e^{u\nu+ip\theta}. \tag{16}$$

This is a unitary character for $\nu \in i\mathbb{R}$.

The classification of irreducible, unitary representations of $G = \mathrm{PSL}_2\mathbb{C}$ is classical. Let $\pi_{\nu,p}$ denote the representation of G obtained by extending the character $\chi_{\nu,p}$ to $B = UT$ trivially along U and then inducing unitarily to G. The unitary irreducible representations of G are then as follows:

- the trivial representation;
- the unitary principal series representations $\pi_{\nu,p}$, for $\nu \in i\mathbb{R}$ and $p \in \mathbb{Z}$;
- the complementary series representations $\pi_{\nu,0}$, for $0 < \nu < 1$.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

The only equivalences among the above irreducible representations of G are that $\pi_{\nu,p} \cong \pi_{-\nu,-p}$ [15, Theorem 16.2]. For a complete description of the principal and complementary series representations including the G-invariant inner product, see [15, §II.4], [17, §2.4.1].

Let $L^2(\Gamma\backslash G)$ be the space of square-integrable functions on $\Gamma\backslash G$. The group G acts on $L^2(\Gamma\backslash G)$ by the right-regular representation. Then, we have the decomposition

$$L^2(\Gamma\backslash G) = \bigoplus_{\pi\in\hat{G}} m_\Gamma(\pi)\pi, \tag{17}$$

where $\hat{G}$ is the set of irreducible, unitary representations of G and the non-negative integer $m_\Gamma(\pi)$ is the multiplicity of π in $L^2(\Gamma\backslash G)$. Since Γ is co-compact, the non-vanishing terms in (17) form a countable sum that may (after the trivial representation) be double indexed by $\pi_{pj} \simeq \pi_{\nu_{pj},p}$, where, for every $p \in \mathbb{Z}$, $|\nu_{pj}| \to \infty$ ($j \to \infty$).

Decomposition (17) into irreducible representations is fundamentally connected to the theory of automorphic forms. Each representation π appearing in $L^2(\Gamma\backslash G)$ corresponds to an irreducible representation space V_π. The Casimir element of G acts on V_π by scalar multiplication, and V_π is spanned by Γ-automorphic functions. For an explicit description, see [18, Chapter 8].

2.2 Non-spherical Trace Formulas

For a co-compact discrete subgroup $\Gamma < G$, the Selberg trace formula relates spectral information about the multiplicities of representations in $L^2(\Gamma\backslash G)$ to geometric information about elements of Γ. This formula results from computing the trace of the resolvent operator in two ways.

When the underlying kernel is bi-K-invariant, this recovers the classical Selberg trace formula on the compact hyperbolic manifold $M = \Gamma\backslash G/K$. For example, in the present rank one case $G = \mathrm{PSL}_2\mathbb{C}$, the trace formula [8, Theorem 5.1] (after removing the Eisenstein, parabolic, and elliptic terms) relates the spectrum of the Laplacian on $L^2(M)$ with the lengths of closed geodesics on M; see also [12, Theorem 10.2] for the more familiar case $G = \mathrm{SL}_2\mathbb{R}$.

Full control over the holonomy of geodesics on M requires a trace formula on $\Gamma\backslash G$. The following trace formula was explicated by Lin and Lipnowski for compact, hyperbolic 3-manifolds. We refer to Section 2.1 for notations.

Theorem 2.1 (Lin–Lipnowski [17, Corollary 2]). Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every smooth, compactly supported function

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

$F : T \to \mathbb{C}$ such that $F(t) = F(t^{-1})$,

$$\sum_{\nu,p} m_\Gamma(\pi_{\nu,p})\hat{F}(\chi_{\nu,p}^{-1}) + \frac{1}{2}\int_T |D(t^{-1})|^{1/2}F(t)\,\mathrm{d}t$$
$$= -\frac{1}{2\pi}\operatorname{vol}(\Gamma\backslash G)\Big(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial\theta^2}\Big)F\Big|_{t=1} + \sum_{[\gamma]} \ell(\gamma_0)|D(t_\gamma^{-1})|^{-1/2}F(t_\gamma),$$

where $\hat{F}(\chi) = \int_T F(t)\chi^{-1}(t)\,\mathrm{d}t$ is the Abel transform and $D(t_\gamma) = (1 - e^{\mathbb{C}\ell(\gamma)})^2(1 - e^{-\mathbb{C}\ell(\gamma)})^2$ is the Weyl discriminant. The first sum is over unitary principal and complementary series representations $\pi_{\nu,p}$, and $m_\Gamma(\pi_{\nu,p})$ refers to the multiplicity of a representation $\pi_{\nu,p}$ in $L^2(\Gamma\backslash G)$. The latter sum is over the nontrivial conjugacy classes $[\gamma]$ of Γ, and $\ell(\gamma_0)$ refers to the length of the geodesic corresponding to the primitive element γ_0 that generates γ.

Strictly speaking, there are two elements, γ_0 and γ_0^{-1}, which generate γ; however, $\ell(\gamma_0) = \ell(\gamma_0^{-1})$, and so here and henceforth, we ignore this distinction.

Lin and Lipnowski used Theorem 2.1 with $F(t_{u,\theta}) = g(u)\cos(\theta)$, where g is even, smooth, and compactly supported, and evaluated the above equation to isolate the representations $\pi_{\nu,\pm 1}$, which in turn gives a handle on the first eigenvalue of the Hodge Laplacian acting on coexact 1-forms. From a more analytic perspective, irreducible representations of $K = \mathrm{PSU}_2$ are classified as $(2\ell + 1)$-dimensional representations τ_ℓ ($\ell \geqslant 0$), and according to the right K-action we have the decompositions $\pi_{\nu,p}|_K \simeq \bigoplus_{\ell=|p|}^\infty \tau_\ell$ and

$$L^2(\Gamma\backslash G) = \bigoplus_{\ell=0}^\infty L^2(\Gamma\backslash G)_\ell. \tag{18}$$

In the following two theorems, we similarly specialize Theorem 2.1 to isolate multiplicities of representations $\pi_{\nu,\pm p} \subset L^2(\Gamma\backslash G)$ whose lowest K-weight vectors occur in a fixed component $L^2(\Gamma\backslash G)_{|p|}$. Theorems 2.2 (which we adapt from [7]) and 2.4 are the "even" and "odd" trace formulas and should be compared with [25, Theorem 6.5]. Indeed, the intrinsic symmetry in Theorem 2.1 imposes two equalities among the four quantities $m_\Gamma(\pi_{\pm\nu,\pm p})$, so that two trace formulas provide for the fullest possible spectral resolution in (18). Note that $p = 0$ corresponds to the familiar spherical Maass forms on $\Gamma\backslash\mathbb{H}^3$, in which case Theorem 2.2 recovers the classical spherical trace formula [8, Theorem 5.1] for compact hyperbolic 3-manifolds.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

Theorem 2.2. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $n \in \mathbb{Z}$. Then, for every smooth, even, compactly supported function $g : \mathbb{R} \to \mathbb{C}$,

$$\begin{aligned}\frac{1}{2}\sum_{\nu}(m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n}))\int_{-\infty}^{\infty} g(u)e^{u\nu}\,\mathrm{d}u \\ + \delta_0(n)\int_{-\infty}^{\infty} g(u)\,e^{u}\,\mathrm{d}u - \frac{1}{2}\delta_{\pm 1}(n)\hat{g}(0) \\ = \frac{1}{2\pi}\,\mathrm{vol}(\Gamma\backslash G)(n^2 g(0) - g''(0)) \\ + \sum_{[\gamma]}\ell(\gamma_0)|1 - e^{\mathbb{C}\ell(\gamma)}|^{-1}|1 - e^{-\mathbb{C}\ell(\gamma)}|^{-1}g(\ell(\gamma))\cos(n\,\mathrm{hol}(\gamma)).\end{aligned} \tag{19}$$

The first sum is over unitary principal (for $n = 0$, also complementary) series representations $\pi_{\nu,\pm n}$, $m_\Gamma(\pi_{\nu,\pm n})$ refers to the multiplicity of a representation $\pi_{\nu,\pm n}$ in $L^2(\Gamma\backslash G)$, and δ_0, $\delta_{\pm 1}$ are as in Section 1.4. The latter sum is over the nontrivial hyperbolic and loxodromic conjugacy classes $[\gamma]$ of Γ, and $\ell(\gamma_0)$ refers to the length of the geodesic corresponding to the primitive element γ_0 which generates γ.

Remark 2.3. The left-hand side in Theorem 2.2 may be rewritten as

$$\begin{aligned}\delta_0(n)\int_{-\infty}^{\infty} g(u)\,\mathrm{d}\varpi_\Gamma^*(u) \\ + \frac{1}{2}\sum_{\nu\in i\mathbb{R}}(m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n}))\hat{g}\Big(\frac{i\nu}{2\pi}\Big) - \frac{1}{2}\delta_{\pm 1}(n)\hat{g}(0),\end{aligned} \tag{20}$$

where the sum is now only over unitary principal series representations $\pi_{\nu,\pm n}$ with $\nu \in i\mathbb{R}$ (including for $n = 0$), and ϖ_Γ^* is the absolutely continuous measure on $\mathbb{R}$ given by

$$\mathrm{d}\varpi_\Gamma^*(u) = \Big(e^u + \sum_{\nu\in(0,1)} m_\Gamma(\pi_{\nu,0})e^{u\nu}\Big)\,\mathrm{d}u, \tag{21}$$

with the latter sum being over the complementary spectrum $\pi_{\nu,0}$ occurring in $L^2(\Gamma\backslash G)$. The form (20) is particularly well suited to geodesic counting, with ϖ_Γ^* acting as the density of the length spectrum $[\gamma]$ of Γ (not necessarily primitive and weighted by $\ell(\gamma_0)w(\gamma)$ with $w(\gamma) \asymp_\Gamma e^{-\ell(\gamma)}$ as in (19) and (25)).

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Proof. Consider the function $F_n : T \to \mathbb{C}$ defined as $F_n(t_{u,\theta}) = g(u)\cos(n\theta)$. This is a smooth, compactly supported function on T invariant under inverses. Therefore, Theorem 2.1 applies to F_n; we will explicate each term.

On the spectral side,

$$\begin{aligned}\hat{F}_n(\chi_{\nu,p}^{-1}) &= \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_0^{2\pi} g(u)\big(e^{u\nu+i(p+n)\theta} + e^{u\nu+i(p-n)\theta}\big)\,\mathrm{d}\theta\,\mathrm{d}u \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty} g(u)e^{u\nu}\bigg(\int_0^{2\pi}\big(e^{i(p+n)\theta} + e^{i(p-n)\theta}\big)\,\mathrm{d}\theta\bigg)\,\mathrm{d}u.\end{aligned}$$

Therefore, for $n \neq 0$, $\hat{F}_n(\chi_{\nu,\pm n}^{-1}) = \frac{1}{2}\int_{-\infty}^{\infty} g(u)e^{u\nu}\,\mathrm{d}u$, and $\hat{F}_0(\chi_{\nu,0}^{-1}) = \int_{-\infty}^{\infty} g(u)e^{u\nu}\,\mathrm{d}u$. If $p \neq \pm n$, then $\hat{F}_n(\chi_{\nu,p}^{-1}) = 0$. The contribution of the trivial representation is

$$\begin{aligned}\frac{1}{2}\int_T |D(t_{u,\theta}^{-1})|^{1/2}F_n(t_{u,\theta})\,\mathrm{d}t_{u,\theta} &= \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_0^{2\pi} |e^{u+i\theta}|\big|1-e^{-(u+i\theta)}\big|^2 g(u)\cos(n\theta)\,\mathrm{d}\theta\,\mathrm{d}u \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_0^{2\pi} (e^u + e^{-u} - 2\cos\theta)g(u)\cos(n\theta)\,\mathrm{d}\theta\,\mathrm{d}u \qquad (22) \\ &= \delta_0(n)\int_{-\infty}^{\infty} g(u)\cosh u\,\mathrm{d}u - \frac{1}{2}\delta_{\pm 1}(n)\int_{-\infty}^{\infty} g(u)\,\mathrm{d}u,\end{aligned}$$

by orthogonality. Note that $\int_{-\infty}^{\infty} g(u)\cosh u\,\mathrm{d}u = \int_{-\infty}^{\infty} g(u)e^u\,\mathrm{d}u$ since g is even.

On the geometric side, the contribution of the nontrivial hyperbolic and loxodromic elements is

$$\ell(\gamma_0)|D(t_\gamma^{-1})|^{-1/2}F_n(t_\gamma) = \ell(\gamma_0)|1-e^{\mathbb{C}\ell(\gamma)}|^{-1}|1-e^{-\mathbb{C}\ell(\gamma)}|^{-1}g(\ell(\gamma))\cos(n\,\mathrm{hol}(\gamma)).$$

For the contribution of the identity element on the geometric side, we have

$$-\frac{1}{2\pi}\,\mathrm{vol}(\Gamma\backslash G)\Big(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial\theta^2}\Big)g(u)\cos(n\theta)\Big|_{u=0,\theta=0} = \frac{1}{2\pi}\,\mathrm{vol}(\Gamma\backslash G)(n^2g(0) - g''(0)). \qquad \blacksquare$$

Now that we have a formula for the sum of multiplicities $m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n})$, we also require an understanding of the difference in these multiplicities (both subject to the symmetry $\pi_{\nu,n} \cong \pi_{-\nu,-n}$). This can be achieved by capturing both the length and the holonomy with an odd function.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Theorem 2.4. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $n \in \mathbb{Z}$. Then, for every smooth, odd, compactly supported function $h : \mathbb{R} \to \mathbb{C}$,

$$\begin{aligned}\frac{1}{2}i\sum_{\nu}(m_\Gamma(\pi_{\nu,n}) - m_\Gamma(\pi_{\nu,-n}))\int_{-\infty}^{\infty} h(u)e^{u\nu}\,\mathrm{d}u \\ = \sum_{[\gamma]}\ell(\gamma_0)|1-e^{\mathbb{C}\ell(\gamma)}|^{-1}|1-e^{-\mathbb{C}\ell(\gamma)}|^{-1}h(\ell(\gamma))\sin(n\,\mathrm{hol}(\gamma)),\end{aligned}$$

where the terms are defined as in Theorem 2.2.

Proof. This time we consider the function $H_n : T \to \mathbb{C}$ given by $H_n(t_{u,\theta}) = h(u)\sin(n\theta)$. This is a smooth, compactly supported function on T invariant under inverses. Therefore, Theorem 2.1 applies to H_n, and we explicate all terms. As for the spectral terms, we compute

$$\begin{aligned}\hat{H}_n(\chi_{\nu,p}^{-1}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^{2\pi} H_n(t_{u,\theta})e^{u\nu+ip\theta}\,\mathrm{d}\theta\,\mathrm{d}u \\ &= \frac{1}{2\pi}\hat{h}\Big(\frac{i\nu}{2\pi}\Big)\int_0^{2\pi}\frac{1}{2i}\big(e^{i(p+n)\theta}-e^{i(p-n)\theta}\big)\,\mathrm{d}\theta.\end{aligned}$$

By orthogonality, this vanishes unless $p = \pm n \neq 0$, in which case

$$\hat{H}_n(\chi_{\nu,n}^{-1}) = \frac{1}{2}i\hat{h}\Big(\frac{i\nu}{2\pi}\Big), \quad \hat{H}_n(\chi_{\nu,-n}^{-1}) = -\frac{1}{2}i\hat{h}\Big(\frac{i\nu}{2\pi}\Big).$$

Thus, the spectral terms contribute $\frac{1}{2}i\sum_\nu(m_\Gamma(\pi_{\nu,n}) - m_\Gamma(\pi_{\nu,-n}))\hat{h}(i\nu/2\pi)$, as advertised.

The terms corresponding to the trivial representation and the identity element vanish. Indeed, using the evaluation (22) from the proof of Theorem 2.2 (which in particular shows that $|D(t_{u,\theta}^{-1})|^{1/2}$ is an even function of θ),

$$\frac{1}{2}\int_T |D(t_{u,\theta}^{-1})|^{1/2}H_n(t_{u,\theta})\,\mathrm{d}t_{u,\theta} = \frac{1}{4\pi}\int_{-\infty}^{\infty}\int_0^{2\pi}(e^u + e^{-u} - 2\cos\theta)h(u)\sin(n\theta)\,\mathrm{d}\theta\,\mathrm{d}u = 0,$$

while

$$-\frac{1}{2\pi}\,\mathrm{vol}(\Gamma\backslash G)\Big(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial\theta^2}\Big)h(u)\sin(n\theta)\Big|_{u=0,\theta=0} = 0. \qquad \blacksquare$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

2.3 Plancherel Measure and Bounds on Spectral Densities

Selberg's trace formula gives a handle on the distribution of the spectrum as the spectral parameters (such as the Laplace eigenvalue) increase, which on quotients of a group such as $G = \mathrm{PSL}_2\mathbb{C}$ is guided by a fixed Plancherel measure, depending on G only and supported on the tempered spectrum of $L^2(G)$. In the present situation, keeping in mind the classification from Section 2.1, let μ_{pl} be the absolutely continuous measure on $i\mathbb{R} \times \mathbb{Z}$ given by

$$\mathrm{d}\mu_{\mathrm{pl}}(\nu, n) = \frac{1}{4\pi^2}(|\nu|^2 + n^2)\,|\mathrm{d}\nu|. \tag{23}$$

Then, the identity term in Theorem 2.2 may be rewritten (using Fourier inversion) in the form

$$\mathrm{vol}(\Gamma\backslash G) \cdot \frac{1}{2}\int_{i\mathbb{R}} \left(\int_{-\infty}^{\infty} g(u)e^{u\nu}\,\mathrm{d}u\right)(\mathrm{d}\mu_{\mathrm{pl}}(\nu, n) + \mathrm{d}\mu_{\mathrm{pl}}(\nu, -n)),$$

which should be compared to the cuspidal term (the first sum on the left-hand side of (19)) and may be understood as the leading (or global) term in its geometric expansion, accounting for the spectral density of G, and similarly in Theorem 2.1.

All we need for our application to Theorem 1.1 is a uniform estimate on the cardinality of the spectrum in a short window (a familiar step in the derivation of Weyl's law). In this section, we use the even trace formula from Theorem 2.2 to prove such a local bound for the density of a particular representation type. We emphasize that, while the implied constants in Proposition 2.5 depend on the discrete subgroup Γ, they are independent of n.

Proposition 2.5. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every $n \in \mathbb{Z}$ and $R \in \mathbb{R}$, the multiplicities $m_\Gamma(\pi_{\nu,n})$ of representations $\pi_{\nu,n}$ in $L^2(\Gamma\backslash G)$ satisfy

$$\sum_{R-1\leqslant|\nu|\leqslant R+1} m_\Gamma(\pi_{\nu,n}) \ll \mathrm{vol}(\Gamma\backslash G) \cdot \int_{R-1\leqslant|\nu|\leqslant R+1} \mathrm{d}\mu_{\mathrm{pl}}(\nu, n) + \mathrm{O}_\Gamma(1)$$

$$\asymp \mathrm{vol}(\Gamma\backslash G) \cdot (R^2 + n^2) + \mathrm{O}_\Gamma(1).$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

Proof. Let g be a smooth, even, non-negative, compactly supported function such that $\hat{g}(t) \geqslant 0$ for all $t \in \mathbb{R} \cup i\mathbb{R}$, as well as $\hat{g}(t) \geqslant 1$ for $|t| \leqslant 1/(2\pi)$. Consider the function

$$g_R(x) = 2\cos(Rx)g(x).$$

Its Fourier transform $\hat{g}_R(t) = \hat{g}(t - R/2\pi) + \hat{g}(t + R/2\pi)$ has the property that $\hat{g}_R(t) \geqslant 1$ for $R - 1 \leqslant 2\pi|t| \leqslant R + 1$. Therefore, we have the bound:

$$\sum_{R-1\leqslant|\nu|\leqslant R+1} m_\Gamma(\pi_{\nu,n}) \leqslant \sum_{\nu} m_\Gamma(\pi_{\nu,n})\hat{g}_R(i\nu/2\pi) + \mathrm{O}_\Gamma(1).$$

By Theorem 2.2, the right-hand side of this estimate equals

$$\begin{aligned} &-\delta_0(n)\int_{-\infty}^{\infty} g_R(u)e^u\,\mathrm{d}u + \frac{1}{2}\delta_{\pm 1}(n)\hat{g}_R(0) + \frac{1}{2\pi}\mathrm{vol}(\Gamma\backslash G)(n^2 g_R(0) - g_R''(0)) \\ &\quad + \sum_{[\gamma]} \ell(\gamma_0)|1 - e^{\mathbb{C}\ell(\gamma)}|^{-1}|1 - e^{-\mathbb{C}\ell(\gamma)}|^{-1} g_R(\ell(\gamma))\cos(n\,\mathrm{hol}\,\gamma) + \mathrm{O}_\Gamma(1). \end{aligned}$$

The first two terms contribute $\mathrm{O}(1)$. For the fourth term, g_R is a compactly supported function with $\mathrm{supp}\, g_R \subseteq \mathrm{supp}\, g$ and $|g_R| \leqslant |g|$, so this sum contains $\mathrm{O}_\Gamma(1)$ terms and contributes $\mathrm{O}_\Gamma(1)$. Further, we calculate that

$$g_R''(x) = 2\cos(Rx)g''(x) - 4R\sin(Rx)g'(x) - 2R^2\cos(Rx)g(x),$$

so that the third term contributes $\frac{1}{\pi}\mathrm{vol}(\Gamma\backslash G)\cdot((n^2 + R^2)g(0) - g''(0)) \ll \mathrm{vol}(\Gamma\backslash G)\cdot(R^2 + n^2 + 1)$. Thus,

$$\sum_{R-1\leqslant|\nu|\leqslant R+1} m_\Gamma(\pi_{\nu,n}) \ll \mathrm{vol}(\Gamma\backslash G)\cdot(R^2 + n^2 + 1) + \mathrm{O}_\Gamma(1),$$

which completes the proof since the remaining claims are immediate. ∎

An immediate corollary of Proposition 2.5 is the following:

Corollary 2.6. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every $n \in \mathbb{Z}$ and $R \geqslant 1$, the multiplicities $m_\Gamma(\pi_{\nu,n})$ of representations $\pi_{\nu,n}$ in $L^2(\Gamma\backslash G)$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

satisfy

$$\sum_{|\nu|\leqslant R} m_\Gamma(\pi_{\nu,n}) \ll \mathrm{vol}(\Gamma\backslash G)\cdot(R^3+n^2R)+\mathrm{O}_\Gamma(R).$$

3 Sampling the Length Spectrum Using the Trace Formula

In this section, we prove estimates on certain sums that naturally appear on the geometric side of the trace formulas in Theorems 2.2 and 2.4 when one is interested in sampling geodesics C_γ on M, controlling their length $\ell(\gamma)$ close to or up to a specific length y and detecting their holonomy with a character $e^{in\,\mathrm{hol}(\gamma)}$.

For suitable even and odd sampling functions $g,h:\mathbb{R}\to\mathbb{R}$, respectively, Theorems 2.2 and 2.4 give a handle on sums

$$\begin{aligned} T_n^{\cos}[g] &= \sum_{[\gamma]} \ell(\gamma_0) g(\ell(\gamma)) w(\gamma)\cos(n\,\mathrm{hol}(\gamma)),\\ T_n^{\sin}[h] &= \sum_{[\gamma]} \ell(\gamma_0) h(\ell(\gamma)) w(\gamma)\sin(n\,\mathrm{hol}(\gamma)), \end{aligned} \tag{24}$$

where the sums are over the nontrivial hyperbolic and loxodromic conjugacy classes $[\gamma]$ of Γ, $\ell(\gamma_0)$ refers to the length of the geodesic corresponding to the primitive element γ_0, which generates γ, and

$$w(\gamma) = |1-e^{\mathbb{C}\ell(\gamma)}|^{-1}|1-e^{-\mathbb{C}\ell(\gamma)}|^{-1}. \tag{25}$$

Let $\psi:\mathbb{R}\to\mathbb{R}_{\geqslant 0}$ be a fixed, smooth non-negative function that is compactly supported on $[-1,1]$ and satisfies $\|\psi\|_1=1$ and $\psi\geqslant\frac{1}{2}$ on $[-\frac{1}{2},\frac{1}{2}]$, and for $\eta>0$, define

$$\psi_\eta(t) := \frac{1}{\eta}\psi\Big(\frac{t}{\eta}\Big). \tag{26}$$

For $y>0$, define the sampling functions $g_{y,\eta},h_{y,\eta}:\mathbb{R}\to\mathbb{R}$ as convolutions

$$g_{y,\eta} = \psi_\eta \star \chi_{[-y,y]}, \quad h_{y,\eta} = \psi_\eta \star (\chi_{[-y,y]}\cdot \mathrm{sgn}). \tag{27}$$

In Section 3.1, we use the non-spherical trace formulas of Theorems 2.2 and 2.4 to prove estimates on $T_n^{\cos}[g_{y,\eta}]$ and $T_n^{\sin}[h_{y,\eta}]$. In Section 3.2, we execute the passage from these

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

smooth counts to the sharp counts

$$\begin{aligned} T_n^{\cos}(y) &:= T_n^{\cos}[\chi_{[-y,y]}] \quad\quad = \sum_{\ell(\gamma)\leqslant y} \ell(\gamma_0) w(\gamma) \cos(n \operatorname{hol}(\gamma)), \\ T_n^{\sin}(y) &:= T_n^{\sin}[\chi_{[-y,y]} \cdot \operatorname{sgn}] = \sum_{\ell(\gamma)\leqslant y} \ell(\gamma_0) w(\gamma) \sin(n \operatorname{hol}(\gamma)). \end{aligned} \tag{28}$$

3.1 Trace Formula Estimates

In this section, we use Theorems 2.2 and 2.4 to prove in Lemma 3.1 estimates on the smooth counts $T_n^{\cos}[g_{Y,\eta}]$ and $T_n^{\sin}[h_{Y,\eta}]$ defined in (24).

Lemma 3.1. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, let $n \in \mathbb{Z}$, and let $Y, \eta > 0$. Then, the sums $T_n^{\cos}[g_{Y,\eta}]$ and $T_n^{\sin}[h_{Y,\eta}]$ defined in (24), with $g_{Y,\eta}$ and $h_{Y,\eta}$ as in (27), satisfy

$$\begin{aligned} T_n^{\cos}[g_{Y,\eta}] &= \delta_0(n) \int_{-\infty}^{\infty} g_{Y,\eta}(u)\, \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\left(\frac{1}{\eta^2} + (1+n^2)\Big(\log^* \frac{1}{\eta} + Y\Big)\right), \\ T_n^{\sin}[h_{Y,\eta}] &= \mathrm{O}_\Gamma\left(\frac{1}{\eta^2} + (1+n^2)\Big(\log^* \frac{1}{\eta} + Y\Big)\right), \end{aligned}$$

where ϖ_Γ^* is as in (21) and $\log^* x = \log(2+x)$.

Proof. Using Theorem 2.2 for $g_{Y,\eta}$, in the form (20), we know that

$$\begin{aligned} T_n^{\cos}[g_{Y,\eta}] - \delta_0(n) \int_{-\infty}^{\infty} g_{Y,\eta}(u)\, \mathrm{d}\varpi_\Gamma^*(u) &= \frac{1}{2} \sum_{\nu \in i\mathbb{R}} (m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n})) \hat{g}_{Y,\eta}\Big(\frac{i\nu}{2\pi}\Big) \\ &\quad - \frac{1}{2}\delta_{\pm 1}(n) \hat{g}_{Y,\eta}(0) - \frac{1}{2\pi} \operatorname{vol}(\Gamma\backslash G)(n^2 g_{Y,\eta}(0) - g_{Y,\eta}''(0)). \end{aligned} \tag{29}$$

With our choice of $g_{Y,\eta}$, the right-hand side of (29) equals

$$\frac{1}{2} \sum_{\nu \in i\mathbb{R}} (m_\Gamma(\pi_{\nu,n}) + m_\Gamma(\pi_{\nu,-n})) \hat{g}_{Y,\eta}\Big(\frac{i\nu}{2\pi}\Big) - \delta_{\pm 1}(n) Y + \mathrm{O}\Big(\operatorname{vol}(\Gamma\backslash G)\Big(n^2 + \frac{1}{\eta^2}\Big)\Big).$$

(When $\eta < Y$, the error term $\mathrm{O}(\operatorname{vol}(\Gamma\backslash G)/\eta^2)$ is not needed here, since then $g_{Y,\eta}''(0) = 0$; however, we incur this term later regardless.)

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

For the spectral contribution, we first compute that the Fourier transform for $\nu = it$ ($t \in \mathbb{R}$) is

$$\hat{g}_{Y,\eta}\Big(\frac{-t}{2\pi}\Big) = 2\hat{\psi}\Big(\frac{-\eta t}{2\pi}\Big)\frac{\sin(tY)}{t}.$$

Using the Schwartz bound $|\hat{\psi}(-\eta t/2\pi)| \ll_m 1/(1+\eta|t|)^m$ with (say) $m = 3$, we obtain

$$\begin{aligned}\frac{1}{2}\sum_{\nu=it} m_\Gamma(\pi_{\nu,\pm n})\hat{g}_{Y,\eta}\Big(\frac{i\nu}{2\pi}\Big) &= \Big(\sum_{0\leqslant|t|<1} + \sum_{1\leqslant|t|\leqslant 1/\eta} + \sum_{|t|>1/\eta}\Big) m_\Gamma(\pi_{it,\pm n})\hat{\psi}\Big(\frac{-\eta t}{2\pi}\Big)\frac{\sin(tY)}{t}\\ &\ll Y\sum_{0\leqslant|t|<1} m_\Gamma(\pi_{it,\pm n}) + \sum_{1\leqslant k\leqslant 1/\eta}\frac{1}{k}\sum_{k\leqslant|t|<k+1} m_\Gamma(\pi_{it,\pm n})\\ &\quad + \sum_{k>1/\eta}\frac{1}{k}\frac{1}{(\eta k)^3}\sum_{k\leqslant|t|<k+1} m_\Gamma(\pi_{it,\pm n}).\end{aligned}$$

Bounding the multiplicities $m_\Gamma(\pi_{\nu,\pm n})$ using the uniform local bound of Proposition 2.5, we find that the above is

$$\begin{aligned}&\ll \mathrm{vol}(\Gamma\backslash G)\Big[Y(1+n^2) + \sum_{k\leqslant\frac{1}{\eta}}\Big(k+\frac{n^2}{k}\Big) + \sum_{k>\frac{1}{\eta}}\Big(\frac{1}{\eta^3k^2}+\frac{n^2}{\eta^3k^4}\Big)\Big]\\ &\quad + \mathrm{O}_\Gamma\Big(Y + \sum_{1\leqslant k\leqslant 1/\eta}\frac{1}{k} + \sum_{k>1/\eta}\frac{1}{\eta^3k^4}\Big)\\ &\ll \mathrm{vol}(\Gamma\backslash G)\Big[Y(1+n^2) + \frac{1}{\eta^2} + n^2\log^*\frac{1}{\eta}\Big] + \mathrm{O}_\Gamma\Big(Y + \log^*\frac{1}{\eta}\Big).\end{aligned}$$

Putting everything together completes the proof for the even case.

The odd case is completely analogous and in fact easier, since the trace formula from Theorem 2.4 has only the principal series spectral and non-identity geometric terms. Indeed, we compute that

$$\hat{h}_{Y,\eta}\Big(\frac{-t}{2\pi}\Big) = 4\hat{\psi}\Big(\frac{-\eta t}{2\pi}\Big)\frac{\sin^2(tY/2)}{it},$$

so that using the Schwartz bound $|\hat{\psi}(-\eta t/2\pi)| \ll_m 1/(1+\eta|t|)^m$ and $|\sin^2(tY/2)/it| \ll \min(tY^2, 1/t) \ll \min(Y, 1/t)$, the estimates proceed as above. ∎

Remark 3.2. Theorems 2.2 and 2.4 allow for good control over the dependence in Γ, as we show in Proposition 2.5 and then using this result in the proof of Lemma 3.1.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

For example, the dependence in the leading terms is often guided by $\mathrm{vol}(\Gamma\backslash G)$ only. This is a very important feature when the group Γ varies or where uniformity in Γ is required (say, for a varying level in a congruence group). Since Γ is fixed for us, from now on we combine all dependence on Γ, as in the statement of Lemma 3.1 and beyond.

3.2 Passage to Sharp Cutoff in Length

In this section, we pass from the smooth counts for $T_n^{\cos}[g_{Y,\eta}]$ and $T_n^{\sin}[h_{Y,\eta}]$ of Lemma 3.1 to a sharp count for $T_n^{\cos}(Y) = T_n^{\cos}[\chi_{[-Y,Y]}]$ and $T_n^{\sin}(Y) = T_n^{\sin}[\chi_{[-Y,Y]} \cdot \mathrm{sgn}]$ as shown in (28). This passage requires further use of Theorems 2.2 and 2.4 as well as taking $\eta > 0$ small in Lemma 3.1, which typically ends up being the main source of the error terms.

The passage to the sharp geodesic count relies primarily on estimating contributions from classes in the transition zone $Y - \eta \leqslant \ell(\gamma) \leqslant Y + \eta$, which contains the range where $g_{Y,\eta}(\ell(\gamma)) \neq \chi_{[-Y,Y]}$. This is achieved in the following key lemma, which is the geometric side analogue, in the length aspect, of Proposition 2.5.

Lemma 3.3. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $Y, \eta > 0$. Then,

$$\sum_{Y-\eta\leqslant\ell(\gamma)\leqslant Y+\eta} \ell(\gamma_0) w(\gamma) \ll \int_{Y-2\eta}^{Y+2\eta} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\Big(\frac{1}{\eta^2} + \eta\Big), \tag{30}$$

where $w(\gamma)$ is as in (25) and ϖ_Γ^* is as in (21).

Proof. We will sample the geodesics in the range $\ell(\gamma) \in [Y - \eta, Y + \eta]$ using an even, majorant function

$$f_{Y,2\eta}(x) = \psi\Big(\frac{x - Y}{2\eta}\Big) + \psi\Big(\frac{x + Y}{2\eta}\Big), \tag{31}$$

with a smooth, non-negative bump function ψ as in (26). By the definition of ψ, $f_{Y,2\eta} : \mathbb{R} \to \mathbb{R}$ is a smooth, even, non-negative, absolutely bounded function supported on $\pm[Y - 2\eta, Y + 2\eta]$ and satisfying $f_{Y,2\eta} \gg 1$ on $\pm[Y - \eta, Y + \eta]$, so that

$$\sum_{Y-\eta\leqslant\ell(\gamma)\leqslant Y+\eta} \ell(\gamma_0) w(\gamma) \ll \sum_{[\gamma]} \ell(\gamma_0) w(\gamma) f_{Y,2\eta}(\ell(\gamma)).$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Using the even trace formula of Theorem 2.2, with $n = 0$ and in the form (20), we obtain

$$\sum_{[\gamma]} \ell(\gamma_0) w(\gamma) f_{Y,2\eta}(\ell(\gamma)) = \int_{-\infty}^{\infty} f_{Y,2\eta}(u)\, \mathrm{d}\varpi_\Gamma^*(u) + \sum_{\nu \in i\mathbb{R}} m_\Gamma(\pi_{\nu,0}) \hat{f}_{Y,2\eta}\Big(\frac{i\nu}{2\pi}\Big) + \frac{1}{2\pi} \operatorname{vol}(\Gamma\backslash G) f''_{Y,2\eta}(0).$$

Note that, at this point, we are only using the familiar spherical trace formula. We estimate the first term using $f_{Y,2\eta} \ll 1$ and the support condition on $f_{Y,2\eta}$. The third term may be absorbed in $\mathrm{O}_\Gamma(1/\eta^2)$. (In fact, in a typical application with $\eta < Y/2$, this term vanishes.)

For the principal series representations $\pi_{it,0}$ ($t \in \mathbb{R}$), the Fourier transform $\hat{f}_{Y,2\eta}(t) = 4\eta \cos(2\pi t Y) \hat{\psi}(2\eta t)$ satisfies the Schwartz bound $|\hat{f}_{Y,2\eta}(-t/2\pi)| \ll \eta/(1+\eta|t|)^4$. Bounding the multiplicities $m_\Gamma(\pi_{it,0})$ using the uniform bound of Proposition 2.5, we may finally bound the contribution of the principal series representations as

$$\begin{aligned}\sum_{\nu=it} \hat{f}_{Y,2\eta}(-t/2\pi) m_\Gamma(\pi_{it,0}) &= \Big(\sum_{0\leqslant k<1/\eta} + \sum_{k>1/\eta}\Big) \sum_{k\leqslant|t|<k+1} \hat{f}_{Y,2\eta}(-t/2\pi) m_\Gamma(\pi_{it,0}) \\ &\ll_\Gamma \eta \sum_{0\leqslant k<1/\eta} (k^2+1) + \frac{1}{\eta^3} \sum_{k>1/\eta} \frac{1}{k^2} \ll \frac{1}{\eta^2} + \eta.\end{aligned}$$

Combining everything completes the proof. ∎

Remark 3.4. In the typical regime for the application of Lemma 3.3, when $\eta \ll 1 \ll Y - \eta$, we have in (30) simply $w(\gamma) \asymp_\Gamma e^{-\ell(\gamma)}$ and $|\mathrm{d}\varpi_\Gamma^*(u)/\mathrm{d}u| \asymp_\Gamma e^u$, so that Lemma 3.3 states that

$$\sum_{Y-\eta\leqslant\ell(\gamma)\leqslant Y+\eta} \ell(\gamma_0) \ll_\Gamma e^Y\Big(\eta e^Y + \frac{1}{\eta^2}\Big).$$

This will be the case, in particular, in the proof of Proposition 3.5.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Proposition 3.5. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $n \in \mathbb{Z}$. Then, for every $Y > 0$, the sums $T_n^{\cos}(Y)$ and $T_n^{\sin}(Y)$ defined in (28) satisfy

$$T_n^{\cos}(Y) = \delta_0(n) \int_{-Y}^{Y} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\big(e^{2Y/3} + n^2 Y\big),$$
$$T_n^{\sin}(Y) = \mathrm{O}_\Gamma\big(e^{2Y/3} + n^2 Y\big).$$

Proof. We will use Lemmata 3.1 and 3.3, with a parameter $\eta > 0$ to be suitably chosen momentarily. According to the definition (27), we have that the functions $g_{Y,\eta}$ and $\chi_{[-Y,Y]}$ agree outside the set $\pm[Y-\eta, Y+\eta]$, on which $|g_{Y,\eta} - \chi_{[-Y,Y]}| = \mathrm{O}(1)$. Therefore,

$$\begin{aligned}\big|T_n^{\cos}(Y) - T_n^{\cos}[g_{Y,\eta}]\big| &= \Big|\sum_{[\gamma]} \ell(\gamma_0)\big(g_{Y,\eta}(\ell(\gamma)) - \chi_{[-Y,Y]}(\ell(\gamma))\big) w(\gamma)\cos(n\,\mathrm{hol}(\gamma))\Big| \\ &\ll \sum_{Y-\eta \leqslant \ell(\gamma) \leqslant Y+\eta} \ell(\gamma_0) w(\gamma).\end{aligned}$$

Therefore, using Lemmata 3.1 and 3.3, we find that

$$\begin{aligned}&\Big|T_n^{\cos}(Y) - \delta_0(n)\int_{-Y}^{Y} \mathrm{d}\varpi_\Gamma^*(u)\Big| \\ &\quad\leqslant \big|T_n^{\cos}(Y) - T_n^{\cos}[g_{Y,\eta}]\big| + \Big|T_n^{\cos}[g_{Y,\eta}] - \delta_0(n)\int_{-\infty}^{\infty} g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma^*(u)\Big| \\ &\qquad + \delta_0(n)\int_{-\infty}^{\infty}\big(g_{Y,\eta}(u) - \chi_{[-Y,Y]}(u)\big)\,\mathrm{d}\varpi_\Gamma^*(u) \\ &\quad\ll \int_{Y-2\eta}^{Y+2\eta} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\Big(\frac{1}{\eta^2} + (1+n^2)\Big(\log^*\frac{1}{\eta} + Y\Big) + \eta\Big).\end{aligned}$$

The statement of Proposition 3.5 is vacuously true for $Y = \mathrm{O}_\Gamma(1)$, so we may assume that $Y \gg_\Gamma 1$. As already mentioned in Remark 3.4 and is clear from the definition (21), $\mathrm{d}\varpi_\Gamma^*(u)/\mathrm{d}u \asymp_\Gamma e^u$ for $u \geqslant 0$ (or $u = \mathrm{O}_\Gamma(1)$). We will choose $\eta \ll_\Gamma 1$; then,

$$\int_{Y-2\eta}^{Y+2\eta} \mathrm{d}\varpi_\Gamma^*(u) \asymp_\Gamma \eta e^Y.$$

The admissible choice $\eta = e^{-Y/3}$ optimizes the error terms and yields Proposition 3.5 for $T_n^{\cos}$.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Similarly, if we additionally require 2η to be less than the minimal geodesic length $\eta_0(\Gamma)$, we can show that

$$\left|T_n^{\sin}(Y) - T_n^{\sin}[h_{Y,\eta}]\right| \ll \sum_{Y-\eta \leqslant \ell(\gamma) \leqslant Y+\eta} \ell(\gamma_0) w(\gamma),$$

and the proof is identical from here on. ■

3.3 Passage to Sharp Cutoff in Holonomy

In this section, we prepare the ground for passage to the sharp count in holonomy. The path to the sharp count is again the geometric counterpart to Proposition 2.5, but this time in the holonomy aspect.

Lemma 3.6. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $Y > 0$, $\theta_0 \in \mathbb{R}$, and $0 < \eta' \leqslant 2\pi$. Then,

$$\sum_{\substack{\ell(\gamma) \leqslant Y \\ \theta_0 - \eta' \leqslant \mathrm{hol}(\gamma) \leqslant \theta_0 + \eta'}} \ell(\gamma_0) w(\gamma) \ll \eta' \int_{-Y}^{Y} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\left(\frac{Y}{\eta'^2}\right), \tag{32}$$

where $w(\gamma)$ is as in (25) and ϖ_Γ^* is as in (21).

Proof. Since Γ is discrete, the claim is vacuously true for $Y < \eta_0(\Gamma)$, so we may assume that $Y \geqslant \eta_0(\Gamma)$. Consider a majorant function $f_{\theta_0,2\eta'} : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}$ given by

$$f_{\theta_0,2\eta'}(t) = \sum_{n \in \mathbb{Z}} \left[\psi\left(\frac{t + 2n\pi - \theta_0}{2\eta'}\right) + \psi\left(\frac{t + 2n\pi + \theta_0}{2\eta'}\right)\right],$$

where ψ is a smooth, non-negative bump function as in (26). This is simply a $2\pi\mathbb{Z}$-periodization of the majorant (31) used in the proof of Lemma 3.3; it is a smooth, even, non-negative absolutely bounded function supported on $\pm[\theta_0 - 2\eta', \theta_0 + 2\eta'] + 2\pi\mathbb{Z}$ and satisfying $f_{\theta_0,2\eta'} \gg 1$ on $\pm[\theta_0 - \eta', \theta_0 + \eta'] + 2\pi\mathbb{Z}$. Since the non-negative sampling function $g_{Y+1,1} = \psi_1 \star \chi_{[-Y-1,Y+1]}$ given in (27) also satisfies $g_{Y+1,1} \gg 1$ for $[-Y, Y]$, we have

$$\sum_{\substack{\ell(\gamma) \leqslant Y \\ \theta_0 - \eta' \leqslant \mathrm{hol}(\gamma) \leqslant \theta_0 + \eta'}} \ell(\gamma_0) w(\gamma) \ll \sum_{[\gamma]} \ell(\gamma_0) w(\gamma) g_{Y+1,1}(\ell(\gamma)) f_{\theta_0,2\eta'}(\mathrm{hol}(\gamma)).$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Using the cosine Fourier expansion for $f_{\theta_0,2\eta'}$ shows that, after a standard unfolding calculation, the right-hand side equals

$$\sum_{[\gamma]} \ell(\gamma_0) w(\gamma) g_{Y+1,1}(\ell(\gamma)) \cdot \frac{4\eta'}{2\pi} \sum_{m\in\mathbb{Z}} \cos(m\theta_0)\hat{\psi}\big(\frac{\eta' m}{\pi}\big)\cos(m\,\mathrm{hol}(\gamma))$$
$$= \frac{2\eta'}{\pi} \sum_{m\in\mathbb{Z}} \cos(m\theta_0)\hat{\psi}\big(\frac{\eta' m}{\pi}\big) T_m^{\cos}[g_{Y+1,1}],$$

using the definition (24). Applying Lemma 3.1 and the Schwartz estimate $|\hat{\psi}(\eta' t/\pi)| \ll_k 1/(1+\eta'|t|)^k$ with (say) $k = 4$, this sum evaluates as

$$\frac{2\eta'}{\pi}\hat{\psi}(0)\int_{-\infty}^{\infty} g_{Y+1,1}(u)\,\mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\Big(\eta' \sum_{m\in\mathbb{Z}} \Big|\hat{\psi}\Big(\frac{\eta' m}{\pi}\Big)\Big|(1+m^2)Y\Big)$$
$$\ll \eta' \int_{-Y-2}^{Y+2} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\Big(\eta' Y\Big[1 + \sum_{1\leqslant m\leqslant 1/\eta'} m^2 + \sum_{m>1/\eta'} \frac{m^2}{(\eta' m)^4}\Big]\Big)$$
$$\ll \eta' \int_{-Y}^{Y} \mathrm{d}\varpi_\Gamma^*(u) + \mathrm{O}_\Gamma\Big(\frac{Y}{\eta'^2}\Big),$$

which completes the proof. ∎

Remark 3.7. The reader will notice that this device for passage to sharp count in holonomy requires the use of Lemma 3.1 (and thus Theorem 2.2) with large n. This mirrors the fact that the passage to sharp count in length requires the use of Lemma 3.1 with small $\eta > 0$, which in turn relies on using Theorem 2.2 with large spectral parameter ν. In both cases, the proof boils down to estimates on the density of the automorphic spectrum $\pi_{\nu,n}$ with spectral parameters increasing in different directions; these are in turn provided by Proposition 2.5 which works over any ball of spectral parameters of radius O(1). This structural parallel underlies the agreement between our results in the length and holonomy aspects.

4 Primitivity and Weights

For $Y > 0$ and $n \in \mathbb{Z}$, consider the sums

$$S_n(Y) = \sum_{\ell(\gamma)\leqslant Y} \ell(\gamma) e^{-\ell(\gamma)+in\,\mathrm{hol}\,\gamma}, \qquad S_n^P(Y) = \sum_{\ell(\gamma)\leqslant Y}^{P} \ell(\gamma) e^{-\ell(\gamma)+in\,\mathrm{hol}\,\gamma}, \tag{33}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

where the summation is over the nontrivial conjugacy classes $[\gamma]$ of Γ, and (here and throughout) the superscript ^{P} indicates that summation is restricted to primitive nontrivial conjugacy classes. The sum $S_n^P(y)$ is of primary interest for counting primitive geodesics with control on holonomy.

On the other hand, an application of trace formula as in Section 3 (see (28) and Proposition 3.5) naturally gives a handle on sums such as $T_n(y)$ and its cousin $T_n^P(y)$ defined by

$$T_n(y) = \sum_{\ell(\gamma)\leqslant y} \ell(\gamma_0) w(\gamma) e^{in\,\mathrm{hol}(\gamma)}, \qquad T_n^P(y) = \sum_{\ell(\gamma)\leqslant y}^{P} \ell(\gamma) w(\gamma) e^{in\,\mathrm{hol}(\gamma)}, \tag{34}$$

where γ_0 is the primitive hyperbolic or loxodromic element that generates γ and the weight $w(\gamma) \asymp_\Gamma e^{-\ell(\gamma)}$ is as in (25). The main result of this section, Lemma 4.1 shows that all four sums defined in (33) and (34) agree up to a very small error term. In Lemma 4.3, we record a similar result for sums with a more general length cutoff.

Lemma 4.1. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $y > 0$, $n \in \mathbb{Z}$. Then, the sums defined in (33) and (34) satisfy:

$$S_n^P(y) = S_n(y) + \mathrm{O}_\Gamma(y), \tag{35}$$

$$S_n^P(y) = T_n^P(y) + \mathrm{O}_\Gamma(y), \tag{36}$$

$$T_n^P(y) = T_n(y) + \mathrm{O}_\Gamma(y), \tag{37}$$

and, consequently,

$$S_n^P(y) = T_n(y) + \mathrm{O}_\Gamma(y). \tag{38}$$

Remark 4.2. The bound (38), which follows from (36) and (37), can be thought of as a statement about removal of unwieldy weights $w(\gamma)$ and imprimitive classes from $T_n(y)$. Since our bounds on $T_n(y)$ in Proposition 3.5 are exponential in y (in particular, $T_0(y) \asymp_\Gamma e^y$), Lemma 4.1 shows that the error terms introduced by these maneuvers are very small in comparison, and also that (cf. (35)) all our statements hold if extended to include imprimitive geodesics. The sources of leading error terms in all our principal results are elsewhere, notably in the passage from smooth to sharp cutoff.

Lemma 4.1 is similar, both in spirit and quantitative strength, to the relations between Chebyshev's functions in the proof of the Prime Number Theorem. The estimate $S_n^P(y) = S_n(y) + \mathrm{O}_\Gamma(y)$ should be compared to the classical estimate $|\psi(x) - \theta(x)| \ll x^{1/2}$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

for $\psi(x), \theta(x) \sim x$ as $x \to \infty$, bearing in mind the weights $\ell(\gamma)e^{-\ell(\gamma)}$ in (33). If those weights were removed by summation by parts as in the proof of Proposition 5.1, the corresponding unweighted sums (say $\tilde{S}_n(y)$ and $\tilde{S}_n^P(y) = K_n(y)$ in (41)) would satisfy $|\tilde{S}_n(y) - \tilde{S}_n^P(y)| \ll_{\Gamma,n} e^y$ as compared to the main term $\tilde{S}_0^P(y) \sim_\Gamma e^{2y}/2y$.

Essential ideas for Lemma 4.1 are due to Sarnak–Wakayama [25, Lemmata 7.1, 7.2]. As for us, a key element of their proof of the equidistribution result (3) is to approximate $S_n^P(y)$, which appears in the spectral (Fourier) decomposition of the sum over primitive conjugacy classes, by $T_n(y)$, which can be approximated by the hyperbolic/loxodromic term in the non-spherical trace formula with an appropriate choice of test function. For a general rank one locally symmetric space of finite volume and negative curvature, Sarnak and Wakayama show that all four sums (33)–(34) have the same asymptotic growth, with a subexponential error term. Lemma 4.1 explicates and sharpens this error term in the context of compact hyperbolic 3-manifolds.

Proof. Note that since Γ is discrete, there is a minimum geodesic length $\eta_0(\Gamma)$. When $y < \eta_0(\Gamma)$, all of the sums (33)–(34) vanish, and Lemma 4.1 holds vacuously; thus, we may assume that $y \geqslant \eta_0(\Gamma)$.

We will require, for $k \geqslant 2$ and $y > 0$, an estimate on the sum $J_k(y)$ defined by

$$J_k(y) := \sum_{\ell(\gamma) \leqslant y}^{P} \ell(\gamma) e^{-k\ell(\gamma)}.$$

Since $w(\gamma) \asymp_\Gamma e^{-\ell(\gamma)}$, Proposition 3.5 with $n = 0$ shows that

$$J^*(y) := \sum_{\ell(\gamma) \leqslant y}^{P} \ell(\gamma) e^{-\ell(\gamma)} \ll_\Gamma e^y.$$

In fact, a more precise bound $J^*(y) \ll e^y + \mathrm{O}_\Gamma\big(e^{\frac{2}{3}y} + e^{\nu_1 y}\big)$, with ν_1 as in (1) and to be omitted if Γ admits no complementary spectrum, follows from Proposition 3.5 with $n = 0$ or (essentially) from the Prime Geodesic Theorem (1), but we will not need this. Using integration by parts,

$$\begin{aligned} J_k(y) &= \int_{\eta_0(\Gamma)^-}^{y+} e^{-(k-1)t} \, \mathrm{d}J^*(t) = e^{-(k-1)t} J^*(t) \Big|_{\eta_0(\Gamma)^-}^{y+} + (k-1) \int_{\eta_0(\Gamma)}^{y} J^*(t) e^{-(k-1)t} \, \mathrm{d}t \\ &\ll_\Gamma \delta_2(k) y + e^{-k\eta_0(\Gamma)}, \end{aligned} \tag{39}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

where δ_2 is as in Section 1.4. With the estimate (39) at our disposal, we proceed to prove (35)–(37).

Proof of (35): Suppose $\tilde{\gamma} \in \Gamma$ is not primitive; then, $\tilde{\gamma} = \gamma^k$ for some primitive $\gamma \in \Gamma$ and $k \geqslant 2$. The length and holonomy are $\ell(\gamma^k) = k\ell(\gamma)$ and $\mathrm{hol}(\gamma^k) = k\,\mathrm{hol}(\gamma)$, respectively. Therefore,

$$|S_n(y) - S_n^P(y)| = \Big| \sum_{k\geqslant 2} \sum_{\ell(\gamma)\leqslant y/k}^{P} k\ell(\gamma) e^{-k\ell(\gamma)+ikn\,\mathrm{hol}(\gamma)} \Big| \leqslant \sum_{k\geqslant 2} k \sum_{\ell(\gamma)\leqslant y/k}^{P} \ell(\gamma) e^{-k\ell(\gamma)} = \sum_{k\geqslant 2} kJ_k(y/k),$$

where in fact the sum truncates at $k \leqslant y/\eta_0(\Gamma)$. Using the estimate (39), we find that

$$|S_n(y) - S_n^P(y)| \leqslant \sum_{k\geqslant 2} kJ_k(y/k) = \mathrm{O}_\Gamma(y),$$

as required.

Proof of (36): First, we have that

$$T_n^P(y) - S_n^P(y) = \sum_{\ell(\gamma)\leqslant y}^{P} \ell(\gamma)\big(w(\gamma) - e^{-\ell(\gamma)}\big)e^{in\,\mathrm{hol}(\gamma)}.$$

To simplify, observe that

$$w(\gamma) = |1 - e^{\mathbb{C}\ell(\gamma)}|^{-1}|1 - e^{-\mathbb{C}\ell(\gamma)}|^{-1} = e^{-\ell(\gamma)}|1 - e^{-\mathbb{C}\ell(\gamma)}|^{-2} = e^{-\ell(\gamma)}\big(1 + \mathrm{O}_\Gamma(e^{-\ell(\gamma)})\big).$$

Therefore, using again the estimate (39),

$$|T_n^P(y) - S_n^P(y)| \ll_\Gamma \sum_{\ell(\gamma)\leqslant y}^{P} \ell(\gamma)e^{-2\ell(\gamma)} = J_2(y) \ll_\Gamma y.$$

Proof of (37): As in the proof of (35), we have

$$T_n(y) - T_n^P(y) = \sum_{k\geqslant 2} \sum_{\ell(\gamma)\leqslant y/k}^{P} \ell(\gamma) w(\gamma^k) e^{ikn\,\mathrm{hol}(\gamma)}.$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Since $w(\gamma^k) = e^{-k\ell(\gamma)}(1 + \mathrm{O}_\Gamma(e^{-k\ell(\gamma)}))$, we thus have that

$$T_n(y) - T_n^P(y) \ll_\Gamma \sum_{k\geqslant 2} \sum_{\ell(\gamma)\leqslant y/k}^{P} \ell(\gamma) e^{-k\ell(\gamma)} = \sum_{k\geqslant 2} J_k(y/k) = \mathrm{O}_\Gamma(y),$$

as in the proof of (35). ∎

For future reference, we also include a version of Lemma 4.1 with more arbitrary (such as smooth) cutoffs. This presents no serious distinction, as the proof of Lemma 4.1 uses the cutoff only to control the set of geodesics entering the estimates, followed by term-wise estimates. For $n \in \mathbb{Z}$ and a bounded, compactly supported function $g : \mathbb{R} \to \mathbb{R}$, define

$$\begin{aligned} S_n^P[g] &= \sum_{[\gamma]}^{P} \ell(\gamma) g(\ell(\gamma)) e^{-\ell(\gamma)+in\,\mathrm{hol}\,\gamma}, \\ T_n[g] &= \sum_{[\gamma]} \ell(\gamma_0) g(\ell(\gamma)) w(\gamma) e^{in\,\mathrm{hol}\,\gamma}, \end{aligned} \tag{40}$$

with notation as in (33) and (34). The following lemma shows that these two sums are also comparatively very close.

Lemma 4.3. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $g : \mathbb{R} \to \mathbb{R}$ be a bounded function supported in $[-y, y]$. Then, the sums $S_n^P[g]$ and $T_n[g]$ defined in (40) satisfy

$$S_n^P[g] = T_n[g] + \mathrm{O}_\Gamma(\|g\|_\infty y).$$

Proof. Defining $S_n[g]$ and $T_n^P[g]$ in the obvious way, we find that

$$\begin{aligned} \left|T_n[g] - T_n^P[g]\right| &= \left|\left(\sum - \sum^{P}\right)_{[\gamma]} \ell(\gamma_0) g(\ell(\gamma)) w(\gamma) e^{in\,\mathrm{hol}(\gamma)}\right| \\ &\leqslant \|g\|_\infty \left(\sum - \sum^{P}\right)_{\ell(\gamma)\leqslant y} \ell(\gamma_0) w(\gamma). \end{aligned}$$

From here, an identical proof to the proof of (37) in Lemma 4.1 shows that

$$T_n[g] - T_n^P[g] \ll_\Gamma \|g\|_\infty y.$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

The proofs that all four sums $S_n[g]$, $S_n^P[g]$, $T_n^P[g]$, and $T_n[g]$ are within $O_\Gamma(\|g\|_\infty Y)$ of each other follow in the same way by bootstrapping the proof of Lemma 4.1, *mutatis mutandis*. ∎

5 Holonomy Character Sums

In this section, we prove estimates on the "holonomy character sums"

$$K_n[g_{Y,\eta}] = \sum_{[\gamma]}^{P} g_{Y,\eta}(\ell(\gamma))e^{in\,\mathrm{hol}(\gamma)}, \qquad K_n(Y) = \sum_{\ell(\gamma)\leqslant Y}^{P} e^{in\,\mathrm{hol}(\gamma)}, \tag{41}$$

where $Y, \eta > 0$, the cutoff function $g_{Y,\eta}$ is as in (27), and the summation is over all nontrivial primitive hyperbolic and loxodromic conjugacy classes $[\gamma]$ of Γ.

The sums $K_n[g_{Y,\eta}]$ and $K_n(Y)$ capture the primitive length spectrum of Γ with a smooth and sharp cutoff up to around $Y > 0$, respectively, weighted by characters $\chi_{0,n}$ (see (16)) on the holonomy group $T \cap \mathrm{PSU}_2 \simeq \mathbb{R}/2\pi\mathbb{Z}$. They play an analogous role to that of Dirichlet character sums in the context of the Prime Number Theorem in Arithmetic Progressions; in particular, asymptotics for the sum $K_0(Y)$ recover the Prime Geodesic Theorem, whereas for $n \neq 0$ the sums $K_n(Y)$ feature substantial cancellation, as shown in the following proposition.

Proposition 5.1 (Holonomy character sums). Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, cocompact, torsion-free subgroup, and let $Y > 0$, $n \in \mathbb{Z}$, and $0 < \eta \leqslant \eta_0$. Then, the sums $K_n[g_{Y,\eta}]$ and $K_n(Y)$ defined in (41) satisfy

$$\begin{aligned} K_n[g_{Y,\eta}] &= \delta_0(n)\int_2^\infty g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + O_{\Gamma,\eta_0}\Big(e^Y\Big(\frac{1}{Y\eta^2}+n^2+1\Big)\Big), \\ K_n(Y) &= \delta_0(n)\int_2^Y \mathrm{d}\varpi_\Gamma(u) + O_\Gamma\Big(\frac{e^{5Y/3}}{Y}+n^2e^Y\Big), \end{aligned} \tag{42}$$

where δ_0 and ϖ_Γ are as in Section 1.4 and (2).

Proof. We begin with $K_n(Y)$, which is technically simpler. Recall the sum $S_n^P(Y)$ defined in (33). Combining Proposition 3.5 and Lemma 4.1, we may write

$$\begin{aligned} S_n^P(Y) = \sum_{\ell(\gamma)\leqslant Y}^{P} \ell(\gamma)e^{-\ell(\gamma)+in\,\mathrm{hol}(\gamma)} &= \delta_0(n)\int_{-Y}^{Y}\mathrm{d}\varpi_\Gamma^*(u) + s_n^P(Y), \\ s_n^P(Y) &= O_\Gamma\big(e^{2Y/3}+n^2Y\big). \end{aligned} \tag{43}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

Since Γ is discrete, there is a minimum geodesic length $\eta_0(\Gamma)$. With an eye toward summation by parts, we first rewrite $K_n(y)$ as an integral, separating out the principal part:

$$K_n(y) = \int_{\eta_0(\Gamma)-}^{y+} \frac{e^t}{t}\, \mathrm{d}S_n^P(t) = \delta_0(n) \int_{\eta_0(\Gamma)-}^{y+} \frac{e^t}{t} \frac{\mathrm{d}}{\mathrm{d}t} \int_{-t}^{t} \mathrm{d}\varpi_\Gamma^*(u)\, \mathrm{d}t + \int_{\eta_0(\Gamma)-}^{y+} \frac{e^t}{t}\, \mathrm{d}s_n^P(t). \tag{44}$$

Recalling (2) and (21), the first term equals

$$\delta_0(n) \int_{\eta_0(\Gamma)}^{y} \frac{e^t}{t} \Big(\frac{\mathrm{d}\varpi_\Gamma^*}{\mathrm{d}u}\Big|_{u=t} + \frac{\mathrm{d}\varpi_\Gamma^*}{\mathrm{d}u}\Big|_{u=-t} \Big)\, \mathrm{d}t = \delta_0(n) \int_2^y \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_\Gamma\Big(\log^* y + \frac{e^{(1-\nu_k)y}}{y} \Big), \tag{45}$$

recalling from (21) the definition of $\mathrm{d}\varpi_\Gamma^*$ and from (2) the notation $1-\nu_j^2$ for the exceptional eigenvalues of the Laplacian, with $0 < \nu_k \leqslant \cdots \leqslant \nu_1 < 1$ (and we set formally $\nu_0 = 1$ if $k = 0$). Using integration by parts and the estimate (43), the second term in (44) is

$$\begin{aligned} \frac{e^t}{t} s_n^P(t)\Big|_{\eta_0(\Gamma)-}^{y+} - \int_{\eta_0(\Gamma)}^{y} \Big(\frac{e^t}{t} - \frac{e^t}{t^2} \Big) s_n^P(t)\, \mathrm{d}t &\ll_\Gamma \frac{e^{5y/3}}{y} + n^2 e^y + \int_{\eta_0(\Gamma)}^{y} \Big(\frac{e^{5t/3}}{t} + n^2 e^t \Big)\, \mathrm{d}t \\ &\ll_\Gamma \frac{e^{5y/3}}{y} + n^2 e^y. \end{aligned} \tag{46}$$

Combining (45) and (46) gives the desired asymptotic for $K_n(y)$.

Now, we turn to $K_n[g_{y,\eta}]$. This case presents a minor technical difficulty in that the first step in summation by parts (44) does not work as cleanly. To address this, we rework the proof of Lemma 3.1 by adjusting the choice of the test function $g_{y,\eta}$ from (27) to a slightly different test function $g_{t,\eta}^\lambda : \mathbb{R} \to \mathbb{C}$ given, for $0 < \eta \leqslant \eta_0 \leqslant t$ and $\lambda \in [0,1]$ as follows:

$$\psi_\eta^\lambda(x) = \psi_\eta(x) e^{\lambda x}, \quad g_{t,\eta}^\lambda = \chi_{[-\eta,t]} \star \psi_\eta^\lambda + \chi_{[-t,\eta]} \star \psi_\eta^{-\lambda}, \tag{47}$$

where ψ_η is as defined in (26). We remark that the asymptotic for $K_n[g_{y,\eta}]$ in (42) holds trivially for $y = \mathrm{O}(1)$ (due to the discreteness of Γ), so from now on we may assume that $y \geqslant \eta_0$. We compute, using the Schwartz bound for $\hat{\psi}$,

$$\begin{aligned} \hat{g}_{t,\eta}^\lambda(\xi) = \sum\nolimits_\pm \hat{\chi}_{\pm[-\eta,t]}(\xi) \hat{\psi}_\eta^{\pm\lambda}(\xi) &= \sum\nolimits_\pm \frac{e^{\pm 2\pi i \eta\xi} - e^{\mp 2\pi i t\xi}}{\pm 2\pi i \xi} \hat{\psi}\Big(\eta\xi \pm \frac{i\eta\lambda}{2\pi} \Big) \\ &\ll_{\eta_0, N} \min\Big(t, \frac{1}{|\xi|}, \frac{1}{|\xi|(\eta|\xi|)^N} \Big). \end{aligned}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

The function $g_{t,\eta}^{\lambda} : \mathbb{R} \to \mathbb{R}$ defined in (47) is smooth, even, and compactly supported, so it may be used in the trace formula of Theorem 2.2. It also satisfies $\hat{g}_{t,\eta}^{\lambda}(0) = 2(t+\eta)\hat{\psi}(i\eta\lambda/2\pi)$, $g_{t,\eta}^{\lambda}(0) = 2\hat{\psi}(i\lambda\eta/2\pi) = \mathrm{O}_{\eta_0}(1)$, $(g_{t,\eta}^{\lambda})''(0) = 0$. Running the proof of Lemma 3.1 with $g_{t,\eta}^{\lambda}$ in place of $g_{Y,\eta}$ gives

$$T_n^{\cos}[g_{t,\eta}^{\lambda}] = \delta_0(n)\int_{-\infty}^{\infty} g_{t,\eta}^{\lambda}(u)\,\mathrm{d}\varpi_{\Gamma}^{*}(u) + \mathrm{O}_{\Gamma,\eta_0}\left(\frac{1}{\eta^2} + (1+n^2)\left(\log^{*}\frac{1}{\eta} + t\right)\right). \tag{48}$$

Defining analogously $h_{t,\eta}^{\lambda} = \chi_{[-\eta,t]} \star \psi_{\eta}^{\lambda} - \chi_{[-t,\eta]} \star \psi_{\eta}^{-\lambda}$, we find as in Lemma 3.1 that

$$T_n^{\sin}[h_{t,\eta}^{\lambda}] = \mathrm{O}_{\Gamma,\eta_0}\left(\frac{1}{\eta^2} + (1+n^2)\left(\log^{*}\frac{1}{\eta} + t\right)\right) \tag{49}$$

and note that $h_{t,\eta}^{\lambda} = g_{t,\eta}^{\lambda} \cdot \operatorname{sgn}$ outside $[-2\eta_0, 2\eta_0]$.

Note that $\|g_{t,\eta}^{\lambda}\|_{\infty} \leqslant 2\hat{\psi}(i\eta\lambda/2\pi) = \mathrm{O}_{\eta_0}(1)$ and $\operatorname{supp} g_{t,\eta}^{\lambda} \subseteq [-t-\eta, t+\eta]$, and recall the sum $S_n^{P}[g_{t,\eta}^{\lambda}]$ defined in (40). Combining (48), (49), and Lemma 4.3, we may write

$$S_n^{P}[g_{t,\eta}^{\lambda}] = \sum_{[\gamma]}^{P} \ell(\gamma) g_{t,\eta}^{\lambda}(\ell(\gamma)) e^{-\ell(\gamma)+in\operatorname{hol}(\gamma)} = \delta_0(n)\int_{-\infty}^{\infty} g_{t,\eta}^{\lambda}(u)\,\mathrm{d}\varpi_{\Gamma}^{*}(u) + s_n^{P}[g_{t,\eta}^{\lambda}],$$

$$s_n^{P}[g_{t,\eta}^{\lambda}] = \mathrm{O}_{\Gamma,\eta_0}\left(\frac{1}{\eta^2} + (1+n^2)\left(\log^{*}\frac{1}{\eta} + t\right)\right). \tag{50}$$

From the definition (47), we have that

$$\begin{aligned}(\mathrm{d}g_{t,\eta}^{\lambda}/\mathrm{d}t)(\ell) &= \psi_{\eta}^{\lambda}(\ell - t) + \psi_{\eta}^{-\lambda}(\ell + t)\\ &= \psi_{\eta}(\ell - t)e^{\lambda(\ell-t)} + \psi_{\eta}(\ell + t)e^{-\lambda(\ell+t)}.\end{aligned}$$

Therefore,

$$\widetilde{K}_n^{\lambda}[g_{Y,\eta}] := \int_{\eta_0+}^{Y} e^{\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}S_n^{P}[g_{t,\eta}^{\lambda}]\,\mathrm{d}t = K_n^{\lambda}[g_{Y,\eta}] + \mathrm{O}_{\Gamma,\eta_0}(1), \tag{51}$$

where

$$K_n^{\lambda}[g_{Y,\eta}] = \sum_{[\gamma]}^{P} \ell(\gamma) g_{Y,\eta}(\ell(\gamma)) e^{(\lambda-1)\ell(\gamma)+in\operatorname{hol}\gamma}$$

and the $\mathrm{O}_{\Gamma,\eta_0}(1)$ term accounts for classes $[\gamma]$ with $\ell(\gamma) \leqslant 2\eta_0$ (in particular, this harmless term may be omitted if $\eta_0 \leqslant \frac{1}{2}\eta_0(\Gamma)$). Following the argument in (44), we

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

separate the integral representing $\widetilde{K}_n^{\lambda}[g_{Y,\eta}]$ as

$$\widetilde{K}_n^{\lambda}[g_{Y,\eta}] = \delta_0(n)\int_{\eta_0}^{Y} e^{\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty} g_{t,\eta}^{\lambda}(u)\,\mathrm{d}\varpi_{\Gamma}^{*}(u)\,\mathrm{d}t + \int_{\eta_0+}^{Y} e^{\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}s_n^P[g_{t,\eta}^{\lambda}]\,\mathrm{d}t,$$

where the first term equals, with notation as in (45),

$$\delta_0(n)\int_2^{\infty} g_{Y,\eta}(u)e^{\lambda u}\,\mathrm{d}\varpi_{\Gamma}^{*}(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(Y + \mathbf{1}_{\lambda>\nu_k+1/Y}\frac{e^{(\lambda-\nu_k)Y}}{\lambda-\nu_k}\Big). \tag{52}$$

For the second term, we use integration by parts and estimate (50) to find that it equals

$$e^{\lambda t}s_n^P[g_{t,\eta}^{\lambda}]\Big|_{\eta_0+}^{Y} - \lambda\int_{\eta_0}^{Y} s_n^P[g_{t,\eta}^{\lambda}]e^{\lambda t}\,\mathrm{d}t \ll_{\Gamma,\eta_0} e^{\lambda Y}\Big(\frac{1}{\eta^2} + (1+n^2)\Big(\log^*\frac{1}{\eta}+Y\Big)\Big). \tag{53}$$

Combining (51), (52), and (53) gives an asymptotic for $K_n^{\lambda}[g_{Y,\eta}]$. Finally, we recover the desired sum $K_n[g_{Y,\eta}]$ defined in (41) as

$$\begin{aligned} K_n[g_{Y,\eta}] &= \sum_{[\gamma]}^{P} e^{in\,\mathrm{hol}(\gamma)}g_{Y,\eta}(\ell(\gamma))\ell(\gamma)e^{-\ell(\gamma)}\Big[\int_0^1 e^{\lambda\ell(\gamma)}\,\mathrm{d}\lambda + \frac{1}{\ell(\gamma)}\Big] \\ &= \int_0^1 K_n^{\lambda}[g_{Y,\eta}]\,\mathrm{d}\lambda + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y}\Big) \\ &= \delta_0(n)\int_2^{\infty} g_{Y,\eta}(u)\,\mathrm{d}\varpi_{\Gamma}(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y}\Big(\frac{1}{\eta^2} + (1+n^2)\Big(\log^*\frac{1}{\eta}+Y\Big)\Big)\Big), \end{aligned} \tag{54}$$

keeping in mind the definition (2), and using, for example, the already proved asymptotic (55) for $K_0(Y)$ at the second step. We may assume that $\log^*(1/\eta) = \mathrm{O}_{\Gamma}(Y)$ since otherwise (in light of (2) and (55)) the error term for $K_n[g_{Y,\eta}]$ in (42) clearly dominates all other terms. With this, (54) completes the proof of Proposition 5.1. ■

Remark 5.2. As already remarked, Proposition 5.1 contains as a special case the Prime Geodesic Theorem in the form

$$\pi_{\Gamma}(Y) = K_0(Y) = \int_2^{Y}\mathrm{d}\varpi_{\Gamma}(u) + \mathrm{O}_{\Gamma}\Big(\frac{e^{5Y/3}}{Y}\Big), \tag{55}$$

which we record here for future reference.

It is also instructive to consider how the asymptotic obtained in Proposition 5.1 for $K_n[g_{Y,\eta}]$ evolves as $\eta > 0$ varies from η of constant size, a case we may think of as a model of summation with a smooth cutoff in length on a $\mathrm{O}(1)$ scale, down to $\eta = e^{-Y/3}$,

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

which essentially corresponds to the sharp cutoff in $K_n(y)$ (cf. proof of Proposition 3.5). For $\eta = \eta_0$ of constant size, Proposition 5.1 states that

$$K_n[g_{y,\eta_0}] = \delta_0(n) \int_2^\infty g_{y,\eta_0}(u)\, \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\big((n^2+1)e^y\big). \tag{56}$$

Thus, for a fixed $n \neq 0$, the sum $K_n[g_{y,\eta_0}]$ consisting of $\sim_{\Gamma,\eta_0} e^{2y}/2y$ terms of unit size exhibits essentially square-root cancellation as $y \to \infty$. In fact, power-saving cancellation in $K_n[g_{y,\eta}]$ and $K_n(y)$ persists in the range $n \ll e^{(1/2-\delta)y}$, a statement which should be compared with the range of uniformity in the prime number theorem for arithmetic progressions to large moduli.

As η decreases and the cutoff in $K_n[g_{y,\eta}]$ becomes steeper, the first error term (which is independent of n) becomes more pronounced and, for $\eta = e^{-y/3}$, essentially recovers the first error term in the asymptotic for the sharp count $K_n(y)$. This term, which dominates for a fixed $n \in \mathbb{Z}$ and $y \to \infty$, is rooted in the passage to the sharp count. The second error term detects for large $n \in \mathbb{Z}$ the influence of the oscillating holonomy factor $e^{in\,\mathrm{hol}(\gamma)}$.

6 Ambient Prime Geodesic Theorems

In this section, we prove our principal results providing counts for prime geodesics on M with control on their length and holonomy simultaneously.

The first of the two main results of this section is the Ambient Prime Geodesic Theorem (Theorem 6.5), which features a sharp cutoff both in length $\ell(\gamma) \leqslant y$ and holonomy, as in (4). For many analytic purposes, including the existence and properties of various limiting distributions (see, e.g., Corollary 6.3 of Theorem 6.2), smooth cutoff results suffice. Since the passage to the sharp count is the leading contributor to error terms, and to emphasize the parallel between length and holonomy aspects, we provide $4 = 2 \times 2$ propositions in Sections 6.1 and 6.2, featuring each combination of smooth/sharp counts in length/holonomy, with explicit error terms depending on the steepness of the smooth cutoff. Further, in the concluding Section 6.3, we consider the short-range ambient counting problems and show how a consistent ambiental passage from smooth to sharp counting leads to stronger corresponding short-range counts with smooth/sharp cutoffs, including our second main result, Theorem 6.12.

To effectuate the transition from a smooth to sharp cutoff in holonomy, we will use, for an arbitrary interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$ and $0 < \eta' \leqslant 2\pi$ the function $f_{J,\eta'} : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}$ given by

$$f_{J,\eta'}(t) = \sum_{n\in\mathbb{Z}} \int_J \psi_{\eta'}(t + 2n\pi - \theta)\, \mathrm{d}\theta, \tag{57}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

with $\psi_{\eta'}$ as in (26). In other words, $f_{J,\eta'}$ is the $2\pi\mathbb{Z}$-periodic convolution $\chi_J \star \psi_{\eta'}$ and plays for holonomy the role of $g_{Y,\eta}$ and $h_{Y,\eta}$ of (27) for lengths. In particular, $f_{[\theta,\theta'],\eta'}$ is a smooth, non-negative cutoff function of height 1, which is supported on $[\theta-\eta',\theta'+\eta']+2\pi\mathbb{Z}$ and agrees with $\chi_{[\theta,\theta']}$ outside the set $([\theta-\eta',\theta+\eta']\cup[\theta'-\eta',\theta'+\eta'])+2\pi\mathbb{Z}$.

6.1 Smooth Count

We prove our prime geodesic theorems in holonomy classes by spectrally decomposing the holonomy and then invoking estimates on holonomy character sums in Proposition 5.1. We observed in Proposition 5.1 and Remark 5.2 that, when lengths are weighted with a smooth function $g_{Y,\eta}$ with not too small $\eta > 0$ (say, of fixed size), so that the cutoff is not too steep, one obtains strong asymptotics for holonomy character sums with very modest error terms, such as essentially square-root cancellation in (56).

Our first proposition is the baseline count for sampling geodesics with a smooth function in both the length and the holonomy. For cutoffs of fixed steepness (equivalently with fixed "uncertainty windows," which we denote by $\eta,\eta' > 0$) in both the length and the holonomy, Proposition 6.1 features an error term of essentially square-root strength, well sharper than the sharp cutoff counts in either direction in Proposition 6.4 and Theorems 6.2 and 6.5 below.

Proposition 6.1. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every smooth function $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$ and every $Y > 0$ and $0 < \eta \leqslant \eta_0$,

$$\begin{aligned}\pi_\Gamma(g_{Y,\eta},f) &:= \sum_{[\gamma]}^{P} f(\mathrm{hol}(\gamma))g_{Y,\eta}(\ell(\gamma))\\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \int_2^\infty g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y\eta^2}\|\hat{f}\|_1 + e^Y\|\hat{f}\|_{2,1}\Big)\\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \pi_\Gamma(g_{Y,\eta}) + \mathrm{O}_{\Gamma,f,\eta}(e^Y),\end{aligned}$$

where ϖ_Γ is as in (2) and $\|\hat{f}\|_{2,1} = \|\hat{f}\|_1 + \|\widehat{f''}\|_1$. In particular, for every interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$, $0 < \eta' \leqslant 2\pi$, and $f_{J,\eta'} : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}$ as in (57),

$$\pi_\Gamma(g_{Y,\eta},f_{J,\eta'}) = \frac{|J|}{2\pi}\int_2^\infty g_{Y,\eta}(u)\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big).$$

Proof. Using the Fourier expansion $f(\theta) = (1/2\pi)\sum_{n\in\mathbb{Z}}\hat{f}(n)e^{in\theta}$ and applying Proposition 5.1 to estimate the resulting holonomy character sums $K_n[g_{Y,\eta}]$ defined in (41), we

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

find that

$$\sum_{[\gamma]}^{P} f(\mathrm{hol}(\gamma))g_{Y,\eta}(\ell(\gamma)) = \frac{1}{2\pi}\sum_{[\gamma]}^{P} g_{Y,\eta}(\ell(\gamma))\sum_{n\in\mathbb{Z}}\hat{f}(n)e^{in\,\mathrm{hol}\,\gamma} = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}}\hat{f}(n)K_n[g_{Y,\eta}]$$
$$= \frac{1}{2\pi}\hat{f}(0)\pi_\Gamma(g_{Y,\eta}) + \mathrm{O}_{\Gamma,\eta_0}\Big(\sum_{n\neq 0}|\hat{f}(n)|e^Y\Big(\frac{1}{Y\eta^2}+n^2+1\Big)\Big),$$

which completes the proof for a general smooth $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, keeping in mind that $\pi_\Gamma(g_{Y,\eta}) = K_0[g_{Y,\eta}]$ and Proposition 5.1. For the specific function $f_{J,\eta'}$ defined in (57), we compute its Fourier coefficients by the usual unfolding argument as

$$\hat{f}_{J,\eta'}(n) = \hat{\chi}_J(n)\hat{\psi}(\eta' n) = \frac{e^{-in\theta'}-e^{-in\theta}}{-in}\hat{\psi}(\eta' n),$$

where $J = [\theta, \theta']$. Using the Schwartz estimate $|\hat{\psi}(\eta' n)| \ll_k 1/(1+\eta'|n|)^k$, we can bound

$$\begin{aligned}\sum_{n\in\mathbb{Z}}|\hat{f}_{J,\eta'}(n)| &\ll |J| + \sum_{n\neq 0}\frac{|\hat{\psi}(\eta' n)|}{n} \ll \sum_{n\leqslant 1/\eta'}\frac{1}{n} + \sum_{n>1/\eta'}\frac{1}{\eta'^2 n^3} \ll \log^*\frac{1}{\eta'},\\ \sum_{n\in\mathbb{Z}}(1+n^2)|\hat{f}_{J,\eta'}(n)| &\ll |J| + \sum_{n\neq 0} n|\hat{\psi}(\eta' n)| \ll \sum_{n\leqslant 1/\eta'} n + \sum_{n>1/\eta'}\frac{n}{\eta'^4 n^4} \ll \frac{1}{\eta'^2}.\end{aligned} \tag{58}$$

This completes the proof. ■

6.2 Passage to Sharp Counts and Prime Geodesic Theorems

In this section, we replace the smooth cutoff in Proposition 6.1 by a sharp cutoff in one or both of the length and holonomy. Theorem 6.2 features the familiar sharp cutoff in length, a hallmark of a traditional Prime Geodesic Theorem, and implies effective equidistribution of holonomy in short intervals of length. To stress the conceptual symmetry between the two parameters, we also prove Proposition 6.4, an asymptotic count with a sharp cutoff in holonomy and smoothly sampled length. Finally, in Theorem 6.5 and its Corollary 6.7, we prove ambient prime geodesic counts with a sharp cutoff in both length and holonomy.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Theorem 6.2. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$ be an arbitrary smooth function. Then, for $y > 0$,

$$\pi_\Gamma(y,f) := \sum_{\ell(\gamma)\leqslant y}^{P} f(\mathrm{hol}(\gamma)) = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \int_2^y \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_\Gamma\Big(\|\hat{f}\|_1 \frac{e^{5y/3}}{y} + \|\widehat{f''}\|_1 e^y\Big)$$
$$= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta \cdot \pi_\Gamma(y) + \mathrm{O}_{\Gamma,f}\Big(\frac{e^{5y/3}}{y}\Big),$$

where ϖ_Γ is as in (2). In particular, for every interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$, $0 < \eta' \leqslant 2\pi$, and $f_{J,\eta'} : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}$ as in (57),

$$\pi_\Gamma(y,f_{J,\eta'}) = \frac{|J|}{2\pi}\pi_\Gamma(y) + \mathrm{O}_\Gamma\Big(\frac{e^{5y/3}}{y}\log^*\frac{1}{\eta'} + \frac{e^y}{\eta'^2}\Big).$$

Proof. Using the Fourier expansion $f(\theta) = (1/2\pi)\sum_{n\in\mathbb{Z}}\hat{f}(n)e^{in\theta}$ and applying Proposition 5.1 to estimate the resulting holonomy character sums $K_n(y)$ defined in (41), we find that

$$\sum_{\ell(\gamma)\leqslant y}^{P} f(\mathrm{hol}(\gamma)) = \frac{1}{2\pi}\sum_{\ell(\gamma)\leqslant y}^{P}\sum_{n\in\mathbb{Z}}\hat{f}(n)e^{in\,\mathrm{hol}\,\gamma} = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}}\hat{f}(n)K_n(y)$$
$$= \frac{1}{2\pi}\hat{f}(0)\pi_\Gamma(y) + \mathrm{O}_\Gamma\Big(\frac{e^{5y/3}}{y}\sum_{n\neq 0}|\hat{f}(n)| + e^y\sum_{n\neq 0}n^2|\hat{f}(n)|\Big),$$

which completes the proof for a general smooth $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, keeping in mind (55). The final claim follows by specializing these bounds to $f_{J,\eta'}$ and using estimates (58). ■

For a fixed smooth $f : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}$, Theorem 6.2 recovers [25, Theorem 1] in the present setting of compact hyperbolic 3-manifolds. An immediate corollary of Theorem 6.2, coupled with the Prime Geodesic Theorem (55), is the following equidistribution statement.

Corollary 6.3 (Equidistribution of holonomy). Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, the holonomy of geodesics on $\Gamma\backslash\mathbb{H}^3$ of length $\ell(\gamma) \leqslant y$ is equidistributed in $\mathbb{R}/2\pi\mathbb{Z}$ as $y \to \infty$. In fact, given any collection of intervals $I_n = [y_n', y_n]$ $(0 \leqslant y_n' \leqslant y_n)$ satisfying $|I_n| / e^{-y_n/3} \to \infty$,

$$\frac{1}{\pi_\Gamma(I_n)}\sum_{\ell(\gamma)\in I_n}\delta_{\{\mathrm{hol}(\gamma)\}} \xrightarrow{\text{weak}^*} \frac{1}{2\pi}\mathrm{d}\theta_{\mathbb{R}/2\pi\mathbb{Z}} \quad (n \to \infty).$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Now we transition from a smooth function on the holonomy to a sharp cutoff, while still using a smooth function on the length. We do so using the smooth cutoff function $f_{J,\eta'}$ defined in (57) and then optimizing the choice of η'.

Proposition 6.4. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every interval $J = [\theta, \theta'] \subseteq \mathbb{R}/2\pi\mathbb{Z}$ and every $Y > 0$ and $0 < \eta \leqslant \eta_0$,

$$\pi_\Gamma(g_{Y,\eta}, J) := \sum_{\mathrm{hol}(\gamma)\in J}^{P} g_{Y,\eta}(\ell(\gamma)) = \frac{|J|}{2\pi}\int_2^\infty g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^{5Y/3}}{Y^{2/3}} + \frac{e^Y}{\eta^2}\Big)$$
$$= \frac{|J|}{2\pi}\pi_\Gamma(g_{Y,\eta}) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^{5Y/3}}{Y^{2/3}} + \frac{e^Y}{\eta^2}\Big).$$

Proof. Our starting point is Proposition 6.1, with a parameter $0 < \eta' \leqslant 2\pi$ to be suitably chosen soon, which yields the estimate

$$\pi_\Gamma(g_{Y,\eta}, f_{J,\eta'}) = \frac{|J|}{2\pi}\int_2^\infty g_{Y,\eta}(u)\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big).$$

According to definition (57), the smooth cutoff function $f_{J,\eta'}$ for $J = [\theta, \theta']$ agrees with χ_J outside the set $([\theta - \eta', \theta + \eta'] \cup [\theta' - \eta', \theta' + \eta']) + 2\pi\mathbb{Z}$, on which $|f_{J,\eta'} - \chi_J| = \mathrm{O}(1)$. We also recall that, according to definition (27), the cutoff function $g_{Y,\eta}$ is supported on $[-Y - \eta, Y + \eta]$ and satisfies $|g_{Y,\eta}| \leqslant 1$. Therefore,

$$\big|\pi_\Gamma(g_{Y,\eta}, J) - \pi_\Gamma(g_{Y,\eta}, f_{J,\eta'})\big| = \Big|\sum_{[\gamma]}^{P}(\chi_J - f_{J,\eta'})(\mathrm{hol}(\gamma))g_{Y,\eta}(\ell(\gamma))\Big|$$
$$\ll \sum_{\substack{\ell(\gamma)\leqslant Y+\eta \\ \theta-\eta'\leqslant\mathrm{hol}(\gamma)\leqslant\theta+\eta'}}^{P} 1 + \sum_{\substack{\ell(\gamma)\leqslant Y+\eta \\ \theta'-\eta'\leqslant\mathrm{hol}(\gamma)\leqslant\theta'+\eta'}}^{P} 1.$$

The latter terms are of the form ready to be estimated using Lemma 3.6, which for every $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ yields

$$\sum_{\substack{\ell(\gamma)\leqslant Y+\eta \\ \theta-\eta'\leqslant\mathrm{hol}(\gamma)\leqslant\theta+\eta'}}^{P} 1 \ll \frac{e^{Y+\eta}}{Y}\sum_{\substack{\ell(\gamma)\leqslant Y+\eta \\ \theta-\eta'\leqslant\mathrm{hol}(\gamma)\leqslant\theta+\eta'}} \ell(\gamma_0)w(\gamma) \ll_{\Gamma,\eta_0} \eta'\frac{e^{2Y}}{Y} + \frac{e^Y}{\eta'^2}. \tag{59}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Using Proposition 6.1 and the input from Lemma 3.6 in the form (59), we have that

$$\begin{aligned}\left|\pi_\Gamma(g_{Y,\eta},J)-\frac{|J|}{2\pi}\int_2^Y g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u)\right| &\\ \leqslant \left|\pi_\Gamma(g_{Y,\eta},J)-\pi_\Gamma(g_{Y,\eta},f_{J,\eta'})\right| + \left|\pi_\Gamma(g_{Y,\eta},f_{J,\eta'})-\frac{|J|}{2\pi}\int_2^\infty g_{Y,\eta}(u)\mathrm{d}\varpi_\Gamma(u)\right| &\\ \ll_{\Gamma,\eta_0} \eta'\frac{e^{2Y}}{Y}+\frac{e^Y}{\eta'^2}+\frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'}.\end{aligned} \tag{60}$$

Taking the admissible choice $\eta' = \min(Y^{1/3}e^{-Y/3}, 2\pi)$ completes the proof of the first claim, since for $Y = \mathrm{O}(1)$ the left-hand side is trivially $\mathrm{O}_{\Gamma,\eta_0}(1)$.

The second claim follows immediately, keeping in mind that by Proposition 5.1

$$\pi_\Gamma(g_{Y,\eta}) = K_0[g_{Y,\eta}] = \int_2^\infty g_{Y,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^Y}{Y\eta^2}+e^Y\Big),$$

or alternatively by adapting the use of Proposition 6.1 in (60). ∎

Our main result, the following Theorem 6.5, features a sharp cutoff in both length and holonomy. Such a theorem can be proved by passing from the remaining smooth to sharp cutoff in either Theorem 6.2 or Proposition 6.4; we choose the former route.

Theorem 6.5 (Ambient Prime Geodesic Theorem). Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, cocompact, torsion-free subgroup. Then, for every $Y > 0$ and every interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$,

$$\begin{aligned}\pi_\Gamma(Y,J) &:= \big|\big\{[\gamma]^P : \ell(\gamma)\leqslant Y,\ \mathrm{hol}(\gamma)\in J\big\}\big| \\ &= \iint\limits_{[2,Y]\times J}\mathrm{d}\varpi_\Gamma(u)\frac{\mathrm{d}\theta}{2\pi}+\mathrm{O}_\Gamma\big(e^{5Y/3}\big) = \frac{|J|}{2\pi}\pi_\Gamma(Y)+\mathrm{O}_\Gamma\big(e^{5Y/3}\big),\end{aligned}$$

where ϖ_Γ is as in (2).

Proof. We begin with Theorem 6.2, with a parameter $0 < \eta' \leqslant 2\pi$ to be suitably chosen soon. Recall from (57) that the smooth cutoff function $f_{J,\eta'}$ for $J = [\theta,\theta']$ agrees with χ_J outside the set $([\theta-\eta',\theta+\eta']\cup[\theta'-\eta',\theta'+\eta'])+2\pi\mathbb{Z}$, on which $|f_{J,\eta'}-\chi_J| = \mathrm{O}(1)$. Arguing

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

as in the proof of Proposition 6.4, we find using Lemma 3.6 that

$$\begin{aligned}\left|\pi_\Gamma(y,J)-\pi_\Gamma(y,f_{J,\eta'})\right| &\ll \sum_{\substack{\ell(\gamma)\leqslant y\\ \theta-\eta'\leqslant \mathrm{hol}(\gamma)\leqslant\theta+\eta'}}^{P} 1+\sum_{\substack{\ell(\gamma)\leqslant y\\ \theta'-\eta'\leqslant \mathrm{hol}(\gamma)\leqslant\theta'+\eta'}}^{P} 1\\ &\ll \frac{e^y}{y}\sum_{\substack{\ell(\gamma)\leqslant y\\ \theta-\eta'\leqslant \mathrm{hol}(\gamma)\leqslant\theta+\eta'\\ \text{or } \theta'-\eta'\leqslant \mathrm{hol}(\gamma)\leqslant\theta'+\eta'}} \ell(\gamma_0)w(\gamma)\ll_\Gamma \eta'\frac{e^{2y}}{y}+\frac{e^y}{\eta'^2}. \end{aligned} \tag{61}$$

Using Theorem 6.2 and the input from Lemma 3.6 in the form (61), we have that

$$\begin{aligned}&\left|\pi_\Gamma(y,J)-\iint\limits_{[2,y]\times J} \mathrm{d}\varpi_\Gamma(u)\frac{\mathrm{d}\theta}{2\pi}\right|\\ &\quad\leqslant \left|\pi_\Gamma(y,J)-\pi_\Gamma(y,f_{J,\eta'})\right|+\left|\pi_\Gamma(y,f_{J,\eta'})-\frac{|J|}{2\pi}\int_{[2,y]}\mathrm{d}\varpi_\Gamma(u)\right|\\ &\quad\ll_\Gamma \eta'\frac{e^{2y}}{y}+\frac{e^y}{\eta'^2}+\frac{e^{5y/3}}{y}\log^*\frac{1}{\eta'}.\end{aligned}$$

Taking the admissible choice $\eta'=\min(e^{-y/3},2\pi)$ completes the proof, noting that for $y=\mathrm{O}(1)$ the left-hand side is $\mathrm{O}_\Gamma(1)$, and keeping in mind (55). ■

Remark 6.6. Observe that Proposition 6.4, which features a sharp cutoff in holonomy and a smooth cutoff in length, parallels Theorem 6.2, which has a smooth cutoff in holonomy and a sharp cutoff in length. This strengthens the perspective that length and holonomy should be counted as a pair. Compared with the smooth cutoff in Proposition 6.1, the smooth to sharp transition in either direction yields a significant contribution to the error term; however, the further passage to a sharp cutoff in both directions in Theorem 6.5 leads to minimal or no increase in the error term.

Conceptually (and tracing through the proofs confirms this rigorously), this is so because using smooth functions (27) and (57) to approximate sharp cutoffs essentially mimics what would be the use of Theorem 2.1 with an (ineligible) sharp cutoff test function $\chi_R(t)$ for a target rectangle $R\subseteq\mathbb{R}\times(\mathbb{R}/2\pi\mathbb{Z})$ of length and holonomy. The effect of a smooth cutoff of wall length $\eta\ll 1$ in one direction is that the essential support (in the sense of Schwartz decay) of the Fourier transform extends up to $\asymp 1/\eta$ in the dual-spectral direction. Since typically $\widehat{\chi_R}(\chi_{\nu,p})\asymp_R((1+|\nu|)(1+|p|))^{-1}$, the contribution of the principal series terms is guided roughly (up to logarithmic factors) by the supremum of the Plancherel measure (23) over the said spectral support, which is in turn symmetric

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

in spectral parameters ν and p and does not increase if the support extends in both parameters rather than just one.

Already as an immediate consequence of Theorem 6.5, we obtain the following ambient short-range count for primitive geodesics on M, with the length and holonomy simultaneously restricted to short intervals. We will further improve upon this asymptotic in Theorem 6.12.

Corollary 6.7. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for any intervals $I = [y', y]$ $(0 \leqslant y' \leqslant y)$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$,

$$\begin{aligned}\pi_\Gamma(I,J) &:= \left|\left\{[\gamma]^P : (\ell(\gamma), \mathrm{hol}(\gamma)) \in I \times J\right\}\right| \\ &= \iint\limits_{I\times J} \mathrm{d}\varpi_\Gamma(u)\frac{\mathrm{d}\theta}{2\pi} + \mathrm{O}_\Gamma\left(e^{5y/3}\right).\end{aligned}$$

Remark 6.8. Corollary 6.7 provides an exponent-saving asymptotic for $\pi_\Gamma(I,J)$ as long as $|I \times J| \gg e^{(-1/3+\delta)y}$ for some $\delta > 0$. We emphasize that the lengths $|I|$ and $|J|$ may be short *independently of each other* in any regime satisfying this combined lower bound.

When Corollary 6.7 gives an asymptotic, this may be rewritten (with obvious shorthand notation) as $\pi_\Gamma(I,J) \sim_\Gamma \pi_\Gamma(I) \cdot |J|/(2\pi) \sim_\Gamma \mathrm{Ei}_\Gamma(I) \cdot |J|/(2\pi)$, which can in turn be restated as an effective equidistribution statement for either the lengths or the holonomies in shrinking intervals of lengths, holonomies, or both.

6.3 Ambient Prime Geodesic Theorems for Shrinking Intervals

Results of Section 6.2 establish structural parallels between the length and holonomy aspects in geodesic counting and indicate that the "ambiental" joint count of primitive classes $[\gamma] \in [\Gamma]$ according to the pair $(\ell(\gamma), \mathrm{hol}(\gamma))$ is perhaps the most natural counting object. In this section, we demonstrate how consistently executing "ambiental" passage from smooth to sharp count leads to ambient prime geodesic theorems, which improve upon the results of Section 6.2 when both intervals of length and holonomy are shrinking. To emphasize the symmetry and analogy to Section 6.2, in Proposition 6.11 we prove counts that are smooth (but possibly steep) in one of the parameters and sharp in the other, and in Theorem 6.12 our headline count that is sharp in both aspects.

For the smooth length cutoff, we use

$$g_{I,\eta} = \psi_\eta \star (\chi_I + \chi_{-I}), \tag{62}$$

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

where ψ_η is defined as in (26) and χ_I and χ_{-I} are characteristic functions, to sample the geodesics with length in an interval $I \subseteq [0, y]$. For the smooth holonomy cutoff on an interval $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$, we use the function $f_{J,\eta'}$ defined in (57), which is the periodization of $\psi_\eta \star \chi_J$. Further, for $I = [a, b]$ and $\eta > 0$, we will write $I_\eta^- = [a-\eta, a+\eta]$, $I_\eta^+ = [b-\eta, b+\eta]$, and $I_\eta = I_\eta^- \cup I \cup I_\eta^+ = [a-\eta, b+\eta]$.

To begin, we modify a consequence of Proposition 6.1 to obtain a smooth count in both the length and holonomy aspects.

Lemma 6.9. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every two intervals $I = [y', y]$ $(0 \leqslant y' < y)$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$, and every $0 < \eta \leqslant \eta_0$ and $0 < \eta' \leqslant 2\pi$,

$$\pi_\Gamma(g_{I,\eta}, f_{J,\eta'}) = \frac{|J|}{2\pi} \int_2^\infty g_{I,\eta}(u)\, \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\frac{e^y}{y\eta^2} \log^* \frac{1}{\eta'} + \frac{e^y}{\eta'^2}\Big),$$

where $g_{I,\eta}$ and $f_{J,\eta'}$ are smooth cutoff functions defined in (62) and (57), and ϖ_Γ is the density (2).

Proof. The result follows immediately from subtracting two instances of Proposition 6.1, noting that $g_{I,\eta} = g_{y,\eta} - g_{y',\eta}$. ∎

The following corollary, which may be thought of as the ambient analogue of Lemmata 3.3 and 3.6, will be used to give an upper bound on the boundary terms when transitioning from a smooth to a sharp cutoff. In our typical application of Corollary 6.10, at least one of the intervals I and J will be very short, and we pick $\eta \asymp \min(|I|, \eta_0)$ or $\eta' \asymp |J|$.

Corollary 6.10. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for every two intervals $I = [y', y]$ $(0 \leqslant y' < y)$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$, we have

$$\pi_\Gamma(I, J) \ll_{\Gamma,\eta_0} (|I| + \eta)(|J| + \eta')\frac{e^{2y}}{y} + \frac{e^y}{y\eta^2} \log^* \frac{1}{\eta'} + \frac{e^y}{\eta'^2}$$

for every choice of $0 < \eta \leqslant \eta_0$ and $0 < \eta' \leqslant 2\pi$.

Proof. It is clear from definitions (57) and (62) that $\chi_{I\cup(-I)} \leqslant g_{I_\eta,\eta}$ and $\chi_{J+2\pi\mathbb{Z}} \leqslant f_{J_{\eta'},\eta'}$. Combining this observation with Lemma 6.9, and keeping in mind that $g_{I_\eta,\eta}$ is supported

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

inside $I_{2\eta} \cup (-I_{2\eta})$, we obtain

$$\begin{aligned}\pi_\Gamma(I,J) &\leqslant \pi_\Gamma(g_{I_\eta,\eta}, f_{J_{\eta'},\eta'}) \\ &\ll_{\Gamma,\eta_0} \frac{|J_{\eta'}|}{2\pi}\int_2^\infty g_{I_\eta,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \frac{e^Y}{y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2} \\ &\ll_{\Gamma,\eta_0} (|I|+\eta)(|J|+\eta')\frac{e^{2Y}}{y} + \frac{e^Y}{y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}.\end{aligned}$$

■

In the following proposition, we pass from the smooth result of Lemma 6.9 to asymptotic counts of geodesics with a sharp cutoff in one of the length and holonomy parameters, retaining a smooth cutoff in the other. We pass to the sharp count using Corollary 6.10, which involves finding an upper bound in the two rectangular regions of ambiguity with at least one short side length, and thus improves on the error term compared with Theorem 6.2 and Proposition 6.4 (see Remark 6.13 below).

Proposition 6.11. Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup, and let ϖ_Γ be the density (2). Then, for any intervals $I = [y', y]$ $(0 \leqslant y' < y)$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$:

(a) For every $0 < \eta' \leqslant 2\pi$ and for $f_{J,\eta'}$ as in (57),

$$\begin{aligned}\pi_\Gamma(I, f_{J,\eta'}) &:= \sum_{\ell(\gamma)\in I}^{P} f_{J,\eta'}(\mathrm{hol}(\gamma)) \\ &= \frac{|J|}{2\pi}\int_I \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_\Gamma\Big((|J|+\eta')^{2/3}\frac{e^{5y/3}}{y}\Big(\log^*\frac{1}{\eta'}\Big)^{1/3} + \frac{e^Y}{\eta'^2}\Big).\end{aligned}$$

(b) For every $0 < \eta \leqslant \eta_0$ and for $g_{I,\eta}$ as in (62),

$$\begin{aligned}\pi_\Gamma(g_{I,\eta}, J) &:= \sum_{\mathrm{hol}(\gamma)\in J}^{P} g_{I,\eta}(\ell(\gamma)) \\ &= \frac{|J|}{2\pi}\int_2^\infty g_{I,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big((|I|+\eta)^{2/3}\frac{e^{5y/3}}{y^{2/3}} + \frac{e^Y}{\eta^2}\Big).\end{aligned}$$

Proof. We approximate $\pi_\Gamma(I, f_{J,\eta'})$ using the smooth cutoff count $\pi_\Gamma(g_{I,\eta}, f_{J,\eta'})$, with $0 < \eta \leqslant \eta_0$ to be suitably chosen later. These counts differ at most by the number of primitive classes $[\gamma]$ with $(\ell(\gamma), \mathrm{hol}(\gamma)) \in (I_\eta^- \cup I_\eta^+) \times J_{\eta'}$. By Corollary 6.10, we have the

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

bound

$$\begin{aligned}|\pi_\Gamma(I,f_{J,\eta'}) - \pi_\Gamma(g_{I,\eta},f_{J,\eta'})| &\leqslant \pi_\Gamma(I_\eta^-,J_{\eta'}) + \pi_\Gamma(I_\eta^+,J_{\eta'}) \\ &\ll_{\Gamma,\eta_0} \eta(|J|+\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}.\end{aligned} \tag{63}$$

Combining the estimate of Lemma 6.9 for $\pi_\Gamma(g_{I,\eta},f_{J,\eta'})$ with the estimate (63) on the region of ambiguity, we have that

$$\begin{aligned}\pi_\Gamma(I,f_{J,\eta'}) &= \frac{|J|}{2\pi}\int_2^\infty g_{I,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\eta(|J|+\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big) \\ &= \frac{|J|}{2\pi}\int_I \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big(\eta(|J|+\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big).\end{aligned}$$

Taking, say, $\eta_0 = 1$, the essentially optimal choice $\eta = \min\left(e^{-Y/3}(\log^*\frac{1}{\eta'})^{1/3}/(|J|+\eta')^{1/3}, \eta_0\right)$ completes the proof of (a).

The proof of (b) is similar: approximating $\pi_\Gamma(g_{I,\eta},J)$ with the smooth count $\pi_\Gamma(g_{I,\eta},f_{J,\eta'})$, with $0 < \eta' \leqslant 2\pi$ to be suitably chosen later, we have that $|\pi_\Gamma(g_{I,\eta},J) - \pi_\Gamma(g_{I,\eta},f_{J,\eta'})| \leqslant \pi_\Gamma(I_\eta,J_{\eta'}^-) + \pi_\Gamma(I_\eta,J_{\eta'}^+)$, so that combining Corollary 6.10 and Lemma 6.9 we conclude

$$\pi_\Gamma(g_{I,\eta},J) = \frac{|J|}{2\pi}\int_2^\infty g_{I,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big((|I|+\eta)\eta'\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big).$$

Choosing $\eta' = \min\left(Y^{1/3}e^{-Y/3}/(|I|+\eta)^{1/3}, 2\pi\right)$ completes the proof. ■

Finally, we present our main theorem, which provides a count of length and holonomy in intervals I and J, respectively. When $I = [0,y]$ and J is fixed, this recovers Theorem 6.5. However, when the lengths of I and J are shrinking, we have a significant improvement.

Theorem 6.12 (Ambient Short-Range Prime Geodesic Theorem). Let $\Gamma < \mathrm{PSL}_2\mathbb{C}$ be a discrete, co-compact, torsion-free subgroup. Then, for any intervals $I = [y',y]$ $(0 \leqslant y' < y)$ and $J \subseteq \mathbb{R}/2\pi\mathbb{Z}$,

$$\begin{aligned}\pi_\Gamma(I,J) &:= \left|\left\{[\gamma]^P : (\ell(\gamma),\mathrm{hol}(\gamma)) \in I\times J\right\}\right| \\ &= \iint_{I\times J} \mathrm{d}\varpi_\Gamma(u)\frac{\mathrm{d}\theta}{2\pi} + \mathrm{O}_\Gamma\Big((|I|+|J|)^{2/3}\frac{e^{5y/3}}{y^{2/3}} + \frac{e^{3y/2}}{y^{1/2}}\Big),\end{aligned}$$

where ϖ_Γ is the density (2).

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Proof. We use the smooth count $\pi_\Gamma(g_{I,\eta}, f_{J,\eta'})$, with parameters $0 < \eta \leqslant \eta_0$ and $0 < \eta' \leqslant 2\pi$ to be suitably chosen later, to approximate the sharp cutoff count $\pi_\Gamma(I,J)$. Using Corollary 6.10 to estimate the boundary terms, as in the proof of Proposition 6.11, we obtain

$$\begin{aligned}|\pi_\Gamma(I,J) - \pi_\Gamma(g_{I,\eta}, f_{J,\eta'})| &\leqslant \pi_\Gamma(I_\eta^-, J_{\eta'}) + \pi_\Gamma(I_\eta^+, J_{\eta'}) + \pi_\Gamma(I_\eta, J_{\eta'}^-) + \pi_\Gamma(I_\eta, J_{\eta'}^+)\\ &\ll_\Gamma (\eta|J| + \eta'|I| + \eta\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}.\end{aligned}$$

Then, using Lemma 6.9 for $\pi_\Gamma(g_{I,\eta}, f_{J,\eta'})$, we achieve the estimate

$$\begin{aligned}\pi_\Gamma(I,J) &= \frac{|J|}{2\pi}\int_2^\infty g_{I,\eta}(u)\,\mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big((\eta|J| + \eta'|I| + \eta\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big)\\ &= \frac{|J|}{2\pi}\int_I \mathrm{d}\varpi_\Gamma(u) + \mathrm{O}_{\Gamma,\eta_0}\Big((\eta|J| + \eta'|I| + \eta\eta')\frac{e^{2Y}}{Y} + \frac{e^Y}{Y\eta^2}\log^*\frac{1}{\eta'} + \frac{e^Y}{\eta'^2}\Big). \qquad (64)\end{aligned}$$

To obtain the result, we use the essentially optimal choices

$$\eta = \min\Big(\frac{Y^{1/3}e^{-Y/3}}{|J|^{1/3}}, Y^{1/4}e^{-Y/4}, \eta_0\Big), \quad \eta' = \min\Big(\frac{Y^{1/3}e^{-Y/3}}{|I|^{1/3}}, Y^{1/4}e^{-Y/4}, 2\pi\Big),$$

taking a fixed $\eta_0 = 1$. Here we note that a brief comparison of the three latter summands in the error term in (64) shows that indeed the term $e^{3Y/2}/Y^{1/2}$ in Theorem 6.12 is the best possible following (64). ∎

Remark 6.13. As already remarked, Theorem 6.12 recovers the long-range Theorem 6.5 and its Corollary 6.7. It provides a substantial improvement as soon as both I and J are short, which is particularly strong if the total boundary length $\asymp |I| + |J|$ is favorably small. For example, if $|I| \asymp |J|$, Theorem 6.12 gives a power-saving asymptotic as long as $|I| \asymp |J| \gg e^{(-1/4+\delta)Y}$, with the error term potentially as small as $\mathrm{O}_{\Gamma,\delta}(e^{(3/2+\delta)Y})$. In general, Theorem 6.12 yields a power-saving asymptotic whenever $|I \times J|^3/(|I| + |J|)^2 \asymp |I \times J| \cdot \min(|I|^2, |J|^2) \gg e^{(-1+\delta)Y}$ for some $\delta > 0$.

Proposition 6.11 similarly improves upon Theorem 6.2 and Proposition 6.4. For example, when sampling geodesics with a sharp cutoff in length in an interval of size $|I| \asymp 1$ and a "mild" holonomy cutoff with $\eta' \asymp |J|$, Proposition 6.11 improves upon Theorem 6.2 for all $e^{(-1/3+\delta)Y} \leqslant |J| \leqslant e^{-\delta Y}$, with the error term as good as $\mathrm{O}_\Gamma(e^{3Y/2})$ for $|I| \asymp 1$ and $\eta', |J| \asymp e^{-Y/4}$.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022

Funding

This work was supported by the National Science Foundation [DMS-1903301 to D. M.].

Acknowledgments

We are very grateful to the anonymous referees for their careful reading and thoughtful comments, which have significantly improved the manuscript, including a crucial observation that led to the improvement in our ambient counts over short ranges. D. M. would like to thank the Max Planck Institute for Mathematics for their support and outstanding working conditions and research atmosphere.

References

[1] Balkanova, O., D. Chatzakos, G. Cherubini, D. Frolenkov, and N. Laaksonen. "Prime geodesic theorem in the 3-dimensional hyperbolic space." *Trans. Amer. Math. Soc.* 372 (2019): 5355–74 MR: 4014279.

[2] Balkanova, O. and D. Frolenkov. "Prime geodesic theorem for the Picard manifold." *Adv. Math.* 375 (2020): 42 MR: 4142089.

[3] Balog, A., A. Biró, G. Cherubini, and N. Laaksonen. "Bykovskii-type theorem for the Picard manifold." *Internat. Math. Res. Notices* (2020): rnaa128.

[4] Brumley, F. and D. Milićević. "Counting cusp forms by analytic conductor." *arXiv e-prints* (2018) arXiv:1805.00633.

[5] Chinburg, T., E. Hamilton, D. D. Long, and A. W. Reid. "Geodesics and commensurability classes of arithmetic hyperbolic 3-manifolds." *Duke Math. J.* 145 (2008): 25–44 MR: 2451288.

[6] Deitmar, A. "A prime geodesic theorem for higher rank spaces." *Geom. Funct. Anal.* 14 (2004): 1238–66 MR: 2135166.

[7] Dever, L. *Non-Spherical Maass Forms on Hyperbolic 3-Manifolds*. Master's thesis. Bryn Mawr College, 2019.

[8] Elstrodt, J., F. Grunewald, and J. Mennicke. "Groups acting on hyperbolic space." *Monographs in Mathematics*. Berlin: Springer, 1998. Harmonic analysis and number theory. MR: 1483315.

[9] Gangolli, R. "The length spectra of some compact manifolds of negative curvature." *J. Differential Geometry* 12 (1977): 403–24 MR: 650997.

[10] Gangolli, R. and G. Warner. "Zeta functions of Selberg's type for some noncompact quotients of symmetric spaces of rank one." *Nagoya Math. J.* 78 (1980): 1–44 MR: 571435.

[11] Garoufalidis, S. and A. W. Reid. "Constructing 1-cusped isospectral non-isometric hyperbolic 3-manifolds." *J. Topol. Anal.* 10 (2018): 1–25 MR: 3737506.

[12] Iwaniec, H. "Introduction to the spectral theory of automorphic forms." *Revista Matemática Iberoamericana*. Madrid: Biblioteca de la Revista Matemática Iberoamericana. [Library of the Revista Matemática Iberoamericana], 1995 MR: 1325466.

[13] Kelmer, D. "Distribution of holonomy about closed geodesics in a product of hyperbolic planes." *Amer. J. Math.* 134 (2012): 1613–53 MR: 2999290.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/rnab048/6381506 by Bryn Mawr College user on 13 June 2022

[14] Kelmer, D. "A refinement of strong multiplicity one for spectra of hyperbolic manifolds." *Trans. Amer. Math. Soc.* 366 (2014): 5925–61 MR: 3256189.

[15] Knapp, A. W. "Representation theory of semisimple groups." *Princeton Landmarks in Mathematics*. Princeton, NJ: Princeton University Press, 2001 An overview based on examples, Reprint of the 1986 original. MR: 1880691.

[16] Kowalski, E. "Families of cusp forms." *Actes de la Conférence "Théorie des Nombres et Applications", Publ. Math. Besançon Algèbre Théorie des Nombres* 5–40. Franche-Comté, Besançon: Presses Univ, 2013 MR: 3220018.

[17] Lin, F. and M. Lipnowski., *The Seiberg-Witten equations and the length spectrum of hyperbolic three-manifolds*, *arXiv e-prints* (2018), arXiv:1810.06346.

[18] Lokvenec-Guleska, H. *Sum formula for SL_2 over imaginary quadratic number fields*. PhD thesis. Universiteit Utrecht, 2004.

[19] Margulis, G. "Certain applications of ergodic theory to the investigation of manifolds of negative curvature." *Funkcional. Anal. i Priložen.* 3 (1969): 89–90 MR: 0257933.

[20] Margulis, G., A. Mohammadi, and H. Oh. "Closed geodesics and holonomies for Kleinian manifolds." *Geom. Funct. Anal.* 24 (2014): 1608–36 MR: 3261636.

[21] Margulis, G., A. Mohammadi, and H. Oh. "Closed geodesics and holonomies for Kleinian manifolds." *arXiv e-prints* (2014) arXiv:1404.5698.

[22] Reid, A. W. "Isospectrality and commensurability of arithmetic hyperbolic 2- and 3-manifolds." *Duke Math. J.* 65 (1992): 215–28 MR: 1150584.

[23] Sarnak, P. "The arithmetic and geometry of some hyperbolic three-manifolds." *Acta Math.* 151 (1983): 253–95 MR: 723012.

[24] Sarnak, P., S. W. Shin, and N. Templier. "Families of L-functions and their symmetry." *Families of Automorphic Forms and the Trace Formula*. Simons Symp. 531–78. Cham: Springer, 2016 MR: 3675175.

[25] Sarnak, P. and M. Wakayama. "Equidistribution of holonomy about closed geodesics." *Duke Math. J.* 100 (1999): 1–57 MR: 1714754.

[26] Soundararajan, K. and M. P. Young. "The prime geodesic theorem." *J. Reine Angew. Math.* 676 (2013): 105–20 MR: 3028757.

[27] Sunada, T. "Riemannian coverings and isospectral manifolds." *Ann. Math.* 121 (1985): 169–86 MR: 782558.

[28] Vignéras, M.-F. "Variétés riemanniennes isospectrales et non isométriques." *Ann. Math.* 112 (1980): 21–32 MR: 584073.

Downloaded from https://academic.oup.com/imrn/advance-article/doi/10.1093/imrn/mab048/6381506 by Bryn Mawr College user on 13 June 2022