EXTREMALITY OF SUBMANIFOLDS OF SYSTEMS OF LINEAR FORMS VIA HEIGHT FUNCTIONS

OSAMA KHALIL

ABSTRACT. Suppose $M_{m,n}$ is the space of real matrices with m rows and n columns and $\mathbf{d} : M_{m,n} \to \mathbb{R}^N$ is the map that assigns to each matrix a tuple of its minors in some fixed order for a suitable N. For an interval $B \subset \mathbb{R}$, a smooth map $\varphi : B \to M_{m,n}$ is strongly non-planar if the derivatives of $\mathbf{d} \circ \varphi$ up to order N span $\mathbb{R}^N$ at Lebesgue almost every point in B. Kleinbock, Margulis and Wang showed that Lebesgue almost every $s \in B$ is not very well approximable, generalizing earlier work of Kleinbock and Margulis in the case $\min(m, n) = 1$. In this article, we provide a proof of this result using systems of integral inequalities introduced in the work of Eskin, Margulis and Mozes. The proof produces a new quantitative non-divergence result for expanding translates of shrinking curves on $\mathrm{SL}(m+n, \mathbb{R})/\mathrm{SL}(m+n, \mathbb{Z})$ by general diagonal elements. One feature of this approach is that it relies on the simpler (C, α)-good property for polynomials.

1. INTRODUCTION

1.1. **Extremality and Historical Context.** A real matrix $Y \in M_{m,n}$ with m rows and n columns is very well approximable (VWA) if there exists $\varepsilon > 0$ and infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$\|Y\mathbf{q} - \mathbf{p}\| < \|\mathbf{q}\|^{-n/m-\varepsilon} \text{ for some } \mathbf{p} \in \mathbb{Z}^m \tag{1.1}$$

where for $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, $\|\mathbf{x}\| = \max_{1 \leqslant i \leqslant m} |x_i|$. We say that Y is very well multiplicatively approximable (VWMA) if there exists $\varepsilon > 0$ and infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$\Pi\left(Y\mathbf{q} - \mathbf{p}\right) < \Pi_+\left(\mathbf{q}\right)^{-1-\varepsilon} \text{ for some } \mathbf{p} \in \mathbb{Z}^m \tag{1.2}$$

where for $\mathbf{x} = (x_i)$,

$$\Pi(\mathbf{x}) = \prod_i |x_i|, \qquad \Pi_+(\mathbf{x}) = \prod_i \max\left\{|x_i|, 1\right\} \tag{1.3}$$

We note that if Y is VWA, then it is VWMA.

By Khinchine's transference principal one has that a matrix Y is VWA if and only if its transpose is. It is a classical fact that the set of VWA matrices in $M_{m,n}$ has Lebesgue measure 0. A similar argument shows that the same holds for VWMA matrices. A more delicate problem is one of determining sufficient conditions on submanifolds of $M_{m,n}$ so that almost every point is not VWA (or VWMA) with respect to the induced Lebesgue volume measure. Submanifolds for which almost every point is not VWA (resp. VWMA) are called *extremal* (resp. *strongly extremal*).

In 1932, Mahler asked whether the curve $(x, x^2, \ldots, x^n)$ is extremal. This conjecture was settled by Sprindžuk in the sixties. This led him to formulate the following conjecture. Suppose $f_1, \ldots, f_n : U \to \mathbb{R}$ are real analytic maps on an open set $U \subseteq \mathbb{R}^d$ which, along with 1, are linearly independent over $\mathbb{R}$. Then, the manifold $\mathbf{f}(U)$ is an extremal submanifold of $\mathbb{R}^n$, where $\mathbf{f} = (f_1, \ldots, f_n)$.

This conjecture was settled in the groundbreaking work of Kleinbock and Margulis in [KM98]. In fact, they prove a much stronger form of this conjecture concerning strong extremality for *non-degenerate* manifolds. A smooth map $\mathbf{f} : U \to \mathbb{R}^n$ is ℓ-*non-degenerate* at $\mathbf{x} \in U$ if the derivatives of $\mathbf{f}$ at $\mathbf{x}$ up to order ℓ span $\mathbb{R}^n$. We say $\mathbf{f}$ is non-degenerate at $\mathbf{x}$ if it is ℓ-non-degenerate for some ℓ and non-degenerate (on U) if it is non-degenerate at almost every $\mathbf{x} \in U$ with respect to the Lebesgue measure. These results treat the case of submanifolds of $M_{1,n} \cong \mathbb{R}^n$.

The work of Kleinbock and Margulis has been generalized in many directions. In [KLW04], sufficient conditions on a large class of measures on $\mathbb{R}^n$ were found to guarantee that they assign 0 mass to the set of VWMA vectors. This class (referred to as friendly measures) includes measures of the form $\mathbf{f}_*\lambda$ where λ is a Lebesgue measure on $U \subseteq \mathbb{R}^d$ and $\mathbf{f} : U \to \mathbb{R}^n$ is a non-degenerate smooth map, in particular generalizing the results of [KM98]. It also includes a wide class of fractal measures arising from iterated function systems and satisfying the open set condition.

Subsequently these results on friendly measures on $\mathbb{R}^n \cong M_{1,n}$ were extended to the space of systems of linear forms $M_{m,n}$ in [KMW10]. We say a measure μ on $M_{m,n}$ is *strongly extremal* if it assigns 0 mass to the set of VWMA matrices and *extremal* if the set of VWA matrices is null with respect to μ. A notable class of strongly extremal measures found in [KMW10] are measures of the form $\mu = \mathbf{f}_*\lambda$ on $M_{m,n}$ where λ is a Lebesgue measure on a domain $U \subseteq \mathbb{R}^d$ and $\mathbf{f} : U \to M_{m,n}$ is a smooth map which is strongly non-planar.

To define the property of strong non-planarity, let $\mathbf{d} : M_{m,n} \to \mathbb{R}^N$ be the map that assigns to each matrix the tuple of the determinants of all of its minors given in some prefixed order, where $N = \binom{m+n}{n} - 1$. We will refer to $\mathbf{d}$ as the minors map. A smooth map $\mathbf{f} : U \to M_{m,n}$ is *strongly non-planar* if $\mathbf{d} \circ \mathbf{f} : U \to \mathbb{R}^N$ is non-degenerate on U.

1.2. **Reduction to Dynamics on the Space of Lattices.** An approach to these number theoretic problems via dynamics on the space of unimodular lattices $\mathrm{SL}(m+n,\mathbb{R})/\mathrm{SL}(m+n,\mathbb{Z})$ was innovated in [KM98]. This approach has also been used in the subsequent generalizations mentioned above. We recall this reduction here.

Let $G = \mathrm{SL}(m+n,\mathbb{R})$ and $\Gamma = \mathrm{SL}(m+n,\mathbb{Z})$ for some $m, n \in \mathbb{N}$. Denote by $\mathcal{A} \subset \mathbb{R}^{m+n}$ the set of tuples $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_{m+n})$ satisfying

$$0 < \mathbf{r}_1, \dots, \mathbf{r}_{m+n}, \qquad \sum_{i=1}^{m} \mathbf{r}_i = 1 = \sum_{j=1}^{n} \mathbf{r}_j \tag{1.4}$$

We shall refer to elements of $\mathcal{A}$ as weights. Denote by $\mathcal{T}$ the set of all dilations of $\mathcal{A}$. More precisely, $\mathcal{T}$ consists of the set of tuples $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_{m+n}) \in \mathbb{R}^{m+n}$ satisfying $\mathbf{t} = t\mathbf{r}$ for some $t > 0$ and $\mathbf{r} \in \mathcal{A}$. In particular, for any $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_{m+n}) \in \mathcal{T}$, we use $|\mathbf{t}| > 0$ to denote

$$|\mathbf{t}| = \sum_{i=1}^{m} \mathbf{t}_i = \sum_{j=1}^{n} \mathbf{t}_{m+j}$$

The set $\mathcal{R} = \{(t/m, \dots, t/m, t/n, \dots, t/n) \in \mathcal{T} : t > 0\}$ will also be of interest to us. For $Y \in M_{m,n}$ and $\mathbf{t} \in \mathcal{T}$, define

$$u(Y) = \begin{pmatrix} \mathrm{I}_m & Y \\ \mathbf{0} & I_n \end{pmatrix}, \qquad g_{\mathbf{t}} = \mathrm{diag}(e^{\mathbf{t}_1}, \dots, e^{\mathbf{t}_m}, e^{-\mathbf{t}_{m+1}}, \dots, -e^{\mathbf{t}_{m+n}}) \tag{1.5}$$

where I_d denotes the identity matrix in dimensions d. In particular, these matrices belong to G.

It was observed in [KM98] that the different notions of well approximability of $Y \in M_{m,n}$ are related to deep excursions of $g_{\mathbf{t}}u(Y)\Gamma$ into the cusp of G/Γ. To make this notion precise, recall that G/Γ can be identified with the space of unimodular lattices in $\mathbb{R}^{m+n}$ via the map $g\Gamma \mapsto g\mathbb{Z}^{m+n}$. Thus, we may think of elements of G/Γ as lattices in $\mathbb{R}^{m+n}$. For a lattice $x \in G/\Gamma$, we say a subgroup $\Lambda \subset x$ is *primitive* if $\mathbb{R}\Lambda \cap x = \Lambda$, where $\mathbb{R}\lambda$ denotes the $\mathbb{R}$-span of Λ. Denote by $d(\Lambda)$ the volume of the torus $\mathbb{R}\Lambda/\Lambda$, where if $\Lambda = \{0\}$, we take $d(\Lambda) = 1$.

For $1 \leqslant k < m+n$, define the following functions on G/Γ.

$$\begin{aligned} \tilde{\alpha}_k(x) &= \sup\left\{ \frac{1}{d(\Lambda)} : \Lambda \text{ is a subgroup of } x \text{ of rank } k \right\} \\ \tilde{\alpha}(x) &= \max_{1 \leqslant k < m+n} \tilde{\alpha}_k(x) \end{aligned} \tag{1.6}$$

By Mahler's compactness criterion, $\tilde{\alpha}$ is a proper function on G/Γ and is thus well-equipped to detect cusp excursions.

Given a subset $\mathcal{C} \subseteq \mathcal{T}$ and $x_0 \in G/\Gamma$, we say $g_{\mathcal{C}}x_0$ has *linear growth* if there exists $\gamma > 0$ such that

$$\tilde{\alpha}_1(g_{\mathbf{t}}x_0) \geqslant e^{\gamma|\mathbf{t}|} \text{ for an unbounded set of } \mathbf{t} \in \mathcal{C} \tag{1.7}$$

This terminology is used in [KMW10]. The following proposition provides the link between extremality and dynamics on the space of lattices.

Proposition 1.1 (Proposition 3.1, [KMW10])**.** *Suppose $Y \in M_{m,n}$. Then,*

(1) Y is VWA $\Leftrightarrow g_{\mathcal{R}}u(Y)\Gamma$ has linear growth.
(2) Y is VWMA $\Leftrightarrow g_{\mathcal{T}}u(Y)\Gamma$ has linear growth.

With this dynamical interpretation of extremality in place, we can state the first main result of this article. The following three functions on the space of weights $\mathcal{A}$ will be convenient for us.

$$\begin{aligned}
\mu(\mathbf{r}) &= \min\{\mathbf{r}_i + \mathbf{r}_{m+j} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\} \\
\nu(\mathbf{r}) &= \max\{\mathbf{r}_i + \mathbf{r}_{m+j} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\} \\
D(\mathbf{r}) &= \left\lceil \frac{\nu(\mathbf{r})}{\mu(\mathbf{r})} N - 1 \right\rceil \\
\lambda(\mathbf{r}) &= \min\{\mathbf{r}_i : 1 \leqslant i \leqslant m+n\}
\end{aligned} \tag{1.8}$$

Recall that $N = \binom{m+n}{n} - 1$ These functions allow us to define the following type of unbounded subsets of $\mathcal{T}$ which we shall be interested in.

Definition 1.2. A subset $\mathcal{C} \subseteq \mathcal{T}$ is said to be *completely expanding* if

$$\inf\{\lambda(\mathbf{t}/|\mathbf{t}|) : \mathbf{t} \in \mathcal{C}\} > 0$$

The *degree* of $\mathcal{C}$, denoted by $D(\mathcal{C})$, is defined as follows.

$$D(\mathcal{C}) = \sup\{D(\mathbf{t}/|\mathbf{t}|) : \mathbf{t} \in \mathcal{C}\}$$

The following is the first main result of this article. We emphasize that this result is a special case of the main results obtained previously in [KM98] and [KMW10].

Theorem 1.3. *Suppose $\mathcal{C} \subseteq \mathcal{T}$ is a completely expanding unbounded subset of $\mathcal{T}$ and suppose $\varphi : B \to M_{m,n}$ is a $(D(\mathcal{C})+1)$-times continuously differentiable curve which is strongly non-planar on B. Then, for every $x_0 \in G/\Gamma$, the set of $s \in B$ such that $g_{\mathcal{C}}u(\varphi(s))x_0$ has linear growth has Lebesgue measure* 0.

Let us record a few remarks on our condition on $\mathcal{C}$ being completely expanding.

Remark 1.4. For any fixed weight $\mathbf{r} \in \mathcal{A}$, a positive ray $t\mathbf{r}$ is completely expanding. In particular, Theorem 1.3 shows that $\varphi_*\lambda(\text{VWA}) = 0$ where λ is the Lebesgue measure on λ.

1.3. **Improving Dirichlet's Theorem.** Another notion in the theory of diophantine approximation of systems is motivated by Dirichlet's theorem in this setting. Given an unbounded subset $\mathcal{C} \subseteq \mathcal{T}$ and $0 < \varepsilon < 1$, we say Dirichlet's theorem cannot be improved for a matrix $Y \in M_{m,n}$ along $\mathcal{C}$ if for all $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_{m+n}) \in \mathcal{C}$ sufficiently large, there exist $\mathbf{p} = (p_1, \dots, p_m) \in \mathbb{Z}^m$ and $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{Z}^n$ so that the following inequalities are satisfied

$$\begin{aligned}
|Y_i \cdot q - p_i| &< \varepsilon e^{-\mathbf{t}_i} \qquad i = 1, \dots, m \\
|q_j| &< \varepsilon e^{\mathbf{t}_{m+j}} \qquad j = 1, \dots, n
\end{aligned} \tag{1.9}$$

where Y_i is the i^{th} row of Y. We denote by $\mathrm{DI}_\varepsilon(\mathcal{C})$ the set of matrices Y for which Dirichlet's theorem can be ε-improved along $\mathcal{C}$. The reason for the terminology is that Dirichlet's classical theorem states that the inequalities (1.9) are always satisfied with $\varepsilon = 1$ and $\mathcal{C} = \mathcal{R}$.

Davenport and Schmidt showed that $\mathrm{DI}_\varepsilon(\mathcal{R})$ has Lebesgue measure 0 for any $\varepsilon < 1$. They also showed that for any $\varepsilon < 4^{-1/3}$, the set of $x \in \mathbb{R}$ such that $(x, x^2) \in \mathrm{DI}_\varepsilon(\mathcal{R}) \subset M_{1,2}$ has Lebesgue measure 0. Using dynamics on the space of lattices, these results were extended by Kleinbock and Weiss in [KW08] where they showed that for any $\varepsilon < 1$, the set $\mathrm{DI}_\varepsilon(\mathcal{C})$ has measure 0 whenever $\mathcal{C}$ is drifting away from the walls, a technical condition similar to being completely expanding (see [KW08, Eq. 1.8]). When $\min(m,n) = 1$, they showed that for any non-degenerate smooth map $\mathbf{f} : U \to \mathbb{R}^n \cong M_{1,n}$, there exists $0 < \varepsilon_0 < 1$ such that $\mathbf{f}_*\lambda(\mathrm{DI}_\varepsilon(\mathcal{T})) = 0$ for any $\varepsilon < \varepsilon_0$, where λ is the Lebesgue measure on $U \subset \mathbb{R}^k$ for some k. In fact, they show that the conclusion holds for a more general class of measures, see Theorem 1.5 in [KW08] for the details.

In [Sha09, Sha10], using Ratner's measure classification theorem and the linearization technique, Shah extended the results of Kleinbock and Weiss for *analytic* non-degenerate maps $\mathbf{f}$ as above in the case $\min(m,n) = 1$ to show that $\mathbf{f}_*\lambda(\mathrm{DI}_\varepsilon(\mathcal{T})) = 0$ for any $\varepsilon < 1$.

The second result of this article is a generalization of the results of Kleinbock and Weiss for smooth strongly non-planar curves in $M_{m,n}$. We emphasize that in the case $\min(m,n) = 1$, Theorem 1.5 below is a special case of the results of Kleinbock and Weiss and for general m and n, it is a special case of a result which was announced in [KMW10, Theorem 8.1].

Theorem 1.5. *Suppose $\mathcal{C} \subseteq \mathcal{T}$ is a completely expanding unbounded subset of $\mathcal{T}$ and suppose $\varphi : B \to M_{m,n}$ is a $(D(\mathcal{C})+1)$-times continuously differentiable curve which is strongly non-planar on B. Then, there exists $\varepsilon_0 > 0$ depending on $\mathcal{C}$ and φ such that for any $\varepsilon < \varepsilon_0$, $\varphi_*\lambda(\mathrm{DI}_\varepsilon(\mathcal{C})) = 0$, where λ is the Lebesgue measure on B.*

We remark that the constant ε_0 in Theorem 1.5 is explicitly computable yet it is reasonable to expect that the result holds with $\varepsilon_0 = 1$.

1.4. **Quantitative Non-divergence of Shrinking Curves.** The key technical result we use to establish Theorems 1.3 and 1.5 is a form of quantitative non-divergence of certain measures on the homogeneous space G/Γ. Specifically, we show, in a quantitative form, that any limit point of the push-forward of parameter measures on shrinking pieces of a curve φ as in Theorems 1.3 and 1.5 by diagonal elements as in (1.5) is a probability measure on G/Γ. This result is new as it does not follow from the previous results of [KM98, KMW10, KW08]. The following is the precise statement.

Theorem 1.6. *Suppose $\mathcal{C} \subseteq \mathcal{T}$ is a completely expanding unbounded subset of $\mathcal{T}$ and suppose $\varphi : B \to M_{m,n}$ is a $(D(\mathcal{C})+1)$-times continuously differentiable curve such that $\mathbf{d} \circ \varphi$ is non-degenerate at some $s_0 \in B$. Then, there exists a constant for $\kappa = \kappa(\mathcal{C}) > 0$ such that for any constants $0 \leqslant \delta < \beta < \kappa$ and any $x_0 \in G/\Gamma$,*

$$\sup_{\mathbf{t} \in \mathcal{C} : J_{\mathbf{t}} \subseteq B} \frac{1}{|J_{\mathbf{t}}|} \int_{J_{\mathbf{t}}} \tilde{\alpha}^{\beta}(g_{\mathbf{t}} u(\varphi(s)) x_0)\, ds < \infty$$

where $J_{\mathbf{t}} = \left[s_0 - e^{-\delta|\mathbf{t}|}, s_0 + e^{-\delta|\mathbf{t}|}\right]$. Moreover, the supremum can be taken to be uniform as the base point x_0 varies in a fixed compact set.

Remark 1.7. The exponent κ in Theorem 1.6 can be calculated explicitly as follows:

$$\kappa(\mathcal{C}) = \inf_{\mathbf{t} \in \mathcal{C}} \min\left\{\delta(\mathbf{t}/|\mathbf{t}|), \alpha(\mathbf{t}/|\mathbf{t}|)\right\} \tag{1.10}$$

where the functions δ and α on the space of weights are defined as follows.

$$\begin{aligned} Q(\mathbf{r}) &:= D(\mathbf{r}) \min(m,n), \qquad & C(\mathbf{r}) &:= 2Q(\mathbf{r})\left(Q(\mathbf{r})+1\right)^{1/Q(\mathbf{r})} \\ \alpha(\mathbf{r}) &:= 1/Q(\mathbf{r}), & \delta(\mathbf{r}) &:= \mu(\mathbf{r})/N \end{aligned} \tag{1.11}$$

and $D(\mathbf{r})$ and $\mu(\mathbf{r})$ were defined in (1.8).

The proof of Theorem 1.6 is given in Section 7.

2. Overview of the Proofs

2.1. **The Approach of Kleinbock and Margulis.** Using Proposition 1.1, the idea of proving strong extremality of a (compactly supported) measure μ on $M_{m,n}$ was reduced in [KM98] and [KMW10] to showing measure estimates of the form

$$\mu\left(\{Y \in M_{m,n} : \tilde{\alpha}_1(g_{\mathbf{t}}u(Y)\Gamma) > A\}\right) \leqslant \text{const. } A^{-\alpha}\mu(M_{m,n}) \tag{2.1}$$

for some $\alpha > 0$ and applying the Borel-Cantelli lemma. Estimates of the form (2.1) are often referred to as quantitative non-divergence. Throughout the rest of the article, we will focus on measures μ arising as a pushforward of Lebesgue measure on bounded intervals B of the real line under smooth maps $\mathbf{f} : B \to M_{m,n}$.

The original approach of [KM98] to establish estimates of the form (2.1) has two elements which we recall here. For the purpose of this discussion, we will restrict to the case $m = 1$. The scheme is very similar in the general case.

The first element is a certain growth property for smooth maps known as the $(C, \alpha)-$good property. A continuous function $f : U \to \mathbb{R}$ is said to be $(C, \alpha)-$good on a bounded set $U \subset \mathbb{R}^k$ if there exist positive constants C and α so that for all $\varepsilon > 0$, the following holds

$$\lambda\left(\{x \in U : |f(x)| < \varepsilon\}\right) \leqslant C\left(\frac{\varepsilon}{\sup_{x\in U}|f(x)|}\right)^{\alpha}\lambda(U) \tag{2.2}$$

where λ is the Lebesgue measure on $\mathbb{R}^k$.

The simplest class of $(C, \alpha)-$good functions on the real line is given by polynomials. This was already observed in earlier work of Dani and Margulis [DM93, Lemma 4.1] and further quantified in [KM98, Proposition 3.2]. In [KM98], the $(C, \alpha)-$good property was established for a wider class of smooth maps. The functions which were shown to satisfy this property arose as the coordinate functions (with respect to the standard basis) of the maps $\mathbf{x} \mapsto g_{\mathbf{t}}u(\mathbf{f}(\mathbf{x}))v$, for certain vectors $v \in \bigwedge^k \mathbb{R}^{n+1}$, $1 \leq k \leq n+1$. When $\min(m, n) = 1$, it was shown in [KM98] that if $\mathbf{f}$ is non-degenerate, then these coordinate functions are $(C, \alpha)-$good. The proof of this fact is substantially more involved in this generality than the case of polynomials.

The approach in this article bypasses this step and relies on the $(C, \alpha)-$good property of polynomials only. Roughly, the idea behind this simplification is that, locally, the behavior of a general smooth non-degenerate curve $\mathbf{f}$ is fully encoded in the behavior of a suitable Taylor-type polynomial approximations of the curve. Hence, understanding the excursions of these polynomial approximations in the cusp of G/Γ is sufficient to understand the evolution of the original curve.

The $(C, \alpha)-$good property is used in [KM98] to establish the measure estimate in (2.1) for functions of the form $f(\mathbf{x}) = \|g_{\mathbf{t}}u(\mathbf{f}(\mathbf{x})v\|^{-1}$ for any vector $v \in \bigwedge^k \mathbb{R}^{n+1}$, $1 \leq k \leq n+1$ in place of the function $\tilde{\alpha}_1$. Recall that one may identify any discrete subgroup Λ of $\mathbb{R}^n$ of rank k with a vector $v_1 \wedge \cdots \wedge v_k \in \bigwedge^k \mathbb{R}^{n+1}$, where $v_1, \ldots, v_k$ is a basis for Λ as a $\mathbb{Z}$ module.

The second element of the argument in the proof of Kleinbock and Margulis converts these estimates into a measure estimate for $\tilde{\alpha}_1$. The proof in [KM98] begins by associating to each lattice $x \in G/\Gamma$ a set $P(x)$ of its (primitive) subgroups of all ranks, partially ordered by inclusion. Roughly speaking, a combination of an inductive argument on the length of totally ordered subsets of $P(x)$ and a covering argument using Besicovitch's covering theorem allows them to show that "maximal bad intervals" produced by different elements of $P(x)$ don't overlap. This argument along with (2.1) for individual vectors conclude the proof of (2.1). We refer the reader to [Kle10] for an exposition of the details of this argument.

2.2. **The Margulis Function.** In this article, we replace the second step in the argument of Kleinbock and Margulis with a method involving systems of integral inequalities involving the functions $\tilde{\alpha}_k$ defined in (1.6). This technique was introduced in [EMM98] with the purpose of quantifying the work of Margulis on the Oppenheim conjecture. However, these functions appeared earlier in some form in the work of Dani and Margulis on the recurrence of unipotent flows on homogeneous spaces. The ideas presented in [EMM98] found generalizations to random walks on homogeneous spaces [EM04, BQ11] and the Teichmüller geodesic flow on strata of quadratic differentials [EM01, EMM15]. They were also recently used in [KKLM17] to prove a sharp upper bound on the dimension of singular systems of linear forms.

3. Notation

Here, we fix a common place for some notation that will be referred to in various parts of the article. Let $N = \binom{m+n}{n} - 1$. The following long list of functions on the space of weights $\mathcal{A}$ was introduced in the introduction and we collect them here for the convenience of the reader. For $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_{m+n})$, we define

$$\mu(\mathbf{r}) := \min \{\mathbf{r}_i + \mathbf{r}_{m+j} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\} \tag{3.1}$$

$$\nu(\mathbf{r}) := \max \{\mathbf{r}_i + \mathbf{r}_{m+j} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\} \tag{3.2}$$

$$D(\mathbf{r}) := \left\lceil \frac{\nu(\mathbf{r})}{\mu(\mathbf{r})} N - 1 \right\rceil \tag{3.3}$$

$$\lambda(\mathbf{r}) := \min \{\mathbf{r}_i : 1 \leqslant i \leqslant m+n\} \tag{3.4}$$

$$\delta(\mathbf{r}) := \mu(\mathbf{r})/N \tag{3.5}$$

$$Q(\mathbf{r}) := D(\mathbf{r}) \min(m, n)$$

$$C(\mathbf{r}) := 2Q(\mathbf{r}) \left(Q(\mathbf{r}) + 1\right)^{1/Q(\mathbf{r})} \tag{3.6}$$

$$\alpha(\mathbf{r}) := 1/Q(\mathbf{r}) \tag{3.7}$$

4. Expansion in Linear Representations

This section is dedicated to proving estimates on the average rate of expansion of curves in linear representations under the action of certain diagonalizable elements.

We begin by recalling the following notion of (C, α)-good functions introduced in [KM98] and used, in different form, in prior work of Dani, Margulis and Shah.

Definition 4.1. A function $f : \mathbb{R}^m \to \mathbb{R}$ is (C, α)-good on some subset $B \subset \mathbb{R}^m$ of finite Lebesgue measure if there exist constants $C, \alpha > 0$ such that for any $\varepsilon > 0$, one has

$$|\{x \in B : |f(x)| < \varepsilon\}| \leq C \left(\frac{\varepsilon}{\sup_{x \in B} |f(x)|} \right)^{\alpha} |B|$$

where $|\cdot|$ denotes the Lebesgue measure.

The following lemma summarizes some basic properties of (C, α)-good functions which will be useful for us. The proof follows directly from the definition.

Lemma 4.2. *Let $C, \alpha > 0$. Then,*

(1) If f is a (C, α)-good function on B, then so is $|f|$.
(2) If $f_1, \dots, f_n$ is a collection of (C, α)-good function on B, then so is $\max_k |f_k|$.

4.1. **The Exterior Power Representation.** In this section, we give a description of the coordinates of the fundamental representation of G on the following vector space

$$V = \bigoplus_{k=1}^{m+n-1} \bigwedge^k \mathbb{R}^{m+n}$$

An element $g \in \mathrm{SL}(m+n, \mathbb{R})$ acts on V via the linear map $\bigoplus_{k=1}^{m+n-1} \bigwedge^k g$.

Consider the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^{m+n}$ where $\mathbf{e}_i$ denotes the i^{th} standard basis element and $\mathbf{v}_j$ denotes the $(m+j)^{th}$ element. Now, suppose $1 \leq l < m+n$ and subsets $R \subseteq \{1, \ldots, m\}$ and $S \subseteq \{1, \ldots, n\}$ with

$$|R| + |S| = l$$

are given. Write $R = \{i_1 < \cdots < i_r\}$ and $S = \{j_1 < \cdots < j_s\}$ and let

$$\mathbf{e}_R \wedge \mathbf{v}_S := \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_r} \wedge \mathbf{v}_{j_1} \wedge \cdots \mathbf{v}_{j_s} \tag{4.1}$$

Then, the collection of monomials $\mathbf{e}_R \wedge \mathbf{v}_S$ gives a basis of $V_l = \bigwedge^l \mathbb{R}^{m+n}$ for each $1 \leq l < m+n$. Note that this basis consists of joint eigenvectors of the linear maps $\bigoplus_{k=1}^{m+n-1} \bigwedge^k g$ where g is a diagonal matrix in $\mathrm{SL}(m+n, \mathbb{R})$.

Recall the definition (3.4) of the functional λ on the space of weights $\mathcal{A}$ (defined in 1.4). Then, we see that $\exp(\lambda(\mathbf{r}))$ is the smallest eigenvalue of $\bigoplus_l \bigwedge^l g_{\mathbf{r}}$ in its action on the following subspace of V:

$$V_{\mathcal{A}}^+ = \operatorname{span}\left\{\mathbf{e}_R, \mathbf{e}_{\{1,\ldots,m\}} \wedge \mathbf{v}_S : R \subseteq \{1, \ldots, m\}, S \subset \{1, \ldots, n\}\right\} \tag{4.2}$$

Suppose $Y \in M_{m,n}$ is given. Let us describe the action of the unipotent elements $u(Y)$ on V. For $1 \leq j \leq n$, denote the j^{th} column of Y by Y_j. If $Y_j = (y_{1,j}, \ldots, y_{m,j})$, then we regard Y_j as element of $\mathbb{R}^{m+n}$ via the identification

$$Y_j \mapsto \sum_{i=1}^{m} y_{i,j} \mathbf{e}_i$$

Then, we see that

$$u(Y)\mathbf{e}_i = \mathbf{e}_i, \qquad u(Y)\mathbf{v}_j = Y_j + \mathbf{v}_j \tag{4.3}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. For the action on higher degrees, let subsets

$$I \subseteq \{1, \ldots, m\}, J \subseteq \{1, \ldots, n\}$$

with $|I| = |J|$ be given. Define $Y_{I,J}$ to be the following determinant

$$Y_{I,J} = \det\left((y_{i,j})_{i \in I, j \in J}\right) \tag{4.4}$$

Note that as I and J range over all subsets of $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$, with $|I| = |J|$, the numbers $Y_{I,J}$ give all the coordinates of $F(Y)$ where $F : M_{m,n} \to \mathbb{R}^N$ is the minors map defined in the introduction. Hence, using (4.3), we see that

$$u(Y) \cdot \mathbf{e}_R \wedge \mathbf{v}_S = \sum_{J \subset S} \sum_{\substack{I \subset \{1,\ldots,m\} \setminus R \\ |I| = |J|}} \pm Y_{I,J}\, \mathbf{e}_{R \cup I} \wedge \mathbf{v}_{S \setminus J} \tag{4.5}$$

where $Y_{I,J}$ are real numbers given by (4.4).

Denote by $\mathcal{B}$ the following collection of vectors

$$\mathcal{B} = \left\{\mathbf{e}_R, \mathbf{e}_{\{1,\ldots,m\}} \wedge \mathbf{v}_S : R \subseteq \{1, \ldots, m\}, S \subset \{1, \ldots, n\}\right\}$$

We denote by $\langle \cdot, \cdot \rangle$ the standard euclidean inner product on V. The following key technical lemma will be useful for us.

Lemma 4.3 (cf. Proposition 5.4 in [KMW10])**.** *Suppose $Y \in M_{m,n}$, $1 \leq l < m+n$ and $v \in V_l \setminus \{0\}$ are given. Let $v_{R,S}$ be a non-zero coordinate of v with respect to the basis in (4.1). Then, there exists an element $w \in \mathcal{B} \cap V_l$ such that $\langle u(Y)v, w\rangle$ is a linear combination of 1 and the components of $F(Y)$ with $v_{R,S}$ appearing as one of the coefficients.*

Proof. We remark that this result was proved in [KMW10] for integral vectors v, however the proof goes through in the general case. In the case $m = 1$, the proof is a simple calculation and we present it here for completeness. To simplify notation, we denote by $\mathbf{e}_i$ the standard basis for $\mathbb{R}^{n+1}$ and for a set $I = \{i_1, \dots, i_k\} \subset \{1, \dots, n+1\}$, we write $\mathbf{e}_I = \mathbf{e}_{i_1} \wedge \cdots \mathbf{e}_{i_k}$. Let $v \in V_l \setminus \{0\}$ and write

$$v = \sum_{I \subset \{1,\dots,n+1\}} v_I \mathbf{e}_I$$

where the sum is over index sets of cardinality l.

Let $Y = (Y_1, \dots, Y_n) \in M_{1,n} \cong \mathbb{R}^n$. First, we note that u_Y fixes $\mathbf{e}_1$ and maps $\mathbf{e}_i$ to $\mathbf{e}_i + Y_i \mathbf{e}_1$ for $i = 2, \dots, n$. This implies the following.

$$u_Y \mathbf{e}_I = \begin{cases} \mathbf{e}_I & 1 \in I, \\ \mathbf{e}_I + \sum_{i \in I} \pm Y_i \mathbf{e}_{(I \cup \{1\}) \setminus \{i\}}, & \text{otherwise.} \end{cases}$$

where the sign depends on I. In particular, we get that.

$$u_Y v = \sum_{\substack{I \subset \{1,\dots,n+1\} \\ 1 \notin I}} v_I \mathbf{e}_I + \sum_{\substack{I \subset \{1,\dots,n+1\} \\ 1 \in I}} \left(v_I + \sum_{i \notin I} \pm v_{(I \cup \{i\}) \setminus \{1\}} Y_i \right) \mathbf{e}_I$$

Now, suppose that $v_J \neq 0$ for some $J \subseteq \{1, \dots, n+1\}$. If $1 \in J$, then the $\mathbf{e}_J$ coefficient in the expression above satisfies the conclusion of the lemma. Otherwise, the conclusion holds for $\mathbf{e}_I$ for any I satisfying

$$I = (J \cup \{1\}) \setminus \{j\}$$

for some $j \in J$. $\square$

4.2. **Non-Planarity and Expansion.** The following Proposition will act as a substitute for [EM04, Lemma 4.2] which was in the context of random walks. The proof of that Lemma rests on the positivity of the top Lyapunov exponent of certain linear cocycles due to Furstenburg and Kesten. In our case, the proof relies the existence of eigenvalues bigger than 1 for the acting diagonal elements in addition to the (C, α)-good property. It will follow similar lines as the proof of Lemma 5.1 in [EMM98]. Recall that $N = N(m,n) = \binom{m+n}{n} - 1$ and that $\mathbf{d} : M_{m,n} \to \mathbb{R}^N$ denotes the minors map.

Proposition 4.4. *Let C, α and ρ be positive constants. Suppose $\psi : B \to M_{m,n}$ is a continuous map from an interval $B \subset \mathbb{R}$ satisfying the following conditions:*

(1) Any linear combination of 1 and the coordinates of $\mathbf{d} \circ \psi$ is (C, α)-good on B,
(2) For all $w = (w_0, \dots, w_N) \in \mathbb{R}^{N+1}$,

$$\sup_{r \in B} \left| w_0 + \sum_{i=1}^{N} w_i (\mathbf{d} \circ \psi)_i (r) \right| \geqslant \rho \, \|w\|$$

where $(\mathbf{d} \circ \psi)_i$ is the i^{th} coordinate of $\mathbf{d} \circ \psi$ and $\|\cdot\|$ is the euclidean norm on $\mathbb{R}^{N+1}$.

Let $V_l = \bigwedge^l \mathbb{R}^{m+n}$ for any $1 \leqslant l < m+n$. Then, for all $\beta \in (0, \alpha)$, there exists a constant $D = D(\beta, \rho, C, \alpha) > 0$ such that for all weights $\mathbf{r} \in \mathcal{A}$, $t > 0$, $1 \leqslant l < m+n$ and all $v \in V_l \setminus \{0\}$,

$$\frac{1}{|B|} \int_B \|g_{t\mathbf{r}} u(\psi(s)) v\|^{-\beta} \, ds \leqslant D e^{-\beta \lambda(\mathbf{r}) t} \|v\|^{-\beta}$$

Proof. Let $v \in V_l \setminus \{0\}$ and write

$$v = \sum_{\substack{I \subseteq \{1,\dots,m\} \\ J \subseteq \{1,\dots,n\} \\ |I|+|J|=l}} v_{I,J} \mathbf{e}_I \wedge \mathbf{v}_J$$

We will use $\|\cdot\|_\infty$ to denote the ℓ^∞-norm on V_l with respect to the basis in (4.1).

Let $V_{\mathcal{A}}^+$ be the subspace of defined in (4.2) and let $\pi_+ : V_l \to V_{\mathcal{A}}^+ \cap V_l$ denote the associated orthogonal projection. Using (4.5), one can check that the coordinates of the vector $u(\psi(\cdot)) \cdot v$ are linear combinations of 1 and the components of the map $\mathbf{d} \circ \psi$. In particular, by condition (1) and Lemma 4.2, we see that $\|\pi_+(u(\psi(r))v)\|_\infty$ is (C, α)-good on B.

Let I and J be index sets satisfying

$$\|v\|_\infty = |v_{I,J}|$$

By Lemma 4.3, there exists a vector $\mathbf{w} \in \mathcal{B} \cap V_l$ such that $\langle \pi_+(u(\psi(\cdot)) \cdot v), w\rangle$ is a linear combination of 1 and the components of $\mathbf{d} \circ \psi$, one of whose coefficients is $v_{I,J}$

Thus, by condition (2), since all norms on V are equivalent, we get that

$$\rho_1 := \sup_{r \in B} \|\pi_+(u(\psi(r))v)\|_\infty \gg \rho \|v\| \tag{4.6}$$

where the implicit constant depends only on the equivalence constant between $\|\cdot\|$ and $\|\cdot\|_\infty$.

Thus, for any $\varepsilon > 0$, by definition of the (C, α)-good property, it follows that

$$|\{r \in B : \|\pi_+(u(\psi(r))v)\|_\infty \leqslant \varepsilon\rho_1\}| \leqslant C\varepsilon^\alpha |B| \tag{4.7}$$

Let $E(v, \varepsilon)$ denote the set on the left-hand side of the above inequality. Suppose a weight $\mathbf{r} \in \mathcal{A}$ and $t > 0$ are given. Observe that for $t > 0$, we have that

$$\|g_{t\mathbf{r}} u(\psi(r))v\|_\infty \geqslant \|g_{t\mathbf{r}} \pi_+(u(\psi(r))v)\|_\infty \geqslant e^{\lambda(\mathbf{r})t} \|\pi_+(u(\psi(r))v)\|_\infty \tag{4.8}$$

Indeed, this follows from the fact that $e^{\lambda(\mathbf{r})t}$ is the smallest eigenvalue of $g_{\mathbf{r}t}$ in its action on $V_{\mathcal{A}}^+$. Fix some $\beta \in (0, \alpha)$. Then, for $n \in \mathbb{N}$, by (4.6), (4.7) and (4.8), we get

$$\begin{aligned}
&\int_{E(v, 2^{-n}\rho_1) \setminus E(v, 2^{-(n+1)}\rho_1)} \|g_t u(\psi(r))v\|_\infty^{-\beta} \, dr \\
&\qquad \leqslant e^{-\beta\lambda(\mathbf{r})t} \int_{E(v, 2^{-n}\rho_1) \setminus E(v, 2^{-(n+1)}\rho_1)} \|\pi_+(u(\psi(r))v)\|_\infty^{-\beta} \, dr \\
&\qquad \leqslant e^{-\beta\lambda(\mathbf{r})t} 2^{\beta(n+1)} \rho_1^{-\beta} C 2^{-\alpha n} |B| \\
&\qquad = \rho_1^{-\beta} 2^\beta C 2^{-(\alpha-\beta)n} e^{-\beta\lambda(\mathbf{r})t} |B| \\
&\qquad \leqslant \rho^{-\beta} \|v\|^{-\beta} 2^\beta C 2^{-(\alpha-\beta)n} e^{-\beta\lambda(\mathbf{r})t} |B|
\end{aligned}$$

Now, note that (4.7) implies that $|E(v, 0)| = 0$. Hence, since

$$B = E(v, 0) \sqcup \left(\bigsqcup_{n \geq 0} E(v, 2^{-n}\rho_1) \setminus E(v, 2^{-(n+1)}\rho_1) \right)$$

we get that

$$\frac{1}{|B|} \int_B \|g_t u(\psi(r))v\|_\infty^{-\beta} \, dr \leqslant \frac{\rho^{-\beta} 2^\beta C}{1 - 2^{\alpha-\beta}} e^{-\beta\lambda(\mathbf{r})t} \|v\|^{-\beta}$$

Thus, the claim of the Proposition follows since all norms are equivalent. □

5. Height Functions and Integral Inequalities

In this section, we establish a system of integral inequalities for the functions $\tilde{\alpha}_i$ on X defined in (1.6). Using this system of integral inequalities, we prove the main integral estimate needed for our main results in Proposition 5.3.

5.1. **A System of Integral Inequalities.** We recall the following Lemma from [EMM15] which will be useful for us:

Lemma 5.1 (Lemma 5.6 in [EMM15]). *Let $x \in X$ and let $\Lambda_1, \Lambda_2 \in P(x)$. Then,*

$$\|\Lambda_1\| \, \|\Lambda_2\| \geqslant \|\Lambda_1 \cap \Lambda_2\| \, \|\Lambda_1 + \Lambda_2\|$$

Recall the definition of the functional $\lambda : \mathcal{A} \to \mathbb{R}_+$ given (3.4) As a consequence of Proposition 4.4, we obtain the following analogue of Lemma 5.7 in [EMM15].

Lemma 5.2. *Suppose $\psi : B \to M_{m,n}$ is a continuous map from a subinterval $B \subset \mathbb{R}$ satisfying conditions (1) and (2) of Proposition 4.4 with positive constants C, α and ρ. Then, for all $\beta \in (0, \alpha)$, there exists a constant $D = D(\beta, \rho, C, \alpha) > 0$ such that for all $t > 0$, there exists a constant $\omega = \omega(t) > 0$ for all $x \in X$, $\mathbf{r} \in \mathcal{A}$ and all $1 \leqslant i \leqslant m + n - 1$ such that*

$$\frac{1}{|B|} \int_B \tilde{\alpha}_i^\beta (g_{\mathbf{r}t} u(\psi(r))x) \, dr \leqslant D e^{-\beta \lambda(\mathbf{r})t} \tilde{\alpha}_i(x) + \omega^{2\beta} \max_{1 \leqslant j \leqslant \min\{i, m+n-i\}} \left(\sqrt{\tilde{\alpha}_{i+j}(x) \tilde{\alpha}_{i-j}(x)} \right)^\beta$$

Proof. Let $\beta \in (0, \alpha)$ and let D be the maximum of the constants in the conclusion of Proposition 4.4 applied to the vector spaces $\bigwedge^l \mathbb{R}^{m+n}$ for each $1 \leq l < m + n$. Let $t > 0$ be given and define ω as follows:

$$\omega = \sup_{\substack{r \in B \\ \mathbf{r} \in \mathcal{A}}} \max \left\{ \|g_{\mathbf{r}t} u(\psi(r))\| \, , \left\| (g_{\mathbf{r}t} u(\psi(r)))^{-1} \right\| \right\} \tag{5.1}$$

where $\|\cdot\|$ is the euclidean norm on V.

Fix some $\mathbf{r} \in \mathcal{A}$. Then, for all $r \in B$ and all $v \in V \setminus \{0\}$,

$$\omega^{-1} \leqslant \frac{\|g_{\mathbf{r}t} u(\psi(r))v\|}{\|v\|} \leqslant \omega \tag{5.2}$$

Now, suppose v is an x-integral monomial in degree i. Then, by Proposition 4.4,

$$\int_B \|g_{\mathbf{r}t} u(\psi(r))v\|^{-\beta} \, dr \leqslant D e^{-\beta \lambda_l(\mathbf{r})t} \|v\|^{-\beta} |B| \tag{5.3}$$

Let Λ_i be a subgroup of the lattice corresponding to x such that

$$\tilde{\alpha}_i(x) = \|\Lambda_i\|^{-1}$$

Following [EMM15], let Ψ denote the finite subset of $P(x)$ of rank i subgroups L of x satisfying

$$\|L\|^{-1} \geq \omega^{-1} \tilde{\alpha}_i(x)$$

The finiteness of Ψ follows from the discreteness of the lattice x. Suppose that Λ_i is the only element of Ψ. In this case, by 5.2, we see that for all $r \in B$, we have

$$\tilde{\alpha}_i(g_t u(\psi(r))x) = \|g_t u(\psi(r)) \Lambda_i\|^{-1}$$

Therefore, in this case, by (5.3), we get

$$\int_B \tilde{\alpha}_i(g_t u(\psi(r))x)^\beta \, dr = \int_B \|g_t u(\psi(r)) \Lambda_i\|^{-\beta} \, dr \leqslant D e^{-\beta \lambda_i(\mathbf{r})t} \tilde{\alpha}_i(x)^\beta |B| \leqslant D e^{-\beta \lambda(\mathbf{r})t} \tilde{\alpha}_i(x)^\beta |B| \tag{5.4}$$

where used the fact that $\lambda = \min_i \lambda_i$. Now, assume that Ψ contains an element $L \neq \Lambda_i$ of rank i. Then, the group $\Lambda_i + L$ has rank $i + j$ for some $j > 0$. Moreover, by definition of Ψ and ω and

Lemma 5.1, for all $r \in B$,

$$\tilde{\alpha}_i(g_{\mathbf{r}t}u(\psi(r))x) \leqslant \omega\tilde{\alpha}_i(x) = \frac{\omega}{\|\Lambda_i\|} \leqslant \frac{\omega^2}{\sqrt{\|\Lambda_i\|\,\|L\|}} \leqslant \frac{\omega^2}{\sqrt{\|\Lambda_i + L\|\,\|\Lambda_i \cap L\|}} \leqslant \omega^2 \max_{1\leqslant j\leqslant \min\{i,m+n-i\}} \sqrt{\tilde{\alpha}_{i+j}(x)\tilde{\alpha}_{i-j}(x)}$$

Combining this estimate with (5.4), we get the desired conclusion. □

5.2. **The Contraction Hypothesis on X.** Using the integral estimates for the functions $\tilde{\alpha}_i$ obtained in the previous section, we obtain an integral estimate for a family of functions $\tilde{\alpha}_{\epsilon,\beta}$, which depend on 2 parameters ϵ and β. We show that the average over the pushforward of a curve of the form $u(\psi(r))x$ under $g_{\mathbf{r}t}$ experiences contraction whenever the value $\tilde{\alpha}_{\epsilon,\beta}(x)$ is sufficiently large. In the terminology of Benoist and Quint, such subharmonic behavior of $\tilde{\alpha}_{\epsilon,\beta}$ is referred to as the contraction hypothesis on the space of lattices.

Following [EMM15], for $\epsilon, \beta > 0$, define the following function for every $x \in X = \mathrm{SL}(m+n,\mathbb{R})/\mathrm{SL}(m+n,\mathbb{Z})$:

$$\tilde{\alpha}_{\epsilon,\beta}(x) := \sum_{i=0}^{m+n} \epsilon^{i(m+n-i)}\tilde{\alpha}_i^{\beta}(x) \tag{5.5}$$

The weights $\epsilon^{i(m+n-i)}$ and exponent β allow us to upgrade the integral estimates in 5.2 to an integral estimate for $\tilde{\alpha}_{\epsilon,\beta}$. The following is the main result of this section.

Proposition 5.3. *Suppose $\psi : B \to M_{m,n}$ is a continuous map from a subinterval $B \subset \mathbb{R}$ satisfying conditions (1) and (2) of Proposition 4.4 with positive constants C, α and ρ. Then, for all $\beta \in (0,\alpha)$, there exists a constant $c_0 = c_0(\beta,\rho,C,\alpha) > 0$ such that for all $t > 0$, there exists $\epsilon = \epsilon(t) > 0$ with the following property: for all $x \in X$ and all $\mathbf{r} \in \mathcal{A}$,*

$$\frac{1}{|B|}\int_B \tilde{\alpha}_{\epsilon,\beta}(g_{\mathbf{r}t}u(\psi(r))x)\,dr \leqslant c_0 e^{-\beta\lambda(\mathbf{r})t}\tilde{\alpha}_{\epsilon,\beta}(x) + 2$$

where $\lambda(\mathbf{r})$ is defined in (3.4).

Proof. Let $\beta \in (0,\alpha)$ and let D be as in Lemma 5.2. Let $t > 0$ and let $\omega > 0$ be as in the same Lemma. Let $\epsilon > 0$ to be chosen later. For each $0 \leqslant i \leqslant m+n$, define $q(i) = i(m+n-i)$. Note that

$$q(i) - \frac{q(i-j)+q(i+j)}{2} = j^2$$

Moreover, for all i, we have

$$\epsilon^{q(i)}\tilde{\alpha}_i^{\beta} \leqslant \tilde{\alpha}_{\varepsilon,\beta}$$

Then, since $\tilde{\alpha}_0 \equiv \tilde{\alpha}_d \equiv 1$, by Lemma 5.2, for all $x \in X$

$$\begin{aligned}\frac{1}{|B|}\int_B \tilde{\alpha}_{\epsilon,\beta}(g_t u(\psi(r))x)\, dr &\leqslant De^{-\beta\lambda(\mathbf{r})t}\tilde{\alpha}_{\epsilon,\beta}(x) + 2 \\ &+ \omega^{2\beta}\sum_{i=1}^{m+n-1}\epsilon^{q(i)}\max_{1\leqslant j\leqslant \min\{i,m+n-i\}}\left(\sqrt{\tilde{\alpha}_{i+j}(x)\tilde{\alpha}_{i-j}(x)}\right)^{\beta} \\ &= De^{-\beta\lambda(\mathbf{r})t}\tilde{\alpha}_{\epsilon,\beta}(x) + 2 \\ &+ \omega^{2\beta}\sum_{i=1}^{m+n-1}\max_{1\leqslant j\leqslant \min\{i,m+n-i\}}\epsilon^{j^2}\sqrt{\epsilon^{q(i+j)}\tilde{\alpha}_{i+j}(x)^{\beta}\epsilon^{q(i-j)}\tilde{\alpha}_{i-j}(x)^{\beta}} \\ &\leqslant De^{-\beta\lambda(\mathbf{r})t}\tilde{\alpha}_{\epsilon,\beta}(x) + 2 + \epsilon(m+n)\omega^{2\beta}\tilde{\alpha}_{\epsilon,\beta}(x)\end{aligned}$$

Finally, choose ϵ as follows

$$\epsilon = \inf_{\mathbf{r}\in\mathcal{A}}\frac{De^{-\beta\lambda(\mathbf{r})t}}{(m+n)\omega^{2\beta}} > 0$$

Here, we use the fact that $\lambda(\mathbf{r}) \leqslant mn$ for all $\mathbf{r} \in \mathcal{A}$. Then, the conclusion follows with $c_0 = 2D$. $\square$

The following Lemma provides us with several properties of the functions $\tilde{\alpha}$ which follow from the definition and Mahler's compactness criterion.

Lemma 5.4. *Let $\epsilon, \beta > 0$ be given. Then, the following holds*

(1) Given a bounded neighborhood $\mathcal{O}$ of identity in G, there exists a constant $C_{\mathcal{O}} > 1$ (independent of ϵ), such that for all $g \in \mathcal{O}$ and all $x \in X$,

$$C_{\mathcal{O}}^{-1}\tilde{\alpha}_{\epsilon,\beta}(x) \leqslant \tilde{\alpha}_{\epsilon,\beta}(gx) \leqslant C_{\mathcal{O}}\tilde{\alpha}_{\epsilon,\beta}(x)$$

(2) For all $M > 0$, the set $\overline{\tilde{\alpha}_{\epsilon,\beta}^{-1}([0,M])}$ is compact.

6. Approximation by Polynomials and Supremum Estimate

This section is dedicated to providing the link between our curve φ and the integral estimates obtained in the previous section. This is done by showing that certain polynomial curves approximating our curve satisfy the conditions of Proposition 4.4.

We keep the same notation $G = \mathrm{SL}(m+n,\mathbb{R})$, $\Gamma = \mathrm{SL}(m+n,\mathbb{Z})$ and $X = G/\Gamma$ as in previous sections. Throughout this section, we will use

$$\varphi : B \to M_{m,n}$$

to denote a continuous map from a compact interval $B \subset \mathbb{R}$ into the space of m by n matrices. To avoid trivialities, we will assume that φ is defined on a neighborhood of B.

6.1. **Approximation of Curves with Polynomials.** With the estimate in Proposition 5.3 in place for a fixed time step t, the following Lemma will allow us to iterate this integral estimate in order to obtain an estimate for g_{nt} for all $n \geq 1$. Since the curve φ is not normalized by g_t, we approximate the curve in a neighborhood of each point with a polynomial coming from its Taylor expansion. Then, we push these polynomial curves forward by g_t.

Recall the definition of the functions μ, ν and D on the space of weights $\mathcal{A}$ given in (3.1), (3.2) and (3.3) respectively.

The following is the main result of this section.

Lemma 6.1. *Let $\tilde{\alpha} : X \to (0,\infty)$ be a function satisfying Lemma 5.4. Then, there exists a constant $C_1 > 1$, depending only on the curve φ such that for all $q \in \mathbb{N}$, $\mathbf{r} \in \mathcal{A}$, $t > 0$: the following holds.*

Suppose φ is $(D(\mathbf{r})+1)$-times continuously differentiable and let $\delta(\mathbf{r}) = \mu(\mathbf{r})/N$. Then, for all subintervals $J \subseteq B$ of radius at least $e^{-\delta(\mathbf{r})(q+1)t}$, one has

$$\int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds \leqslant C_1 \int_J \int_{-1}^{1} \tilde{\alpha}(g_t u(P_{q,s,\mathbf{r}}(r)) g_{q\mathbf{r}t} u(\varphi(s))x)\, dr\, ds \tag{6.1}$$

where for $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_{m+n})$, $P_{q,s,\mathbf{r}}(r)$ is the matrix in $M_{m,n}$ whose (i,j)-entry is given by

$$(P_{q,s,\mathbf{r}}(r))_{i,j} = e^{qt(\mathbf{r}_i+\mathbf{r}_{m+j})} \sum_{k=1}^{D(\mathbf{r})} \frac{(re^{-\delta(\mathbf{r})(q+1)t})^k}{k!} \varphi_{i,j}^{(k)}(s) \tag{6.2}$$

Proof. First, we note that for all $r \in [-1,1]$, we have

$$J \subseteq J \pm re^{-\delta(\mathbf{r})(q+1)t} \tag{6.3}$$

Using positivity of $\tilde{\alpha}$, (6.3) and a change of variable, we get

$$\begin{aligned}
\int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds &= \int_0^1 \int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds\, dr \\
&\leqslant \int_0^1 \int_{J\pm re^{-\delta(\mathbf{r})(q+1)t}} \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds\, dr \qquad \text{by 6.3} \\
&= \int_{-1}^1 \int_{J+re^{-\delta(\mathbf{r})(q+1)t}} \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds\, dr \\
&= \int_{-1}^1 \int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s+re^{-\delta(\mathbf{r})(q+1)t}))x)\, ds\, dr
\end{aligned}$$

Then, using the Taylor expansion of φ up to degree $D(\mathbf{r})$, we get

$$\int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds \leqslant \int_J \int_{-1}^1 \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s) + Q_{q,s,\mathbf{r}}(r) + O(e^{-\delta(\mathbf{r})(q+1)t(D(\mathbf{r})+1)}))x)\, dr\, ds$$

where $Q_{q,s,\mathbf{r}}(r)$ is the matrix in $M_{m,n}$ whose (i,j)-entry is given by

$$(Q_{q,s,\mathbf{r}}(r))_{i,j} = \sum_{k=1}^{D(\mathbf{r})} \frac{(re^{-\delta(\mathbf{r})(q+1)t})^k}{k!} \varphi_{i,j}^{(k)}(s) \tag{6.4}$$

Moreover, by definition of $g_{q\mathbf{r}t}$ and $u(Y)$, we have

$$g_{q\mathbf{r}t}u(Y)g_{-q\mathbf{r}t} = u((e^{qt(\mathbf{r}_i+\mathbf{r}_{m+j})}Y_{i,j})_{1\le i\le m, 1\le j\le n})$$

But, it follows from the definitions that $\mathbf{r}_i + \mathbf{r}_{m+j} \le \nu(\mathbf{r}) \le \delta(\mathbf{r})(D(\mathbf{r})+1)$ for all i and j. Thus, we get

$$\begin{aligned}
\int_J \tilde{\alpha}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\, ds &\leqslant \int_J \int_{-1}^1 \tilde{\alpha}(u(O(1))g_{(q+1)\mathbf{r}t}u(\varphi(s) + Q_{q,s,\mathbf{r}}(r))x)\, dr\, ds \\
&= \int_J \int_{-1}^1 \tilde{\alpha}(u(O(1))g_{\mathbf{r}t}u(P_{q,s,\mathbf{r}}(r))g_{q\mathbf{r}t}u(\varphi(s))x)\, dr\, ds
\end{aligned}$$

Note that $u(O(1))$ belongs to a bounded neighborhood of identity independently of t, $\mathbf{r}$ and q. Hence, by (1) of Lemma 5.4, there exists a constant $C_1 > 1$ such that for all $y \in X$,

$$C_1^{-1}\tilde{\alpha}((y) \leqslant \tilde{\alpha}((u(O(1))y) \leqslant C_1\tilde{\alpha}((y)$$

This concludes the proof. ∎

6.2. **A Lower Estimate for the Supremum.** The aim of this section is to verify that the polynomial approximations of our curves satisfy the conditions of Proposition 4.4. Recall the definition of the minors map $\mathbf{d} : M_{m,n} \to \mathbb{R}^N$ whose components are given in (4.4), where $N = \binom{m+n}{n} - 1$. We also recall the definition of the functions μ, ν and D on the space of weights $\mathcal{A}$ given in (3.1), (3.2) and (3.3) respectively. The following Proposition is the precise statement.

Proposition 6.2. *Suppose φ is $(D(\mathbf{r})+1)$-times continuously differentiable. Let $P_{q,s,\mathbf{r}} : [-1,1] \to M_{m,n}$ be its Taylor polynomial defined in (6.2). Then,*

(1) Any linear combination of 1 and the components of the map $\mathbf{d} \circ P_{q,s,\mathbf{r}}$ is (C,α)-good on $[-1,1]$, where for $Q = D(\mathbf{r}) \min(m,n)$,

$$C = 2Q(Q+1)^{1/Q}, \qquad \alpha = 1/Q$$

(2) Assume that φ is strongly non-planar on B. Then, there exists $\rho > 0$, depending only on the curve φ, such that for all $s \in B$, $q \in \mathbb{N}$, $\mathbf{r} \in \mathcal{A}$ and $t > 0$, the following holds for all $w \in \mathbb{R}^{N+1}$:

$$\sup_{r \in [-1,1]} |\langle w, (1, \mathbf{d} \circ P_{q,s,\mathbf{r}}(r))\rangle| \geqslant \rho \|w\|$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product and $\|\cdot\|$ is the euclidean norm on $\mathbb{R}^{N+1}$.

It was shown in [KM98] that polynomials are (C,α)-good for constants C and α depending only on their degrees. More precisely, one has

Lemma 6.3 (Proposition 3.2 in [KM98], Lemma 4.1 in [DM93])**.** *For any $k \in \mathbb{N}$, any polynomial in $\mathbb{R}[x]$ of degree at most k is $(2k(k+1)^{1/k}, 1/k)$-good on any interval in $\mathbb{R}$.*

Thus, the remaining task is to find a uniform constant $\rho > 0$ which satisfies condition (2) of Proposition 4.4. The following observation will be useful for us.

Lemma 6.4. *Suppose $b > 0$, $Y = (Y_{i,j}) \in M_{m,n}$ and $\mathbf{r} \in \mathcal{A}$ are given. Consider the matrix $Z \in M_m, n$ whose (i,j)-entry are given by $Z_{i,j} = b^{\mathbf{r}_i + \mathbf{r}_{m+j}} Y_{i,j}$ for $1 \le i \le m$ and $1 \le j \le n$. Let $I \subseteq \{1, \ldots, m\}$ and $J \subseteq \{1, \ldots, n\}$ be such that $|I| = |J|$. Then,*

$$\det(Z_{i,j})_{i \in I, j \in J} = b^{\sigma(I,J)} \det(Y_{i,j})_{i \in I, j \in J}$$

where $\sigma(I,J) := \sum_{i \in I, j \in J} \mathbf{r}_i + \mathbf{r}_{m+j}$.

Proof. We proceed by induction on the cardinality of the index sets I and J. When $|I| = |J| = 1$, there is nothing to prove. Suppose the statement holds for all sets I' and J' with $1 \le |I'| = |J'| = k < \min(m,n)$ and let $I \subseteq \{1, \ldots, m\}$ and $J \subseteq \{1, \ldots, n\}$ be such that $|I| = |J| = k+1$. Using Laplace expansion of the determinant of the submatrix $(Z_{i,j})_{i \in I, j \in J}$ along the i_0 row for some $i_0 \in I$, we get

$$\det(Z_{i,j})_{i \in I, j \in J} = \sum_{j \in J} Z_{i_0,j} C_{i_0,j} \tag{6.5}$$

where $C_{i_0,j}$ denotes the cofactor of $(Z_{i,j})_{i \in I, j \in J}$ with the i_0 row and j column removed. By our induction hypothesis, we get

$$C_{i_0,j} = b^{\sigma(I \setminus \{i_0\}, J \setminus \{j\})} D_{i_0,j} \tag{6.6}$$

where $D_{i_0,j}$ denotes the cofactor of $(Y_{i,j})_{i \in I, j \in J}$ with the i_0 row and j column removed. Thus, plugging (6.6) into (6.5) and using the definition of $Z_{i_0,j}$, we get

$$\begin{aligned} \det(Z_{i,j})_{i \in I, j \in J} &= \sum_{j \in J} b^{\mathbf{r}_{i_0} + \mathbf{r}_{m+j}} Y_{i_0,j} b^{\sigma(I \setminus \{i_0\}, J \setminus \{j\})} D_{i_0,j} \\ &= b^{\sigma(I,J)} \sum_{j \in J} Y_{i_0,j} D_{i_0,j} = b^{\sigma(I,J)} \det(Y_{i,j})_{i \in I, j \in J} \end{aligned}$$

which is the desired conclusion. ∎

We are now ready for the proof of the Proposition. We will need the following notation. For a matrix $Y = (Y_{i,j}) \in M_{m,n}$ and subsets $I \subseteq \{1,\dots,m\}$ and $J \subseteq \{1,\dots,n\}$ such that $|I| = |J|$, we use the following notation

$$(\mathbf{d}(Y))_{I,J} = \det(Y_{i,j})_{i\in I, j\in J} \tag{6.7}$$

Proof of Proposition 6.2. Since $P_{q,s,\mathbf{r}}$ is a polynomial map of degree at most $D(\mathbf{r})$, then the map $\mathbf{d}\circ P_{q,s,\mathbf{r}}$ is a polynomial of degree at most $D(\mathbf{r})\min(m,n)$. Thus, (1) follows by Lemma 6.3.

Suppose that (2) does not hold. Then, we can find a sequence of vectors $w^{(\ell)} \in \mathbb{R}^{N+1}$ of norm 1, natural numbers q_ℓ, points $s_\ell \in [-1,1]$, $t_\ell > 0$ and weights $\mathbf{r}_\ell$ such that

$$\sup_{|r|\leqslant 1} \left| \langle w^{(\ell)}, (1, \mathbf{d}\circ P_{q_\ell,s_\ell,\mathbf{r}_\ell}(r))\rangle \right| < \frac{1}{\ell} \tag{6.8}$$

By Lemma 6.4, applied with $b = e^{q_\ell t_\ell}$, for all subsets $I \subseteq \{1,\dots,m\}$ and $J \subseteq \{1,\dots,n\}$ such that $|I| = |J|$, we have

$$(\mathbf{d}\circ P_{q_\ell,s_\ell,\mathbf{r}_\ell}(r))_{I,J} = e^{q_\ell t_\ell \sigma(I,J)} \left(\mathbf{d}\circ Q_{q_\ell,s_\ell,\mathbf{r}_\ell}(r)\right)_{I,J}$$

where $\sigma(I,J) := \sum_{i\in I, j\in J} \mathbf{r}_i + \mathbf{r}_{m+j}$ and $Q_{q_\ell,s_\ell,\mathbf{r}_\ell}(r)$ is the matrix whose (i,j)-entry is given by

$$(Q_{q_\ell,s_\ell,\mathbf{r}_\ell}(r))_{i,j} = \sum_{k=1}^{D(\mathbf{r})} \frac{(re^{-\delta(\mathbf{r})(q_\ell+1)t_\ell})^k}{k!} \varphi_{i,j}^{(k)}(s_\ell)$$

where $\delta(\mathbf{r}) = \mu(\mathbf{r})/N$ and $\mu(\mathbf{r})$ is defined in (3.1). Recall that we are fixing some order on the set of pairs of index subsets I and J which we are using to define the minors map $\mathbf{d}$. In particular, we may use such pairs to index coordinates of vectors of $\mathbb{R}^N$. Hence, for each ℓ, we may write the vectors $w^{(\ell)}$ in coordinates in the following form

$$w^{(\ell)} = \left(w_0^{(\ell)}, (w_{I,J}^{(\ell)})_{\substack{I\subseteq\{1,\dots,m\}\\ J\subseteq\{1,\dots,n\}}} \right)$$

For each ℓ, we define vectors $v^{(\ell)}$ by setting $v_0^{(\ell)} = w_0^{(\ell)}$ and for each pair of index sets I and J, we set

$$v_{I,J}^{\ell} = e^{q_\ell t_\ell \sigma(I,J)} w_{I,J}^{(\ell)}$$

For simplicity of notation, we define the following functions

$$f_\ell(r) := \langle v^{(\ell)}, (1, \mathbf{d}\circ Q_{q_\ell,s_\ell,\mathbf{r}_\ell}(re^{\delta(\mathbf{r})(q_\ell+1)t}))\rangle$$

Thus, (6.8) becomes

$$\sup_{|r|\leqslant e^{-\delta(\mathbf{r})(q_\ell+1)t_\ell}} |f_\ell(r)| < \frac{1}{\ell} \tag{6.9}$$

Assume for the moment that

$$q_\ell t_\ell \to \infty$$

Hence, using say the centered difference approximation formula of the p^{th} derivative $f_\ell^{(p)}$, Taylor's theorem and (6.9), we see that

$$\left| f_\ell^{(p)}(0) \right| \leqslant \frac{2^p}{\ell} e^{p\delta(\mathbf{r})(q_\ell+1))t_\ell} + O(e^{-2\delta(\mathbf{r})(q_\ell+1)t_\ell}) \tag{6.10}$$

For $w = (w_0,\dots,w_N) \in \mathbb{R}^{N+1}$, let $\overline{w}$ denote the following

$$\overline{w} = (w_1,\dots,w_N) \in \mathbb{R}^N$$

Note that for each $1 \le p \le N$, by the chain rule, we have

$$f_\ell^{(p)}(0) = \langle \overline{v}^{(\ell)}, D^{(p)}\mathbf{d}\circ\varphi^{(p)}(s_\ell)\rangle \tag{6.11}$$

where we $D^{(p)}\mathbf{d}$ denotes the p^{th} derivative of the map $\mathbf{d}$. Here, we implicitly identified $M_{m,n}$ and $\mathbb{R}^N$ with their tangent spaces at the relevant points.

We can collect the information given by (6.9) in the following concise form. Consider the $(N \times N)-$matrix $\Phi(s_\ell)$ whose p^{th} row is given by $D^{(p)}\mathbf{d} \circ \varphi^{(p)}(s_\ell)$. Then, using (6.11), we see that the inequalities (6.9) imply

$$\|\Phi(s_\ell)\overline{v}_\ell\|_\infty \leqslant \frac{2^N}{\ell} e^{N\delta(\mathbf{r})(q_\ell+1))t_\ell} + O(e^{-2\delta(\mathbf{r})(q_\ell+1)t_\ell})$$

where $\|\cdot\|_\infty$ is the ℓ^∞-norm on $\mathbb{R}^N$ with respect to the standard basis. By equivalence of norms on $\mathbb{R}^N$ and the fact that $N\delta(\mathbf{r}) = \mu(\mathbf{r})$, this implies

$$\|\Phi(s_\ell)\overline{v}_\ell\| \ll \frac{2^N}{\ell} e^{\mu(\mathbf{r})(q_\ell+1))t_\ell} + O(e^{-2\delta(\mathbf{r})(q_\ell+1)t_\ell}) \tag{6.12}$$

where $\mu(\mathbf{r})$ is defined in (3.2). Since φ is assumed to be strongly non-planar, this implies that for all ℓ, the matrix $\Phi(s_\ell)$ is invertible. Thus, in particular, we obtain

$$\left\|\Phi(s_\ell)\overline{v}^{(\ell)}\right\| \geqslant \left\|\Phi(s_\ell)^{-1}\right\|^{-1} \left\|\overline{v}^{(\ell)}\right\| \tag{6.13}$$

Note that $\mu(\mathbf{r}) \leq \sigma(I,J)$ for all non-empty index sets I and J with $|I| = |J|$. Thus, we get that

$$e^{\mu(\mathbf{r})(q_\ell+1))t_\ell} \left\|\overline{w}^{(\ell)}\right\| \ll e^{\mu(\mathbf{r})(q_\ell+1))t_\ell} \left\|\overline{w}^{(\ell)}\right\|_\infty \leqslant \left\|\overline{v}^{(\ell)}\right\|_\infty \ll \left\|\overline{v}^{(\ell)}\right\| \tag{6.14}$$

Finally, since $P_{q_\ell,s_\ell,\mathbf{r}_\ell}(0) = \mathbf{0}$, then inequality (6.8) implies that

$$|w_0^{(\ell)}| \leqslant \frac{1}{\ell} \to 0$$

In particular, since $\left\|w^{(\ell)}\right\| = 1$, it follows that for ℓ sufficiently large

$$\left\|\overline{w}^{(\ell)}\right\| \gg \left\|w^{(\ell)}\right\| = 1 \tag{6.15}$$

Therefore, combining (6.12), (6.13), (6.14) and (6.15), we obtain

$$\left\|\Phi(s_\ell)^{-1}\right\|^{-1} \ll \frac{2^N}{\ell} + O(e^{-(2\delta(\mathbf{r})+\mu(\mathbf{r}))(q_\ell+1)t_\ell}) \xrightarrow{\ell\to\infty} 0 \tag{6.16}$$

Note that here we use the fact that $\mu(\mathbf{r}) > 0$. However, the compactness of the interval B, strong non-planarity of φ on B and the continuity of the map $s \mapsto \left\|\Phi(s)^{-1}\right\|$, implies that $\left\|\Phi(s_\ell)^{-1}\right\|^{-1}$ is bounded below, contradicting (6.16).

Now, suppose that for all ℓ,

$$\sup_\ell q_\ell t_\ell < \infty$$

Then, we can approximate the derivatives of f_ℓ at 0 using points at distance $\leq \ell^{-1/N+1}$. These points will belong to the interval $|r| \leq e^{-\delta(\mathbf{r})(q_\ell+1)t_\ell}$ for ℓ sufficiently large. In particular, instead of inequality (6.10), we get

$$\left|f_\ell^{(p)}(0)\right| \leqslant \frac{2^p}{\ell}\ell^{p/N+1} + O(\ell^{-2/N+1}) \to 0$$

for all $p \leq N$. Then, one checks that the same argument as in the previous case yields a contradiction. This completes the proof. $\square$

Remark 6.5. We note that the proof of Proposition 6.2 shows that ρ in condition (2) depends only on the degree of non-degeneracy of φ measured using bounds on the function $s \mapsto \|\Phi(s)\|$ and the estimate in (6.16) utilized in the proof.

7. Integral Estimates and Non-divergence of Shrinking Curves

In this section, we provide a proof of Theorem 7.1. Recall the definition of the functions $\mu, D, \lambda, \delta, C$ and α on the space of weights $\mathcal{A}$ given in the introduction. Recall that $N = \binom{m+n}{n} - 1$.

The following is the main result of this section which implies non-divergence of push-forwards of shrinking strongly planar curves and Theorem 1.6 in the introduction.

Theorem 7.1. *Suppose a weight $\mathbf{r} \in \mathcal{A}$ is given and let $D = \max\{N, D(\mathbf{r})\}$. Suppose $\varphi : B \to M_{m,n}$ is a strongly non-planar C^{D+1}-curve on a compact interval B. For every $\beta \in (0, \min(\delta(\mathbf{r}), \alpha(\mathbf{r})))$, there exists $t_0 \geqslant 1$, depending on φ, β, $\lambda(\mathbf{r})$ and $C(\mathbf{r})$ such that the following holds: for every $t \geqslant t_0$, there exists $\epsilon = \epsilon(t) > 0$ such that for all $\delta \in [0, \beta)$, $x \in X$ and $q \in \mathbb{N}$, one has*

$$\sup_{q \geqslant 1, s_0 \in B} \frac{1}{|J_q|} \int_{J_q + s_0} \tilde{\alpha}_{\epsilon,\beta}(g_{q\mathbf{r}t} u(\varphi(s))x)\, ds < \infty \tag{7.1}$$

where $J_q := [-e^{-\delta qt}, e^{-\delta qt}]$. Moreover, the supremum can be taken to be uniform over basepoints x in a fixed compact set.

Proof. The statement will be proved by induction on q. We begin by choosing the necessary constants. Fix some $\beta \in (0, \min(\delta(\mathbf{r}), \alpha(\mathbf{r})))$ and $\delta \in [0, \beta)$. For $t > 0$, let $\epsilon(t) > 0$ be as in Proposition 5.3 and define

$$M_0(t) := \sup_{s_0 \in B} \tilde{\alpha}_{\epsilon,\beta}(u(\varphi(s))x) \tag{7.2}$$

We note that from the definition of the function $\tilde{\alpha}_{\epsilon,\beta}$ and compactness of B, $M_0(t)$ is a decreasing function of t which is bounded below by 2.

Let C_1 and c_0 be the constants provided by Lemma 6.1 and Proposition 5.3 respectively. Note that these constants depend only on φ and β. Choose $t_0 > 0$ so that

$$2C_1 c_0 = e^{(\beta - \delta)\lambda(\mathbf{r}) t_0}$$

Suppose $q \geq 1$, $t > t_0$ and $s_0 \in B$ are given. We define $M > 0$ as follows

$$M = \max\left\{ 2C_1 c_0 e^{-\beta\lambda(\mathbf{r})t} M_0(t) + 4C_1, 4C_1 \left(1 - 2C_1 c_0 e^{-(\beta-\delta)\lambda(\mathbf{r})t}\right)^{-1} \right\} \tag{7.3}$$

We claim that M gives an upper bound on the supremum in (7.1).

By Lemma 6.1, we have

$$\int_{J_{q+1}+s_0} \tilde{\alpha}_{\epsilon,\beta}(g_{(q+1)\mathbf{r}t} u(\varphi(s))x)\, ds \leqslant C_1 \int_{J_{q+1}+s_0} \int_{-1}^{1} \tilde{\alpha}_{\epsilon,\beta}(g_t u(P_{q,s,\mathbf{r}}(r)) g_{q\mathbf{r}t} u(\varphi(s))x)\, dr\, ds \tag{7.4}$$

where

$$(P_{q,s,\mathbf{r}}(r))_{i,j} = e^{qt(\mathbf{r}_i + \mathbf{r}_{m+j})} \sum_{k=1}^{D(\mathbf{r})} \frac{(re^{-\delta(\mathbf{r})(q+1)t})^k}{k!} \varphi_{i,j}^{(k)}(s) \tag{7.5}$$

By Proposition 6.2, the polynomials in (6.2) satisfy conditions (1) and (2) of Proposition 4.4. Thus, the integral estimate in Proposition 5.3 applies with $f(r) = P_{q,s,\mathbf{r}}(r)$. Therefore, by (7.4), we get

$$\int_{J_{q+1}+s_0} \tilde{\alpha}_{\epsilon,\beta}(g_{(q+1)\mathbf{r}t} u(\varphi(s))x)\, ds \leqslant 2C_1 c_0 e^{-\beta\lambda(\mathbf{r})t} \int_{J_{q+1}+s_0} \tilde{\alpha}_{\epsilon,\beta}(g_{q\mathbf{r}t} u(\varphi(s))x)\, ds + 4C_1 |J_{q+1}| \tag{7.6}$$

Hence, for $q = 0$, since $M_0(t) \leq M_0(t_0)$, it follows that

$$\begin{aligned} \frac{1}{|J_1|} \int_{J_1+s_0} \tilde{\alpha}_{\epsilon,\beta}(g_t u(\varphi(s))x)\, ds &\leqslant 2C_1 c_0 e^{-\beta\lambda(\mathbf{r})t} \frac{1}{|J_1|} \int_{J_1+s_0} \tilde{\alpha}_{\epsilon,\beta}(u(\varphi(s))x)\, ds + 4C_1 \\ &\leqslant 2C_1 c_0 e^{-\beta\lambda(\mathbf{r})t} M_0(t) + 4C_1 \leqslant M \end{aligned}$$

Now, suppose that the claim holds for all $1 \leq k < q+1$. Applying (7.6) and the fact that $J_{q+1} \subset J_q$, we get

$$\begin{aligned}\int_{J_{q+1}+s_0} \tilde{\alpha}_{\epsilon,\beta}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\,ds &\leqslant 2C_1c_0e^{-\beta\lambda(\mathbf{r})t}\int_{J_{q+1}+s_0}\tilde{\alpha}_{\epsilon,\beta}(g_{q\mathbf{r}t}u(\varphi(s))x)\,ds + 4C_1|J_{q+1}| \\ &\leqslant 2C_1c_0e^{-\beta\lambda(\mathbf{r})t}\int_{J_q+s_0}\tilde{\alpha}_{\epsilon,\beta}(g_{q\mathbf{r}t}u(\varphi(s))x)\,ds + 4C_1|J_{q+1}|\end{aligned}$$

By the induction hypothesis, since $|J_q|/|J_{q+1}| = e^{\delta\lambda(\mathbf{r})t}$, this implies

$$\frac{1}{|J_{q+1}|}\int_{J_{q+1}+s_0}\tilde{\alpha}_{\epsilon,\beta}(g_{(q+1)\mathbf{r}t}u(\varphi(s))x)\,ds \leqslant 2C_1c_0e^{-\beta\lambda(\mathbf{r})t}e^{\delta\lambda(\mathbf{r})t}M + 4C_1 \leqslant M$$

This completes the proof. □

References

[BQ11] Yves Benoist and Jean-Francois Quint, *Random walks on finite volume homogeneous spaces*, Inventiones mathematicae **187** (2011), no. 1, 37–59.

[DM93] S. G. Dani and G. A. Margulis, *Limit distributions of orbits of unipotent flows and values of quadratic forms*, I. M. Gelfand Seminar, Adv. Soviet Math., vol. 16, Amer. Math. Soc., Providence, RI, 1993, pp. 91–137. MR 1237827

[EM01] Alex Eskin and Howard Masur, *Asymptotic formulas on flat surfaces*, Ergodic Theory Dynam. Systems **21** (2001), no. 2, 443–478. MR 1827113

[EM04] Alex Eskin and Gregory Margulis, *Recurrence properties of random walks on finite volume homogeneous manifolds*, Random Walks and Geometry: Proceedings of a Workshop at the Erwin Schrödinger Institute, Vienna, June 18-July 13, 2001, Walter de Gruyter, 2004, p. 431.

[EMM98] Alex Eskin, Gregory Margulis, and Shahar Mozes, *Upper bounds and asymptotics in a quantitative version of the oppenheim conjecture*, Annals of Mathematics **147** (1998), no. 1, 93–141.

[EMM15] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi, *Isolation, equidistribution, and orbit closures for the* $\mathrm{SL}(2,\mathbb{R})$ *action on moduli space*, Ann. of Math. (2) **182** (2015), no. 2, 673–721. MR 3418528

[KKLM17] S. Kadyrov, D. Kleinbock, E. Lindenstrauss, and G. A. Margulis, *Singular systems of linear forms and non-escape of mass in the space of lattices*, J. Anal. Math. **133** (2017), 253–277. MR 3736492

[Kle10] Dmitry Kleinbock, *Quantitative nondivergence and its Diophantine applications*, Homogeneous flows, moduli spaces and arithmetic, Clay Math. Proc., vol. 10, Amer. Math. Soc., Providence, RI, 2010, pp. 131–153. MR 2648694

[KLW04] Dmitry Kleinbock, Elon Lindenstrauss, and Barak Weiss, *On fractal measures and Diophantine approximation*, Selecta Math. (N.S.) **10** (2004), no. 4, 479–523. MR 2134453

[KM98] D. Y. Kleinbock and G. A. Margulis, *Flows on homogeneous spaces and diophantine approximation on manifolds*, The Annals of Mathematics **148** (1998), no. 1, 339.

[KMW10] Dmitry Kleinbock, Gregory Margulis, and Junbo Wang, *Metric Diophantine approximation for systems of linear forms via dynamics*, Int. J. Number Theory **6** (2010), no. 5, 1139–1168. MR 2679461

[KW08] Dmitry Kleinbock and Barak Weiss, *Dirichlet's theorem on Diophantine approximation and homogeneous flows*, J. Mod. Dyn. **2** (2008), no. 1, 43–62. MR 2366229

[Sha09] Nimish A. Shah, *Equidistribution of expanding translates of curves and Dirichlet's theorem on Diophantine approximation*, Invent. Math. **177** (2009), no. 3, 509–532. MR 2534098

[Sha10] ———, *Expanding translates of curves and Dirichlet-Minkowski theorem on linear forms*, J. Amer. Math. Soc. **23** (2010), no. 2, 563–589. MR 2601043