Lattices, submanifolds and diophantine approximations
(Nicolas de Saxcé, Winter 2024)

Yuxiang Jiao
https://yuxiangjiao.github.io

Contents

§1 Classical results and general settings

Theorem 1.1

For every $\theta \in \mathbb{R} \setminus \mathbb{Q}$, there exists infinitely many $p/q \in \mathbb{Q}$ such that $|\theta - p/q| \leqslant 1/q^2$.

The first proof (continued fractions). Let $\theta_0 = \theta$ and $a_0 = \lfloor \theta_0 \rfloor$. For every $i \geqslant 1$, we define inductively that

$$\theta_i = \frac{1}{\theta_{i-1} - a_{i-1}}, \quad a_i = \lfloor \theta_i \rfloor.$$

We can check that

$$\theta = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \frac{1}{\theta_n}}}.$$

We have the following two facts:

1. $\begin{bmatrix} 1 \\ \theta \end{bmatrix} \mathbb{R} = \begin{bmatrix} 0 & 1 \\ 1 & a_0 \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 1 \\ \theta_n \end{bmatrix} \mathbb{R}.$
2. Let $p_n/q_n = a_0 + \frac{1}{\ddots + 1/a_n}$, then $\begin{bmatrix} q_n & q_{n-1} \\ p_n & p_{n-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & a_0 \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & a_n \end{bmatrix}.$

In particular, for every n, $\theta \in [p_n/q_n, p_{n+1}/q_{n+q}]$ (maybe reverse order). Then

$$\left| \theta - \frac{p_n}{q_n} \right| \leqslant \left| \frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}} \right| = \frac{1}{q_n q_{n+1}} \leqslant \frac{1}{q_n^2}.$$

□

Exercise 1.2. (1) Show that $q_{n+1} = a_{n+1}q_n + q_{n-1}$, and deduce that there are infinitely many n such that $q_{n+1} \geqslant \phi \cdot q_n$, where $\phi = (1+\sqrt{5})/2$.
(2) Conclude that there are infinitely many p_n/q_n such that $|\theta - p_n/q_n| \leqslant 1/(\sqrt{5}q_n^2)$.
(3) Check that the constant $\sqrt{5}$ is optimal.

The second proof (using Dirichlet's theorem).

Theorem 1.3 (Dirichlet)
For every $\theta \in \mathbb{R}$ and $Q \geqslant 1$, there exists $q \in \{1, \cdots, Q\}$ and $p \in \mathbb{Z}$ such that

$$\left|\theta - \frac{p}{q}\right| \leqslant \frac{1}{qQ} \leqslant \frac{1}{q^2}.$$

□

Definition 1.4. For $\theta \in \mathbb{R}$, we define its Diophantine exponent as

$$\beta(\theta) := \sup\left\{\beta > 0 : \exists p/q \text{ arbitrarily close to } \theta \text{ with } |\theta - p/q| \leqslant q^{-\beta}\right\}.$$

There are several basic properties:
(D) By Dirichlet's theorem, $\beta(\theta) \geqslant 2$ for every $\theta \in \mathbb{R}$.
(BC) By Borel-Cantelli lemma, $\beta(\theta) = 2$ for almost every $\theta \in \mathbb{R}$.
(R) Roth showed that $\beta(\theta) = 2$ for every $\theta \in \overline{\mathbb{Q}} \setminus \mathbb{Q}$.

Exercise 1.5 (Liouville)**.** Show that if $f(\theta) = 0$ for some $f \in \mathbb{Z}[X] \setminus \{0\}$, then $\beta(\theta) \leqslant (\deg f)$ if $f \notin \mathbb{Q}$.

Approximation in $\mathbb{R}^n$.

Let $\theta = \begin{bmatrix}\theta_1 & \cdots & \theta_n\end{bmatrix}^t \in \mathbb{R}^n$. We can consider several types of approximations:
- **Simultaneous approximations**: $|\theta_i - p_i/q| \leqslant q^{-\beta}$ for $i = 1, \cdots, n$.
- **Linear form approximations**: $|q - p_1\theta_1 - \cdots - p_n\theta_n| \leqslant q^{-\beta+1}$.

Here, the simultaneous approximation can also be considered as a **projective approximations**. Let $x = \mathbb{R}\begin{bmatrix}1 \\ \theta\end{bmatrix} \subset \mathbb{R}^d$, which is a point in $\mathbb{P}(\mathbb{R}^d)$. Let v be an element in $\mathbb{P}(\mathbb{Q}^d) \subset \mathbb{P}(\mathbb{R}^d)$. Then v is also a rational line in $\mathbb{R}^d$, which can be written as $\mathbb{R}\mathbf{v}$ for some primitive $\mathbf{v} = \begin{bmatrix}q & p_1 & \cdots & p_n\end{bmatrix}^t \in \mathbb{Z}^d$. The **height** of v is given by $H(v) := \|\mathbf{v}\|$. We want to study $d(x,v) \leqslant H(v)^{-\beta}$. Here the distance is understood in the projective space.

Theorem 1.6
(D) For every $x \in \mathbb{P}(\mathbb{R}^d)$, $\beta(x) \geqslant d/(d-1)$.
(BC) For almost every $x \in \mathbb{P}(\mathbb{R}^d)$, $\beta(x) = d/(d-1)$.
(R-S) For every $x \in \mathbb{P}(\overline{\mathbb{Q}}^d)$ not in any proper rational subspace, $\beta(x) = d/(d-1)$.

Exercise 1.7. Check (D) and (BC).

Theorem 1.8 (Subspace theorem, Schmidt, 1970s)

Let $L \in \mathrm{GL}(d, \overline{\mathbb{Q}})$ and write $L_1, \cdots, L_d$ for the rows of L. For every $\varepsilon > 0$, all solutions $\mathbf{v} \in \mathbb{Z}^d$ satisfying the inequality

$$|L_1(\mathbf{v}) \cdots L_d(\mathbf{v})| \leqslant \|\mathbf{v}\|^{-\varepsilon}$$

are contained in a finite union of $\mathbb{Q}$-hyperplanes.

Exercise 1.9. Check the theorem when $L \in \mathrm{GL}(d, \mathbb{Q})$.

Proof of (RS) assuming the subspace theorem. Write $x = \mathbb{R}[1\ \ \theta_2\ \ \cdots\ \ \theta_d]^t$ with $\theta_i \in \overline{\mathbb{Q}}$. Take

$$L = \begin{bmatrix} 1 & & & \\ -\theta_2 & 1 & & \\ \vdots & & \ddots & \\ -\theta_d & & \cdots & 1 \end{bmatrix}.$$

Assume that $d(x, v) \leqslant H(v)^{-\beta}$ for some $v \in \mathbb{P}(\mathbb{Q}^d)$. Take $\mathbf{v} \in \mathbb{Z}^d$ corresponding to v. Then $L_1(\mathbf{v}) = |q|$ and $L_i(\mathbf{v}) = |-q\theta_i + p_i|$ for $i \geqslant 2$. By the assumption, we have $L_i(\mathbf{v}) \leqslant \|\mathbf{v}\| H(v)^{-\beta}$ for every $i \geqslant 2$. Hence $|L_1(\mathbf{v}) \cdots L_d(\mathbf{v})| \leqslant \|v\|^{d-(d-1)\beta}$. If $d - (d-1)\beta > 0$ then $\mathbf{v}$ belongs to a finite union of $\mathbb{Q}$-hyperplanes $V_1 \cup \cdots \cup V_k$. But $x \notin \mathbb{P}(V_1 \cup \cdots \cup V_k)$, so $d(x, v)$ is bounded away from 0. There are only finitely many v with bounded height. A contradiction. ☐

Exercise 1.10. Prove (RS) for linear form approximations.

Approximation by linear subspaces.

Schmidt's question. Fix integers $q \leqslant k \leqslant \ell < d$. Given an ℓ-dimensional subspace $x \in \mathbb{R}^d$. Study k-dimensional rational subspace v lying close to x.

Definition 1.11 (distance). $d(v, x) := \max \{ d(\mathbf{u}, x) : \mathbf{u} \in v, \|\mathbf{u}\| = 1 \}$.

Notation 1.12. Denote X_ℓ to be the grassmannian variety of ℓ-dimensional subspaces in $\mathbb{R}^d$. Let $X_k(\mathbb{Q})$ to be the $\mathbb{Q}$-points in X_k (corresponding to $\mathbb{Q}$-subspaces).

Definition 1.13 (height). For every $v \in X_k(\mathbb{Q})$, the intersection $v \cap \mathbb{Z}^d$ is a subgroup of $\mathbb{Z}^d$, which can be written as $\mathbb{Z}v_1 \oplus \cdots \mathbb{Z}v_k$. The **height** of v is defined to be

$$H(v) := \mathrm{vol}(v_1 \wedge \cdots \wedge v_k) = \mathrm{vol}(v/(v \cap \mathbb{Z}^d)).$$

Proposition 1.14

There exists $C = C(d)$ such that $N_d(H) := \# \{ v \in X_k(\mathbb{Q}) : H(v) \leqslant H \}$ satisfies

$$C^{-1} H^d \leqslant N_d(H) \leqslant C H^d.$$

Exercise 1.15. Check this for $k = 1$ and $k = d - 1$.

Theorem 1.16
(D) For every $x \in X_\ell(\mathbb{R})$, $\beta_k(x) \geqslant \frac{d}{k(d-\ell)}$.
(BC) For almost every $x \in X_\ell(\mathbb{R})$, $\beta_k(x) = \frac{d}{k(d-\ell)}$.
(R) For every $x \in X_\ell(\overline{\mathbb{Q}})$ not contained in any proper rational pencil, $\beta_k(x) = \frac{d}{k(d-\ell)}$.

Definition 1.17. A **pencil** in X_ℓ is the a subset

$$\mathscr{P}_{W,r} := \{ x \in X_\ell(\mathbb{R}) : \dim x \cap W \geqslant r \},$$

where $W \subset \mathbb{R}^d$ is a rational subspace and $r \geqslant 1$.

Now we explain the intuition of this theorem. For every $v \in X_k(\mathbb{Q})$ and $\varepsilon > 0$. The set $\{ x \in X_\ell(\mathbb{R}) : d(v,x) \leqslant \varepsilon \}$ is an ε-neighborhood of $E_v = \{ x \in X_\ell(\mathbb{R}) : v \subset x \}$. Here E_v is a submanifold of $X_\ell(\mathbb{R})$ and $\dim E_v = (d-\ell)(\ell-k)$. Then $\operatorname{codim} E_v = k(d-\ell)$ and hence $\operatorname{vol}\{ x : d(v,x) \leqslant \varepsilon \} \asymp \varepsilon^{k(d-\ell)}$.

On the other hand, the number of $v \in X_k(\mathbb{Q})$ with $H(v) \leqslant H$ is approximately H^d. So that expected value for ε satisfies $H^d \varepsilon^{k(d-\ell)} = 1$. This gives $\varepsilon = H^{-\frac{d}{k(d-\ell)}}$.

Exercise 1.18. Use this argument to show that $\beta(x) \leqslant \frac{d}{k(d-\ell)}$ for almost every x.

§2 The correspondence between lattices and subspaces

Lattices in $\mathbb{R}^d$

Proposition 2.1

If Λ is a discrete subgroup of $\mathbb{R}^d$, then there exists k linearly independent vectors $v_1, \cdots, v_k$ such that $\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_k$

Proof. Take $v_1 \in \Lambda$ with minimal norm. Consider $P_{v_1^\perp}(\Lambda)$, which is a discrete subgroup of $v_1^\perp$ since v_1 is the shortest vector. By induction, we may write

$$P_{v_1^\perp}(\Lambda) = \mathbb{Z}P_{v_1^\perp}(v_2) \oplus \cdots \oplus \mathbb{Z}P_{v_1^\perp}(v_k).$$

Then $\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_k$. □

Definition 2.2. A **lattice** in $\mathbb{R}^d$ is a discrete subgroup of rank d. We can write $\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d$ with (v_i) a basis of $\mathbb{R}^d$ in this case.

Definition 2.3. The **first minimum** of a lattice is $\lambda_1(\Lambda) := \min \{ \|v\| : v \in \Lambda \setminus \{0\} \}$. The **co-volum**e of Λ is $\operatorname{covol}\Lambda = \operatorname{vol}(v_1 \wedge \cdots \wedge v_k)$, where $v_1, \cdots, v_k$ is given above.

Theorem 2.4 (Minkowski I)

Let Δ be a lattice in $\mathbb{R}^d$. If C is a convex symmetric set in $\mathbb{R}^d$ with $\operatorname{vol} C > 2^d \operatorname{covol}\Delta$, then $C \cap \Delta \neq \{0\}$. In particular, $\lambda_1(\Delta)^d \leqslant \frac{2^d}{\operatorname{vol} B(0,1)} \operatorname{covol}\Delta$.

Proof. Consider $\Delta_q = \frac{1}{q}\Delta$ for $q \in \mathbb{N}_+$. The number of points in $\Delta_q \cap \frac{C}{2}$ is approximately $q^d \frac{\operatorname{vol}(C)}{2^d \operatorname{covol}\Delta}$. If $\operatorname{vol} C > 2^d \operatorname{covol}\Delta$, for q large enough, there exists $v_1, v_2 \in \Delta_1 \cap \frac{C}{2}$ with the same image in $\Delta_q/\Delta \cong (\mathbb{Z}/q\mathbb{Z})^d$. Then $0 \neq v_1 - v_2 \in \Delta \cap C$. □

Definition 2.5. The **successive minima** of Δ is $\lambda_1(\Delta) \leqslant \cdots \leqslant \lambda_d(\Delta)$, where

$$\lambda_i(\Delta) := \inf \{ \lambda > 0 : \Delta \cap B(0, \lambda) \text{ contains } i \text{ linearly independent vectors} \} .$$

Theorem 2.6 (Minkowski II)

$\operatorname{covol}\Delta \leqslant \lambda_1(\Delta) \cdots \lambda_d(\Delta) \leqslant \frac{2^d}{\operatorname{vol} B(0,1)} \operatorname{covol}\Delta$.

Proof. If $v_1, \cdots, v_d$ are linearly independent with $\|v_i\| = \lambda_i$, then $\Delta' = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d < \Delta$. Hence $\lambda_1 \cdots \lambda_d \geqslant \operatorname{covol}(\Delta') \geqslant \operatorname{covol}(\Delta)$.

For the converse, we first construct an orthogonal basis $u_1, \cdots, u_d$ satisfying

$$\operatorname{span}\{ u_1, \cdots, u_i \} = \operatorname{span}\{ v_1, \cdots, v_i \}, \quad \forall 1 \leqslant i \leqslant d.$$

Let $T : u_i \mapsto \lambda_i^{-1}(\Delta) u_i$. We denote $\Delta_T = T\Delta$.

Claim 2.7. $\lambda_1(\Delta_T) \geqslant 1$.

Proof. Indeed, for every $v \in \Delta$, write $v = \sum_{i=1}^{I} \alpha_i v_i$ with $\alpha_I \neq 0$. Since v is linearly independent with $(v_1, \cdots, v_{I-1})$, $\|v\| \geqslant \lambda_I(\Delta)$. Therefore,

$$\|Tv\| \geqslant \frac{\|v\|}{\|T^{-1}|_{\operatorname{span}\{ v_1, \cdots, v_I \}}\|} = \frac{\|v\|}{\|T^{-1}|_{\operatorname{span}\{ u_1, \cdots, u_I \}}\|} \geqslant \frac{\lambda_I(\Delta)}{\lambda_I(\Delta)} = 1.$$

□

Now we apply Minkowski I to Δ_T, we obtain

$$1 \leqslant \frac{2^d \operatorname{covol}\Delta_T}{\operatorname{vol} B(0,1)} = \frac{2^d \operatorname{covol}\Delta}{\lambda_1(\Delta) \cdots \lambda_d(\Delta) \operatorname{vol} B(0,1)}.$$

□

Remark 2.8 We proved this theorem for euclidean norm above. But it is true in general for any norm with

$$\frac{\operatorname{covol}\Delta}{d!} \leqslant \lambda_1(\Delta) \cdots \lambda_d(\Delta) \leqslant \frac{2^d}{\operatorname{vol} B(0,1)} \operatorname{covol}\Delta.$$

Dani's correspondence

Let $x \in X_\ell(\mathbb{R})$. We want to study the diophantine exponent $\beta_k(x)$. Let $G = \operatorname{SL}(d, \mathbb{R})$ and $P = \operatorname{Stab}_G(x_0)$ where $x_0 = \operatorname{span}\{ e_1, \cdots, e_\ell \} \in X_\ell(X)$. Then G acts transitively on $X_\ell(\mathbb{R}) \cong P\backslash G$, here the isomorphism is given by $gx_0 \mapsto Pg^{-1}$.

Notation 2.9. For $x \in X_\ell(\mathbb{R})$, let $u_x \in G$ be such that $x = Pu_x$ (hence $u_x x = x_0$).

The **zooming flow** is given by

$$a_t = \begin{bmatrix} e^{-t/\ell} & & & & & \\ & \ddots & & & & \\ & & e^{-t/\ell} & & & \\ & & & e^{t/(d-\ell)} & & \\ & & & & \ddots & \\ & & & & & e^{t/(d-\ell)} \end{bmatrix}, \quad t \in \mathbb{R}.$$

Proposition 2.10 (Dani's correspondence, version 1)
For $x \in X_\ell(\mathbb{R})$, let $\Delta_x = u_x\mathbb{Z}^d$ be the lattice in $\mathbb{R}^d$. Let

$$\gamma_1(x) := \limsup_{t\in+\infty} -\frac{1}{t}\log\lambda_1(a_t\Delta_x).$$

Then

$$\beta_1(x) = \frac{d}{(d-\ell)(1-\ell\gamma_1(x))}.$$

Applications.

(1) Lower bound on β. Minkowski's first theorem shows that $\lambda_1(a_t\Delta_x) \lesssim 1$. Hence $\gamma_1(x) \geqslant 0$ and $\beta_1(x) \geqslant \frac{d}{d-\ell}$.

(2) Let Ω be the space of unimodular lattices in $\mathbb{R}^d$. Then $\Omega \cong \mathrm{SL}(d,\mathbb{R})/\mathrm{SL}(d,\mathbb{Z}) = G/\Gamma$ and it admits a finite G-invariant measure m_Ω. For $f \in C_c(\mathbb{R}^d)$, we define

$$\widetilde{f}(\Delta) := \sum_{\text{primitive } v\in\Delta} f(v).$$

Then $\int_\Omega \widetilde{f}\,\mathrm{d}m_\Omega = \int_{\mathbb{R}^d} f$. Take $f = \mathbb{1}_{B(0,\varepsilon)}$, then $\widetilde{f}(\Delta) \geqslant \mathbb{1}_{\lambda_1(\Delta)\leqslant\varepsilon}$. Therefore,

$$m_\Omega(\{\,\lambda_1 \leqslant \varepsilon\,\}) \leqslant \int \widetilde{f} = \int f \lesssim \varepsilon^d.$$

Claim 2.11. For almost every $\Delta \in \Omega$, $\lim_{t\to+\infty} \frac{1}{t}\log\lambda_1(a_t\Delta) = 0$.

Proof. For every $\varepsilon > 0$, we aim to show $\lambda_1(a_t\Delta) \geqslant e^{-\varepsilon t}$ for t large enough. It is enough to check for $t \in \mathbb{N}$. Note that

$$|\,\{\,\Delta : \lambda_1(a_t\Delta) \leqslant e^{-\varepsilon t}\,\}\,| = |\,\{\,\Delta : \lambda_1(\Delta) \leqslant e^{-\varepsilon t}\,\}\,| \lesssim \varepsilon^{-d\varepsilon t}.$$

By Borel-Cantelli lemma, we have $\limsup -\frac{1}{t}\log\lambda_1(a_t\Delta) \leqslant 0$. Hence the limit is 0 because $\lambda_1(a_t\Delta) \leqslant 1$ for every t. This implies that $\lambda_1(x) = 0$ for almost every x. $\square$

Exterior powers. For $0 \leqslant k \leqslant d$, the exterior power $\wedge^k\mathbb{R}^d$ is a vector space with basis e_I where $I \subset \{\,1,\cdots,d\,\}$ and $\#I = k$. If $I = \{\,i_1 < \cdots < i_k\,\}$ then $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$.

Exercise 2.12. If $\wedge^k\mathbb{R}^d$ is endowed with the euclidean structure making e_I an orthonormal basis, then, for $W = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_k$ a discrete subgroup of $\mathbb{R}^d$, we have $|W| = \|v_1 \wedge \cdots \wedge v_k\|$, where $|W|$ denotes the covolume of W in its real span.

Note that a_t acts on $\wedge^k\mathbb{R}^d$ with eigenvalues $\exp(-(\frac{k}{\ell}-i\frac{d}{\ell(d-\ell)})t)$ for $0\leqslant i\leqslant k$. An element e_I is an eigenvector corresponding to the eigenvalue $\exp(-(\frac{k}{\ell}-i\frac{d}{\ell(d-\ell)})t)$ if and only if $\#(I\setminus\{1,\cdots,\ell\})=i$. We write $\pi_+:\wedge^k\mathbb{R}^d\to\wedge^k\mathbb{R}^d$ to be the projection to the eigenspace with the eigenvalue $e^{-kt/\ell}$ (parallel to other eigenspaces).

Proposition 2.13 (Dani's correspondence, version 2)
For $x\in X_\ell(\mathbb{R})$, let

$$\gamma_k(x):=\sup\left\{\gamma\in\mathbb{R}:\begin{array}{c}\exists t>0\text{ large},\ \exists w\in a_t\wedge^k u_x\mathbb{Z}^d\text{ with}\\ \|w\|\leqslant e^{-\gamma t},\ \|\pi_+ w\|\geqslant\frac{1}{2}\|w\|\end{array}\right\}.$$

Then

$$\beta_k(x)=\frac{d}{(d-\ell)(k-\ell\gamma_k(x))}.$$

Proof. Assume $\beta<\beta_k(x)$, then there exists $v\in X_k(\mathbb{Q})$ close to x with $d(v,x)\leqslant H(v)^{-\beta}$. Take $\mathbf{v}\in\wedge^k\mathbb{Z}^d$ representing v. We want to make $\|a_tu_x\mathbf{v}\|$ small. We write $u_x\mathbf{v}=\mathbf{v}^{(0)}+\mathbf{v}^{(1)}+\cdots$ such that $a_t\mathbf{v}^{(i)}=\exp(-(\frac{k}{\ell}-i\frac{d}{\ell(d-\ell)})t)\mathbf{v}^{(i)}$.

Lemma 2.14
If v is close to x, then $\|\mathbf{v}^{(0)}\|\asymp H(v)$, $\|\mathbf{v}^{(1)}\|\asymp H(v)d(v,x)$ and $\|\mathbf{v}^{(p)}\|\lesssim H(v)d(v,x)^p$ for every $p\geqslant 2$.

Proof. Fix x and so does u_x. Then $H(v)=\|\mathbf{v}\|\asymp\max_i\|\mathbf{v}^{(i)}\|$. Note that $d(v,x)=d_{X_k}(v,E_x)$ where $E_x=\{y\in X_k:y\subset x\}$. We have

$$d(v,x)\asymp d(u_xv,u_xx)=d_{X_k}(u_xv,E_{x_0})\asymp d(\frac{u_x\mathbf{v}}{\|u_x\mathbf{v}\|},\wedge^k\operatorname{span}\{e_1,\cdots,e_\ell\}).$$

Note that $\wedge^kE_\ell$ is exactly the eigenspace of a_t with the eigenvalue $e^{-kt/\ell}$. Therefore,

$$d(v,x)\asymp\frac{1}{\|u_x\mathbf{v}\|}\max_{i\geqslant1}\|\mathbf{v}^{(i)}\|\asymp\frac{1}{H(v)}\max_{i\geqslant1}\|\mathbf{v}^{(i)}\|.$$

If $d(v,x)$ is small enough, then $\max_{i\geqslant1}\|\mathbf{v}^{(i)}\|$ is much smaller than $H(v)\asymp\max_{i\geqslant0}\|\mathbf{v}^i\|$. Therefore, $\mathbf{v}^{(0)}$ is the main term and $\|\mathbf{v}^{(0)}\|\asymp H(v)$.

Besides, we also obtain $\max_{i\geqslant1}\|\mathbf{v}^{(i)}\|\lesssim H(v)d(v,x)$. Now we demonstrate the remaining two estimates. For simplicity, we assume that $k=\ell$. After some appropriate rotations, we may assume that $\pi_+(u_x\mathbf{v})$ is parallel to $e_1\wedge\cdots\wedge e_\ell$. Then we write (**we cheat here**)

$$\frac{u_x\mathbf{v}}{\|u_x\mathbf{v}\|}=\begin{bmatrix}\mathrm{id}&0\\(u_{ij})&\mathrm{id}\end{bmatrix}(e_1\wedge\cdots\wedge e_\ell)$$

with $u_{ij}\in\mathbb{R}$ small. So we have $d(v,x)\asymp\max_{i,j}|u_{ij}|$. But then

$$\frac{\mathbf{v}^{(1)}}{\|\mathbf{v}^{(0)}\|}=\sum\pm u_{ij}\cdot e_{\{1,\cdots,\ell\}\setminus\{j\}\cup\{i\}}$$

is with norm $\asymp\max|u_{ij}|\asymp d(v,x)$. For $p\geqslant2$, we can find that $\|\mathbf{v}^{(p)}\|/\|\mathbf{v}^{(0)}\|$ is a homogeneous polynomial of deg p, so we have $\|\mathbf{v}^{(p)}\|\lesssim\|\mathbf{v}^{(0)}\|(\max|u_{ij}|)^p\asymp H(v)d(v,x)^p$. $\square$

So we have
$$\|a_t u_x \mathbf{v}\| \asymp H(v) \max\left\{ e^{-\frac{kt}{\ell}}, e^{-(\frac{k}{\ell}-\frac{d}{\ell(d-\ell)})t} d(x,v), \cdots \right\}.$$
Take $t > 0$ so that $e^{\frac{dt}{\ell(d-\ell)}} = H(v)^\beta$. Then
$$\|a_t u_x \mathbf{v}\| \lesssim H(v) e^{-\frac{kt}{\ell}} = e^{-(\frac{k}{\ell}-\frac{1}{\beta}\frac{d}{\ell(d-\ell)})t}.$$
Thus $\gamma_k(x) \geqslant \frac{k}{\ell} - \frac{1}{\beta}\frac{d}{\ell(d-\ell)}$.

For the converse direction, assume that $\|a_t u_x \mathbf{v}\| \leqslant e^{-\gamma t}$ and $\|\pi_+(a_t u_x \mathbf{v})\| \gtrsim \|a_t u_x \mathbf{v}\|$. Using the above computation, this yields:
$$e^{-\gamma t} \gtrsim H(v) \max\left\{ e^{-\frac{kt}{\ell}}, e^{-(\frac{k}{\ell}-\frac{d}{\ell(d-\ell)})t} d(x,v) \right\} \quad \text{and} \quad e^{-(\frac{k}{\ell}-\frac{d}{\ell(d-\ell)})t} \|\mathbf{v}^{(1)}\| \lesssim e^{-\frac{kt}{\ell}} \|\mathbf{v}^{(0)}\|.$$
Therefore, $H(v) \lesssim e^{(\frac{k}{\ell}-\gamma)t}$ and $d(x,v) \lesssim H(v)^{-\frac{d}{(d-\ell)(k-\ell\gamma)}}$. □

During the proof of Lemma 2.14, we assume implicitly that $\mathbf{v}$ was decomposable. That is $\mathbf{v} = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$ for some $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{R}^d$. This is always possible thanks to the following lemma:

Lemma 2.15 (Mahler)

If Δ is a lattice in $\mathbb{R}^d$, then the successive minima of $\wedge^k \Delta$ are essentially (up to a multiplicative constant) equal to the
$$\lambda_I(\Delta) = \lambda_{i_1}(\Delta) \cdots \lambda_{i_k}(\Delta), \quad I \subset \{1, \cdots, \ell\}, \#I = k,$$
and achieved by decomposable vectors.

Proof. Assume Δ is unimodular and hence so is $\wedge^k \Delta$. If $\|v_i\| = \lambda_i(\Delta)$ with $v_1, \cdots, v_d$ linearly independent, then $v_I = v_{i_1} \wedge \cdots \wedge v_{i_k}$ satisfies $\|v_I\| \leqslant \lambda_I(\Delta)$. But by Minkowski II, $\prod_I \lambda_I(\Delta) \lesssim 1$ and hence $\|v_I\| \asymp \lambda_I(\Delta)$ for each I. □

Going back to the correspondence, if there exits $w \in \wedge^k a_t u_x \mathbb{Z}^d$ with $\|w\| \leqslant e^{-\gamma t}$ (i.e. $\lambda_1(\wedge^k a_t u_x \mathbb{Z}^d) \leqslant e^{-\gamma t}$) and $\|\pi_+(w)\| \gtrsim \|w\|$, then we can find such w with that is decomposable.

§3 Algebraic subspaces

Grayson polygon and Harder-Narasimhan filtration.

Let Δ be a lattice in $\mathbb{R}^d$, let $\mu_i(\Delta) = \min\{ |V| : V < \Delta, V \cong \mathbb{Z}^i \}$ be the successive covolumes of Δ.

Definition 3.1. The **Grayson polygon** C_Δ is the maximal convex function on $[0,d]$ whose graph has below each point $(i, \log \mu_i(\Delta))$.

Proposition 3.2 (Harder-Narasimhan filtration)

If C_Δ has angle at the point i then there exists $V_i < \Delta$ of rank i with $|V_i| = \log \mu_i(\Delta)$. Moreover, if $I = \{ i_1 < \cdots < i_k \}$ is the set of angle points then
$$\{0\} < V_{i_1} < \cdots < V_{i_k} < \Delta.$$

Definition 3.3. Let $\mathbb{K}$ be a field with characteristic 0. A map $\tau : \mathrm{Gr}(\mathbb{K}^d) \to \mathbb{R}$ is **submodular** if

$$\tau(V \cap W) + \tau(V + W) \leqslant \tau(V) + \tau(W), \quad \forall V, W \subset \mathbb{K}^d.$$

Example 3.4

If Δ is a lattice in $\mathbb{R}^d$ then $\{$ primitive subgroups of $\Delta \} \hookrightarrow \mathrm{Gr}(\mathbb{Q}^d)$. Then $\tau(V) = \log|V|$ is submodular, or equivalently $|V \cap W| \cdot |V + W| \leqslant |V| \cdot |W|$.

Exercise 3.5. Check this inequality.

Lemma 3.6 (Submodularity)

Let $\tau : \mathrm{Gr}(\mathbb{K}^d) \to \mathbb{R}$ be submodular with $\tau(\{0\}) = 0$. Then there exists a unique maximal subspace with

$$\frac{\tau(V)}{\dim V} = \inf\left\{ \frac{\tau(W)}{\dim W} : W \subset \mathbb{K}^d \right\}.$$

Proof. Assume for simplicity that V, W both attain the infimum a. Then

$$\tau(V + W) \leqslant a(\dim V + \dim W) - a\dim(V \cap W) = a\dim(V + W).$$

This proves the lemma. □

Theorem 3.7

If $\tau : \mathrm{Gr}(\mathbb{K}^d) \to \mathbb{R}$ is submodular with $\tau(0) = 0$. Define its Grayson polygon C_τ as the maximal convex function on $[0, d]$ lying below all points $(\dim W, \tau(W))$. If C_τ has angle at i, then there is a unique V_i such that $\dim V_i = i$ and $C_\tau(i)$, and if $I = \{i_1 < \cdots < i_k\}$ is the set of angle points for C_τ then we have a HN-filtration

$$\{0\} < V_{i_1} < \cdots < V_{i_k} < \mathbb{K}^d.$$

Remark 3.8 By Minkowski II, $\mu_i(\Delta) \asymp \lambda_1(\Delta) \cdots \lambda_i(\Delta)$. So the shapes of C_Δ are (up to a additive constant) equal to $(\log \lambda_1(\Delta), \cdots, \log \lambda_d(\Delta))$.

Parametric subspace theorem.

Aim 3.9. Given $\Delta \subset \mathbb{R}^d$ a lattice, describe $C_{a_t\Delta}$ for $t > 0$, where $a_t = \mathrm{diag}(e^{\alpha_1 t}, \cdots, e^{\alpha_d t})$.

Theorem 3.10 (Parametric subspace theorem)
Assume that $\Delta = L\mathbb{Z}^d$ with $L \in \mathrm{GL}(d, \overline{\mathbb{Q}})$. Then there exists C_∞ such that

$$\lim_{t\to+\infty} \frac{1}{t} C_{a_t\Delta} = C_\infty.$$

Moreover, if $I = \{i_1 < \cdots < i_k\}$ are the angles of C_∞ then there exists a filtration $\{0\} < V_{i_1} < \cdots < V_{i_k} < \mathbb{R}^d$ such that for every $t > 0$ large enough and for every s, $a_t L V_{i_s}$ contains the first i_s successive minima of $a_t L\mathbb{Z}^d$.

§4 Rational approximation to linear subspaces

Definition 4.1. For $W < \mathbb{R}^d$, the **expansion rate** of W under the flow $a_t L$ is

$$\tau_L(W) := \lim_{t\to+\infty} \frac{1}{t} \log \|a_t L w\|,$$

where $w \in \wedge^{\dim W}\mathbb{R}^d$ represents W.

Remark 4.2 $\tau_L(W)$ is the logarithm of the largest eigenvalue occurring in the decomposition of Lw along the eigenspaces of a_t in $\wedge^{\dim W}\mathbb{R}^d$.

Remark 4.3 If Λ_W is a lattice in W, then $|a_t L\Lambda_W| \asymp e^{\tau_L(W)}|\Lambda_W|$.

Exercise 4.4. $\tau_L : \mathrm{Gr}(\mathbb{Q}^d) \to \mathbb{R}$ is submodular.

Theorem 4.5 (Precision on the parametric subspace theorem.)
C_∞ is the Grayson polygon associated to τ_L and the HN filtration also corresponds.

Proof. V_{i_1} minimizes the rate $\frac{\tau_L(V_{i_1})}{i_1} = \min_V \frac{\tau_L(V)}{\dim V}$ and any V satisfying $\frac{\tau_L(V_{i_1})}{i_1} = \frac{\tau_L(V)}{\dim V}$ is a subspace of V_{i_1}. Observe that $|a_t L V_{i_1}(\mathbb{Z})| \asymp e^{t\tau_L(V_{i_1})}|V_{i_1}(\mathbb{Z})|$. So by Minkowski's first theorem, there exists $v \in a_t L V_{i_1}(\mathbb{Z})$ with $\|v\| \lesssim e^{t\frac{\tau_L(V_{i_1})}{i_1}}$. This shows that for every $t > 0$ large, $\lambda_1(a_t L\mathbb{Z}^d) \lesssim e^{t\frac{\tau_L(V_{i_1})}{i_1}}$. So we have $\frac{1}{t}\log \mu_{i_1}(a_t L\mathbb{Z}^d) \leqslant \tau_L(V_{i_1}) + o(1)$.

To check that $\frac{1}{t}C_t \to C_\infty$ on $[0, i_1]$, all we need to show is that

$$\lambda_1(a_t L\mathbb{Z}^d) \geqslant e^{t(\frac{\tau_L(V_{i_1})}{i_1} - \varepsilon)}$$

for every $\varepsilon > 0$ and $t > 0$ large enough. Let $V \leqslant \mathbb{Q}^d$ of minimal dimension such that there exists arbitrarily large t with $v \in V(\mathbb{Z})$ satisfying $\|a_t L v\| \leqslant e^{t(\frac{\tau_L(V_{i_1})}{i_1} - \varepsilon)}$. Let $k = \dim V$. We apply the subspace theorem.

Let $L_1, \cdots, L_d$ be the rows of L. Let j_1 be minimal such that $L_{j_1}|_V \neq 0$. We then find $j_1, \cdots, j_k$ inductively such that $L_{j_1}|_V, \cdots, L_{j_k}|_V$ are linearly independent. Then $\tau_L(V) = A_{j_1} + \cdots + A_{j_k}$.

We have

$$\begin{aligned}\|L_{j_1}(v)\cdots L_{j_k}(v)\| \leqslant & e^{-\tau_L(V)t}\prod_{s=1}^{k}\left|e^{A_{j_s}}tL_{j_s}(v)\right| \\ \leqslant & e^{\tau_L(V)t}\prod_{s=1}^{k}\|a_tLv\| \leqslant e^{\tau_L(V)t}e^{kt(\frac{\tau_L(V_{i_1})}{i_1}-\varepsilon)} \\ \leqslant & e^{-kt(\varepsilon-o(1))} \leqslant \|v\|^{-\varepsilon'}.\end{aligned}$$

So all such v must belong to a finite union of proper subspaces of V. By the minimality of V, there can be such solutions only for bounded t. Hence we obtain that $\frac{1}{t}C_{a_t\Delta} \to \tau_L$ on $[0, i_1]$. Then we apply an induction and we are done. □

Application to rational approximation to linear subspaces.

Let $x \in X_\ell(\overline{\mathbb{Q}})$ and $u_x \in \mathrm{SL}(d, \overline{\mathbb{Q}})$ such that $x = Pu_x$ ($x = u_x^{-1}\operatorname{span}\{e_1, \cdots, e_\ell\}$). We want to understand the successive minima of $a_tu_x\mathbb{Z}^d$. For $W \leqslant \mathbb{Z}^d$, write $\tau_x(W) = \tau_{u_x}(W)$. Then

$$\tau_x(W) = -\frac{\dim x\cap W}{\ell} + \frac{\dim W - \dim x\cap W}{d-\ell}.$$

So

$$\frac{\tau_x(W)}{\dim W} = \frac{1}{d-\ell} - \frac{\dim x\cap W}{\dim W}\cdot\frac{d}{\ell(d-\ell)}.$$

To minimize this, one has to maximize $\frac{\dim x\cap W}{\dim W}$.

Example 4.6

V_{i_1} is the unique subspace such that $\frac{\dim x\cap V_{i_1}}{\dim V_{i_1}} = \max_{W\leqslant\mathbb{Q}^d}\frac{\dim x\cap W}{\dim W}$.

Recall that a pencil for $W \subset \mathbb{Q}^d$ and $r \geqslant 1$ is

$$\mathscr{P}_{W,r} = \{\, x \in X_\ell(\mathbb{R}) : \dim x\cap W \geqslant r \,\}.$$

We say the pencil is **constraining** if $\frac{r}{\dim W} > \frac{\ell}{d}$.

Corollary 4.7

If $x \in X_\ell(\overline{\mathbb{Q}})$ is not in any constraining rational pencil, then $\beta_k(x) = \frac{d}{k(d-\ell)}$.

Proof. By the example above, $V_{i_1} = \mathbb{Q}^d$. So the filtration is trivial and $C_\infty = 0$. Hence for every $i = 1, \cdots, d$, $\lambda_i(a_tu_x\mathbb{Z}^d) = e^{o(t)}$. But recall that the successive minima of $\wedge^k a_tu_x\mathbb{Z}^d$ are essentially the $\lambda_I = \lambda_{i_1}\cdots\lambda_{i_k} = e^{o(t)}$. So $\wedge^k a_tu_x\mathbb{Z}^d$ has a nice basis consisting of vectors of length $e^{o(t)}$. One of them must satisfy $\|\pi_+(w)\| \gtrsim \|w\|$ so $\gamma_k(x) \geqslant 0$. Hence we obtain $\beta_k(x) \geqslant \frac{d}{k(d-\ell)}$. But we also know that $\wedge^k a_tu_x\mathbb{Z}^d$ contains no vector of norm less than $e^{\varepsilon t}$, so $\gamma_k(x) \leqslant 0$ and hence $\beta_k(x) = \frac{d}{k(d-\ell)}$. □

For general cases, V_{i_1} is the maximal maximizing $\frac{\dim x\cap V_{i_1}}{i_1} = \frac{\ell_1}{i_1}$; V_{i_2} is maximal maximizing $\frac{\dim x\cap V_{i_2} - \dim x\cap V_{i_1}}{i_2 - i_1} = \frac{\ell_2-\ell_1}{i_2-i_1}, \cdots$. To understand the successive minimas of $\wedge^k a_tu_x\mathbb{Z}^d$, we decompose

$$\wedge^k \mathbb{Q} = \bigoplus_{k_1 \leqslant k_2 \leqslant \cdots \leqslant k_s = k} \underbrace{\bigwedge^{k_1} V_{i_1} \wedge \bigwedge^{k_2 - k_1} (V_{i_2}/V_{i_1}) \wedge \cdots \wedge \bigwedge^{k_s - k_{s-1}} (V_{i_s}/V_{i_{s-1}})}_{\text{denoted by } W_{\underline{k}} = W_{k_1, \cdots, k_s}}.$$

The logarithm of the successive minmas in $a_t u_x W_{\underline{k}}$ are essentially equal to

$$\Lambda_{\underline{k}} = \frac{k}{d-\ell} - \frac{d}{\ell(d-\ell)} \left(\frac{k_1 \ell_1}{i_1} + \frac{(k_2 - k_1)(\ell_2 - \ell_1)}{i_2 - i_1} + \cdots + \frac{(k_s - k_{s-1})(\ell_s - \ell_{s-1})}{i_s - i_{s-1}} \right).$$

To minimizing $\Lambda_{\underline{k}}$, one should take $k_1 = i_1, k_2 = i_2, \cdots, k_s = \min\{ i_s, k \}$. But then, one might not have $\|\pi_+(w)\| \gtrsim \|w\|$. To ensure this, it is necessary to have $k_r \leqslant \ell_r$ for every r. Indeed, otherwise, $u_x W_{\underline{k}} \cap \wedge^k \operatorname{span}\{ e_1, \cdots, e_\ell \} = \{ 0 \}$. Then we have $\|u_x v - \pi_+ u_x v\| \geqslant c\|u_x v\|$. So

$$\|a_t u_x v\| \geqslant c^{-(\frac{k}{\ell} - \frac{d}{\ell(d-\ell)})t} \|u_x v\| \gtrsim e^{\frac{dt}{\ell(d-t)}} \|\pi_+(a_t u_x v)\|$$

for $v \in W_{\underline{k}}$. This is not as desired.

Best possible choice is therefore $k_r = \min\{ \ell_r, k \}$ for every i. Then we get the correct value. For example,

$$\gamma_\ell(x) = -\frac{\ell}{d-\ell} + \frac{d}{\ell(d-\ell)} \sum_{r=1}^{s} \frac{(\ell_r - \ell_{r-1})^2}{i_r - i_{r-1}}.$$

Finally, we can prove the first item in Theorem 1.16. It suffices to show that $\gamma_k \geqslant 0$. We consider a simpler case that $k = \ell$.

Proof. For $k = \ell$, by Cauchy-Schwartz, we have

$$d \sum_{r=1}^{s} \frac{(\ell_r - \ell_{s-1})^2}{i_r - i_{r-1}} \geqslant \left(\sum (\ell_r - \ell_{r-1}) \right) \geqslant \ell^2.$$

Hence $\gamma_\ell \geqslant 0$. □