Quantitative Recurrence of unipotent flows.

Monday, January 26, 2015
10:43 AM

In 70's Margulis proved any non-cocompact lattice in a higher rank simple Lie group, e.g. $SL_n(\mathbb{R})$ for $n \geq 3$, is arithmetic, i.e. it is more or less $\mathbb{G}(\mathbb{Z})$ for some algebraic $\mathbb{Q}$-group $\mathbb{G}$ s.t. $\mathbb{G} \simeq SL_n / \mathbb{R}$. His proof was based on recurrence of unipotent flows in $\Omega^{(1)}(\mathbb{R}^n)$.

Later Dani-Margulis and Kleinbock-Margulis gave quantitive versions of the non-divergence of polynomial maps.

(I am following D. Kleinbock's Clay notes.)

Before stating the main theorem, let me give a few definitions and lemmas.

<u>Def</u>. A subgroup Λ of $\mathbb{Z}^n$ is called a <u>primitive</u> subgroup if $\exists\, v_1, \dots, v_n \in \mathbb{Z}^n$ s.t.

$$(1)\ \Lambda = \bigoplus_{i=1}^{m} \mathbb{Z} v_i, \quad (2)\ \mathbb{Z}^n = \bigoplus_{i=1}^{n} \mathbb{Z} v_i.$$

<u>Lemma</u>. Suppose Λ is a discrete subgroup of $\mathbb{R}^n$.

m

And $\Lambda = \bigoplus_{i=1}^{\cdots} \mathbb{Z} v_i$. Then

$$\mathrm{vol}\left(\left\{\sum_{i=1}^{m} a_i v_i \;\middle|\; |a_i| \le 1/2\right\}\right)^2$$

is independent of the choice of $v_1, \dots, v_m$. It is denoted by $d(\Lambda)$. Moreover, if $\Lambda \subseteq \mathbb{Z}^n$ is a subgroup, then $d(\Lambda) \in \mathbb{Z}^+$. If Λ_1 is a finite-index subgroup of Λ_2, then $d(\Lambda_1) = [\Lambda_2 : \Lambda_1]^2 \, d(\Lambda_2)$. And $d(\Lambda) = \mathrm{vol}\left(\mathbb{R}\text{-span}\,\Lambda / \Lambda\right)^2$.

<u>Pf.</u> Let $v_1', \dots, v_m'$ be another $\mathbb{Z}$-basis of Λ. So

$$\exists\, \gamma \in GL_m(\mathbb{Z}) \text{ s.t. } v_i' = \sum_{j=1}^{m} \gamma_{ji} \, v_j$$

$$\Rightarrow [v_1' \cdots v_m'] = [v_1 \cdots v_m]\, \gamma$$

By Gram-Schmidt, $[v_1 \cdots v_m] = \underbrace{[u_1 \cdots u_m]}_{\text{orthonormal}} \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_m \end{bmatrix} \begin{bmatrix} 1 & & * \\ & \ddots & \\ & & 1 \end{bmatrix}$

$$\Rightarrow \mathrm{vol}(\mathcal{F}(v_1, \dots, v_m))^2 = \prod_{i=1}^{m} a_i^2 = \underset{n\times m}{\det}\left([v_1 \cdots v_m]^t [v_1 \cdots v_m]\right)$$

$$\Rightarrow \mathrm{vol}(\mathcal{F}(v_1', \dots, v_m'))^2 = \det\left(\gamma^t [v_1 \cdots v_m]^t [v_1 \cdots v_m] \gamma\right)$$

$$= \det(\gamma)^2 \, \mathrm{vol}(\mathcal{F}(v_1, \dots, v_m))^2$$

$$= \mathrm{vol}(\mathcal{F}(v_1,\dots,v_m))^2.$$

$\Longrightarrow$ it is independent of the choice of a $\mathbb{Z}$-basis.

- If $\Lambda \subseteq \mathbb{Z}^n$, then $\Lambda = \bigoplus_{i=1}^{m} \mathbb{Z}\, v_i$ for some $v_i \in \mathbb{Z}^n$.

$\Rightarrow d(\Lambda) = \det([v_1 \cdots v_m]^t [v_1 \cdots v_m]) \in \mathbb{Z}^+$.

- By Weil's formula, $\mathrm{vol}(\mathcal{F}(v_1,\dots,v_m)) = \displaystyle\int_{\Lambda_{\mathbb{R}}/\Lambda} I(\mathbb{1}_{\mathcal{F}})([x])\, d[x]$

$$= \mathrm{vol}(\Lambda_{\mathbb{R}}/\Lambda).$$

$\Rightarrow d(\Lambda) = \mathrm{vol}(\Lambda_{\mathbb{R}}/\Lambda)^2.$

- $\Lambda_1 \subseteq \Lambda_2$ finite index $\Longrightarrow$ $\Lambda_{1,\mathbb{R}}/\Lambda_1 \twoheadrightarrow \Lambda_{2,\mathbb{R}}/\Lambda_2$ is a degree $[\Lambda_2 : \Lambda_1]$ covering map.

$(\Lambda_{1,\mathbb{R}} = \Lambda_{2,\mathbb{R}})$

$\Longrightarrow \mathrm{vol}(\Lambda_{1,\mathbb{R}}/\Lambda_1) = [\Lambda_2 : \Lambda_1]\, \mathrm{vol}(\Lambda_{2,\mathbb{R}}/\Lambda_2)$. $\blacksquare$

<u>Def.</u> $\{u_t\}_{t\in\mathbb{R}} \subseteq GL_m(\mathbb{R})$ is called a <u>unipotent flow</u> if

(a) $\forall t_1, t_2$, $u_{t_1} \cdot u_{t_2} = u_{t_1+t_2}$,

(b) $\forall t$, u_t is unipotent

<u>Remark</u>. u is unipotent $\Longleftrightarrow$ $n = u - I$ is nilpotent

$\Rightarrow \log u = \log(I+n) = n - \frac{n^2}{2} + \frac{n^3}{3} - \cdots + (-1)^{m} \frac{n^{m-1}}{m-1}$

$=: \underline{u}$ is (well-defined and) nilpotent.

$\Rightarrow$ If $\{u_t\}$ is a (real) unipotent flow, then

$$u_t = \exp(t\,\underline{u}) \quad (\text{where } \underline{u} = \log u_1 .)$$

$$= \sum_{i=1}^{m-1} (-1)^{i-1} t^i \underline{u}^i$$

__Def.__ A function $u : [a,b] \longrightarrow GL_m(\mathbb{R})$ is called a polynomial map if $(u(t))_{ij}$ is a poly. for any i,j.

Moreover $\deg u := \max_{i,j} (\deg u_{ij})$.

__Remark__. A unipotent flow $\{u_t\} \subseteq GL_m(\mathbb{R})$ is a poly. map of $\deg \leq m-1$.

__Lemma__. If $u : [a,b] \longrightarrow GL_m(\mathbb{R})$ is a degree $\underline{d}$ poly. map, then $t \longmapsto d(u(t)\Lambda_0)$ is a polynomial of degree at most $\underline{2md}$ for any discrete subgroup $\Lambda_0 \subseteq \mathbb{R}^m$.

__Pf__. $d(u(x)\Lambda_0) = \det\left[[v_1 \cdots v_k]^t\, u(x)^t\, u(x)\, [v_1 \cdots v_k] \right]$

where $\Lambda_0 = \bigoplus^{k} \mathbb{Z} v_i$. ∎

where $\Lambda_0 = \bigoplus_{i=1}^{k} \mathbb{Z} v_i$. $\blacksquare$

Theorem (Quantitative non-divergence)

(Dani–Margulis, Kleinbock–Margulis)

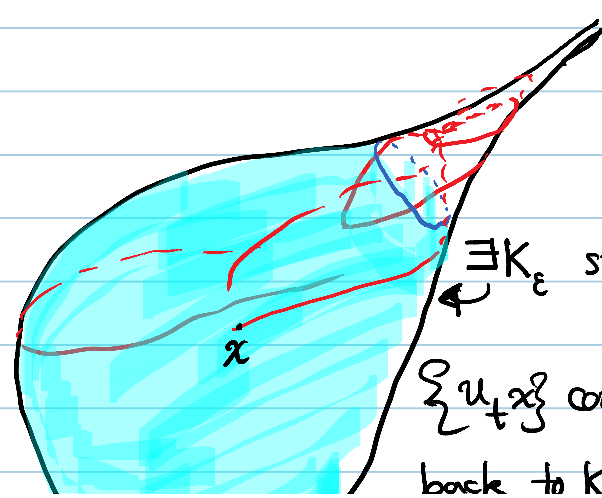
Suppose $\Lambda_0 \in \Omega^{(1)}(\mathbb{R}^n)$, $u : \mathbb{R} \to SL_m(\mathbb{R})$ is a degree d_0 polynomial map.

(*) Suppose $\exists \rho < 1$ s.t. for any <u>primitive subgroup</u> Δ of Λ_0 we have

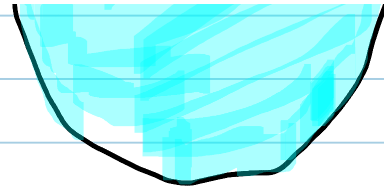
$$\max_{0 \le x \le T} d(u(x)\Delta) \ge \rho^{\operatorname{rank}(\Delta)} .$$

Then for any $0 < \varepsilon^2 < \rho$ we have

$$\ell(\{ t \in [0,T] \mid \delta(u_t \Lambda_0) < \varepsilon \}) \ll_{m,d_0} \left(\frac{\varepsilon}{\sqrt{\rho}}\right)^{1/d_0 m} .$$



$\exists K_\varepsilon$ s.t. $\{u_t x\}$ comes back to K.



$\{u_t x\}$ works back to K_ε. $1-\varepsilon$ portion of time.

This theorem is much easier to be proved in $\mathbb{R}^2$. Let's try to prove it in $\mathbb{R}^2$, and $u_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$.

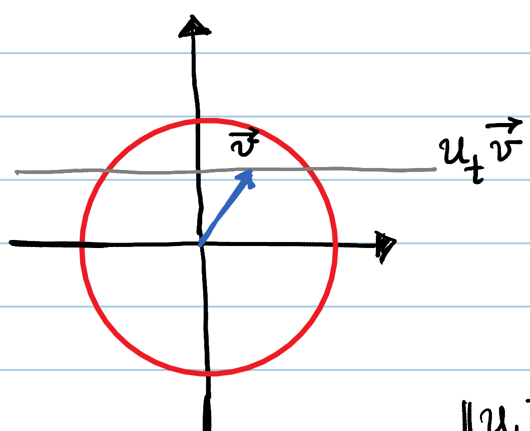
Let v_1 and v_2 be two linearly independent vectors in $\mathbb{R}^2$

$\Rightarrow \|\vec{v}_1 \times \vec{v}_2\| \geq 1 \Rightarrow \|\vec{v}_1\| \, \|\vec{v}_2\| \geq 1$

$\Rightarrow \underline{\exists i}, \ \|\vec{v}_i\| \geq 1$.

$\Rightarrow \Lambda \cap$ open ball of radius $1 \subseteq$ a real line.

So we need to understand what $\underline{u_t}$ does to a single "small" vector.

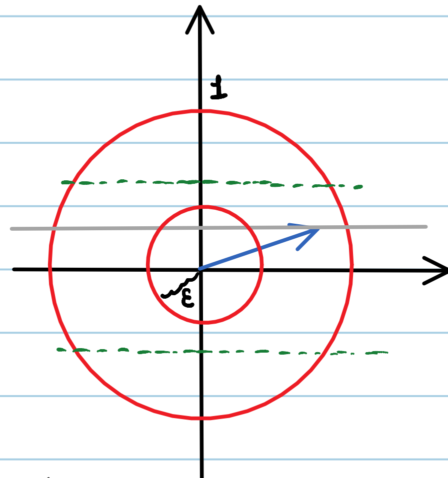


as long as $\underline{u_t\vec{v}}$ is in the ball of radius 1 we have

$\|u_t\vec{v}\| = \delta(u_t\Lambda)$ if

$\|\vec{v}\| = \delta(\Lambda)$.

Let's have a picture where $\delta(u_t\Lambda)$ is very small for some values of t.



So as you can see only $O(\varepsilon)$ portion of time $u_t\Lambda$ has a vector whose length $\leq \varepsilon$. So the interval $[0,T]$ is divided to $O(1)$ intervals where $\delta(u_t\Lambda) = \|u_t v\|$ for a fixed v. And inside this $O(1)$-sized interval there is a subinterval of size $O(\varepsilon)$ where $u_t\Lambda$ has a vector of size at most ε. [There are also shorter intervals where $\delta(u_t v)$ remains large, e.g. $\geq 1/2$.]

The above description is valid as long as $\vec{v}$ is NOT on the x-axis. If it is, then $u_t\vec{v} = \vec{v}$ so $\delta(u_t\Lambda)$ does NOT change. Indeed we get a closed orbit for u_t:

$$u_t \begin{bmatrix} \alpha & x \\ 0 & \alpha^{-1} \end{bmatrix} SL_2(\mathbb{Z}) = \begin{bmatrix} \alpha & x+t\alpha^{-1} \\ 0 & \alpha^{-1} \end{bmatrix} SL_2(\mathbb{Z})$$

$$= \begin{bmatrix} \alpha & \\ & \alpha^{-1} \end{bmatrix} \begin{bmatrix} 1 & \alpha^{-1}x + t\alpha^{-2} \\ 0 & 1 \end{bmatrix} SL_2(\mathbb{Z})$$

$\Rightarrow$ periodic, with period α^2.

In higher dimension, difficulty arises as we can have many "small vectors". As a result, it would be hard to keep track of potentially "bad" vectors and "catch them in the act".

Let's reiterate the main points of the 2x2 case:

① For $t \in [0, T]$, if $u_t g_0 \mathbb{Z}^2$ has a vector $u_t g_0 v_t$ s.t.

$$\varepsilon \ll \| u_t g_0 v_t \| \ll 1$$

Then $\varepsilon \leq \delta(u_t g_0 \mathbb{Z}^2) \rightsquigarrow$ we can "mark" t with the vector $v_t \in \mathbb{Z}^2$ and it "protects" us from having ε-small vectors.

② For a fixed vector v,

$$\| u_t g_0 v \| = \Theta(1) \Rightarrow \frac{|\{ t \in [0, T] \mid \| u_t g_0 v \| \ll \varepsilon \}|}{T} = O(\varepsilon).$$

③ These vectors "partition" $[0,T]$.

In higher dimension we cannot mark time with a single vector

i.e. knowing any kind of information on $\|u_t g_0 \vec{v}\|$

for a single vector $\vec{v} \in \mathbb{Z}^n$ cannot protect us from having

ε-small vectors in $u_t g_0 \mathbb{Z}^n$.

We instead work with partial flags. A few initial definitions

- Let $h(t) := u_t g_0$. Then $h(t)$ is also a degree d_0 polynomial map.

Def.: $\mathcal{P}(\mathbb{Z}^n) := \{\Delta \leq \mathbb{Z}^n \mid \Delta$ is a primitive subgroup$\}$.

- $\mathcal{F} = \{\Delta_1, \dots, \Delta_m\} \subseteq \mathcal{P}(\mathbb{Z}^n)$ is called a partial flag of length $m-1$ if $\Delta_1 \subsetneq \cdots \subsetneq \Delta_m$.

- Length of a subset $\mathcal{P}$ of $\mathcal{P}(\mathbb{Z}^n)$ is

$$\max\{m \mid \exists \text{ a partial flag of length } m \subseteq \mathcal{P}\}.$$

- $\Delta \in \mathcal{P}(\mathbb{Z}^n)$ is said to be compatible with a partial flag $\mathcal{F}$ if $\forall \Delta' \in \mathcal{F}$, either $\Delta \subseteq \Delta'$ or $\Delta' \subseteq \Delta$.

Example. We are particularly interested in subsets

$$\mathcal{P}_{\mathcal{F}} := \{\Delta \in \mathcal{P}(\mathbb{Z}^n) \mid \Delta \text{ is compatible with } \mathcal{F}, \ \Delta \notin \mathcal{F}\}.$$

where $\mathcal{F}$ is a partial flag.

$\mathcal{P}_{\emptyset} = \mathcal{P}(\mathbb{Z}^n)$; $\mathcal{P}_{\{0 \subseteq \langle e_1 \rangle \subseteq \cdots \subseteq \langle e_1, \cdots, e_n \rangle\}} = \emptyset$.

Length of $\mathcal{P}_{\mathcal{F}}$ = $n-$ length of $\mathcal{F} - 1$

(length of $\emptyset := -1$).

Magical definition (ε-protected)

We say t is ε-protected relative to $\mathcal{P}$ if $\exists$ a partial flag $\mathcal{F} \subseteq \mathcal{P}$ s.t.

(In $\mathcal{F}$) $\forall \Delta \in \mathcal{F}$, $\varepsilon \rho^{\operatorname{rank}\Delta} \le d(h(t)\Delta) \le \rho^{\operatorname{rank}\Delta}$.

(In $\mathcal{P}_{\mathcal{F}} \cap \mathcal{P}$) $\forall \Delta' \in \mathcal{P}_{\mathcal{F}} \cap \mathcal{P}$, $d(h(t)\Delta') \ge \rho^{\operatorname{rank}\Delta}$.

Lemma. Suppose $0 < \varepsilon^2 < \rho$ and t is $\frac{\varepsilon^2}{\rho}$-prot.

Lemma. Suppose $0 < \varepsilon < \sqrt{\rho}$ and $\mathcal{F}$ is $\frac{\varepsilon^2}{\rho}$-prot. relative to $\mathcal{P}(\mathbb{Z}^n)$. Then $\delta(h(t)\mathbb{Z}^n) \geq \varepsilon$.

Pf. $\exists$ a partial flag $\mathcal{F} = \{\Delta_1, \dots, \Delta_m\}$ s.t. the above conditions hold.

$\forall\, v \in \mathbb{Z}^n \setminus \{0\}$, let $\underline{i}$ be the smallest integer s.t. $v \in \Delta_i$.

__Case 1__. $\Delta_i = \Delta_{i-1} \oplus \mathbb{Z} v$

$$\Rightarrow \frac{\varepsilon^2}{\rho} \rho^{\operatorname{rank} \Delta_i} \leq d(h(t)\Delta_i) \leq d(h(t)\Delta_{i-1})\, d(h(t)v)$$
$$\leq \rho^{\operatorname{rank} \Delta_{i-1}} \cdot \|h(t)v\|^2$$

$$\Rightarrow \quad \varepsilon \leq \|h(t)v\|.$$

__Case 2.__ $\Delta_{i-1} \subsetneq \Delta_{i-1} \oplus \mathbb{Z} v \subsetneq \Delta_i$

$$\Rightarrow \quad \rho^{\operatorname{rank}\Delta_{i-1}+1} \leq d(h(t)\Delta_{i-1} + h(t)\mathbb{Z}v)$$
$$\leq d(h(t)\Delta_{i-1})\, \|h(t)v\|^2$$
$$\leq \rho^{\operatorname{rank}\Delta_{i-1}} \|h(t)v\|^2$$

$$\Rightarrow \quad \varepsilon \leq \sqrt{\rho} \leq \|h(t)v\|. \qquad \blacksquare$$

The above lemma roughly says that we can mark $\underline{\underline{t}}$ with partial flags as we did with vectors in $n=2$ case.

The following is the main property of a polynomial map that plays the role of ②.

Proposition. Let $f(x)$ be a polynomial of degree $\underline{d_0}$. Then for any interval I we have

$$\frac{1}{|I|}\left|\{t\in I \mid |f(t)|\le \varepsilon \max|f(I)|\}\right| \le d_0(d_0+1)^{1/d_0}\varepsilon^{1/d_0}.$$

Pf. Suppose $\ell := |\{t\in I \mid |f(t)|\le\varepsilon\max|f(I)|\}|$.

Since this set is a finite union of closed intervals,

$\exists\, x_0,\dots,x_{d_0}\in I$ s.t. ① $|f(x_i)|\le\varepsilon\max|f(I)|$

② $x_0<x_1<\cdots<x_{d_0}$.

③ $x_{i+1}-x_i\ge \ell/d_0$.

By Lagrange interpolation,

$$f(x)=\sum_{i=0}^{d_0} f(x_i)\prod_{j=0}^{d_0}\frac{(x-x_j)}{(x_i-x_j)}$$

$$f(x) = \sum_{i=0} f(x_i) \prod_{\substack{j=0 \\ i\neq j}} \frac{\quad}{(x_i - x_j)}$$

$$\Longrightarrow \quad \max |f(I)| \leq \varepsilon \max |f(I)| \cdot (d_0+1) \left(\ell/d_0\right)^{-d_0}$$

By shifting and rescaling we change $\underline{f}$, but not the ratio on the left hand side. So we can assume $I=[0,1]$

$$\Longrightarrow \quad \ell^{d_0} \leq (d_0+1)\, d_0^{d_0}\, \varepsilon \Longrightarrow \ell \leq (d_0+1)^{1/d_0}\, d_0\, \varepsilon^{1/d_0}. \quad \blacksquare$$

Now that we have almost all the needed tools, let's start the process:

<u>First Observation</u> It is enough to prove the following:

Suppose $0<\varepsilon^2<\rho<1$; $\forall \Delta \in \mathcal{P}(\mathbb{Z}^n)$

$$\circledast \qquad \max_{t\in[0,T]} d(h(t)\Delta) \geq \rho^{\operatorname{rank}(\Delta)}.$$

Then $\left|\{t\in[0,T] \mid t \text{ is NOT } \tfrac{\varepsilon^2}{\rho}\text{-protected relative to } \mathcal{P}(\mathbb{Z}^n)\}\right| \ll \left(\frac{\varepsilon}{\sqrt{\rho}}\right)^{1/d_0 n} T$

Instead we prove the above for $\mathcal{P} \subseteq \mathcal{P}(\mathbb{Z}^n)$. And use induction on the length of $\mathcal{P}$.

Start the process of marking:

$$t \leadsto \{\Delta \in \mathcal{P} \mid d(h(t)\Delta) \leq \rho^{\text{rank}\,\Delta}\} =: S(t)$$

is a finite set. (Possibly empty.)

If it is empty, t is ε/ρ protected.

Suppose $\exists\, \Delta \in \mathcal{P}$, $d(h(t)\,\Delta) \leq \rho^{\text{rank}(\Delta)}$.

$\forall \Delta \in S(t)$, find the largest interval $I_{t,\Delta} = [0,T] \cap [t-r, t+r]$

$$\text{s.t. } d(h(s)\,\Delta) \leq \rho^{\text{rank}\,\Delta} \quad \forall\, s \in I_{t,\Delta}.$$

By ⊛ and continuity it is clear that

$$\max_{s \in I_{t,\Delta}} d(h(s)\,\Delta) = \rho^{\text{rank}\,\Delta}.$$

Since $S(t)$ is a finite set, we can choose the largest $I_{t,\Delta}$. It will be denoted by I_t. So

$$\forall t,\ S(t) \neq \emptyset \Rightarrow \forall \Delta \in \mathcal{P},\ \max d(h(I_t)\,\Delta) \geq \rho^{\text{rank}\,\Delta}.$$

[If not, $\exists\, \Delta \in \mathcal{P}$, $\max d(h(I_t)\Delta) < \rho^{\text{rank}\,\Delta}$

$\Rightarrow d(h(t)\Delta) < \rho^{\operatorname{rank}\Delta}$

$\Rightarrow \Delta \in S(t) \Rightarrow \max d(h(I_{t\Delta})\Delta) = \rho^{\operatorname{rank}\Delta}$ ※].

Now we have a covering $\{I_t\}$ of the set $E := \{t \in [0,T] \mid S(t) \neq \emptyset\}$, and marking $t \longrightarrow \Delta_t$. And $\forall t \in S$, $\Delta \in \mathcal{P}$, $\max d(h(I_t)\Delta) \geq \rho^{\operatorname{rank}\Delta}$.

. For $t \in E$, consider $\mathcal{P}_{\{\Delta_t\}}$. As we mentioned earlier length of $\mathcal{P}_{\{\Delta_t\}}$ is less than length of $\mathcal{P}$. We know

$\forall \Delta \in \mathcal{P}_{\{\Delta_t\}}$, $\max d(h(I_t)\Delta) \geq \rho^{\operatorname{rank}\Delta}$.

So by the induction hypothesis we have I

$$\frac{1}{|I_t|}\left|\{s \in I_t \mid s \text{ is not } \tfrac{\varepsilon^2}{\rho}\text{-protected relative to } \mathcal{P}_{\{\Delta_t\}}\}\right| \ll \varepsilon^{1/d_0 n}.$$

Now that we know only a "small" portion of I_t is NOT ε^2/ρ-protected relative to $\mathcal{P}_{\{\Delta_t\}}$. We can assume that $s \in I_t$ is ε^2/ρ-protected relative to $\mathcal{P}_{\{\Delta_t\}}$. So there is a partial flag $\mathcal{F}_s \subseteq \mathcal{P}_{\{\Delta_t\}}$ s.t.

III $(\text{In } \mathcal{F}_s)$ $\forall \Delta \in \mathcal{F}$, $\left(\frac{\varepsilon^2}{\rho}\right) \rho^{\operatorname{rank}\Delta} \leq d(h(s)\Delta) \leq \rho^{\operatorname{rank}\Delta}$,

$\left(\text{In } \mathcal{P}_{\mathcal{F}_s \cup \{\Delta_t\}} \cap \mathcal{P}\right)$ $\forall \Delta \in \mathcal{P}_{\mathcal{F} \cup \{\Delta_t\}} \cap \mathcal{P}$, $\quad d(h(s)\Delta) \geq \rho^{\operatorname{rank}\Delta}$.

On the other hand, we have proved that

II

$$\frac{1}{|I_t|} \left| \{ s \in I_t \mid \quad d(h(s)\Delta_t) \leq \frac{\varepsilon^2}{\rho} \cdot \rho^{\operatorname{rank}\Delta_t} \} \right|$$

$$= \frac{\varepsilon^2}{\rho} \max d(h(I_t)\Delta_t)$$

$$\ll$$

$$(\varepsilon/\sqrt{\rho})^{1/d_0 n}.$$

If s is outside of I and II, then by III

s is $\frac{\varepsilon^2}{\rho}$-protected relative to $\mathcal{P}$ (using $\mathcal{F}_s \cup \{\Delta_t\}$).

Hence

$$\frac{1}{|I_t|} \left| \{ s \in I_t \mid s \text{ is NOT } \varepsilon^2/\rho\text{-protected relative to } \mathcal{P} \} \right| \ll (\varepsilon/\sqrt{\rho})^{1/d_0 n}.$$

- $\{I_t\}$ is a covering of E by intervals. So it has a subcover $\{I_{t_i}\}_{i=1}^{\infty}$ s.t. each point is covered by at most two intervals, i.e. $2\mathbb{1}_E \geq \sum \mathbb{1}_{I_{t_i}} \geq \mathbb{1}_E$. (why?)

most two intervals, i.e. $\mathbb{1}_E \leq \sum \mathbb{1}_{I_{t_i}} \leq 2 \cdot \mathbb{1}_E$. (why?)

. So we are done. $\blacksquare$

<u>**Remark**</u> The existence of such subcover is a special case of Besicovitch Covering Theorem. See your exercise for more details.