ROTH DIOPHANTINE APPROXIMATION, CONCISION ATTEMPTED

DANIEL RUI

ABSTRACT. The proof of Roth's thm. on Diophantine approx. of algebraic $\alpha \in \overline{\mathbb{Q}}$ is widely said to be difficult. Sadly, I wasn't able to simplify much. I followed Gaurish(2015)'s thorough writeup. **Key ideas:** (1) Chernoff concentration makes $\frac{n}{2}$ "order of vanishing (Orv)" at $\vec{\alpha}$ for n-var poly $P(\vec{x})$ easy to cook up. (2) Poly $P(\vec{x})$ eval at point $(\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n}) =: \frac{\vec{a}}{\vec{b}}$, get value $\in \mathbb{Z}[\prod \frac{1}{b^{\vec{r}}}]$, which is discrete(!!!), leading to P vanishing to significant order at nearby $\frac{\vec{a}}{\vec{b}}$. (3) Generalized Wronskians G of P are very amenable to inductive bounding of Orv, and there is tight relation b/t Orv of G and P.

CONTENTS

1 Part 1/2: High Order of Vanishing (Orv) 2
1.1 Central Estimate (introduce $\phi^{\vec{r}}(\vec{i})$) . . . 2
1.2 Chernoff — indices concentrated at $\frac{n}{2}$ . . . 3
1.3 Cook up P with high Orv at $\vec{\alpha}$ (using Chernoff and Siegel lemma) . . . 3
1.4 Trickle-down vanishing (high vanishing spreads to nearby) . . . 4
1.4.1 Weirder conditions on $\vec{r}$ . . . 5
2 Part 2/2: Low Order of Vanishing 5
2.1 Preliminaries on Wronskian . . . 5
2.1.1 1-variable z . . . 5
2.1.2 Multi-variable $z_1, \dots, z_h$ . . . 6
2.2 Roth low-vansishing thm. proof — Base case $n = 1$ . . . 6
2.3 Roth low-vanishing thm. proof — Induction step $n - 1$ to n . . . 7
2.3.1 Relate Orv of P and U at $\frac{\vec{a}}{\vec{b}}$; get sandwich bound on Orv P . . . 8
2.3.2 Finish Roth low-vanishing thm proof . . . 9
2.4 Final contradiction (thus proving Roth Thm.) . . . 10
3 Appendices 11
3.1 Appx. on "Orv of multi-var Poly (at point) (w.r.t. measuring stick ϕ)" . . . 11
3.2 Appx. on Wronskians . . . 12
3.2.1 1-var . . . 12
3.2.2 Multi-variable . . . 12

E-mail address: danrui@math.ucla.edu.
Funding: ∅.

Trying to condense proof of Roth's diophantine approximation theorem.

Notation. $\vec{\bullet}$ is always len n vector (unless o.w. indicated). $P(\vec{x}) = \sum_{0 \leqslant \vec{j} \leqslant \vec{r}} q(\vec{j}) \vec{x}^{\vec{j}} \in \mathbb{Z}[\vec{x}]$, "qoefficients". $\|P\|_\infty$ is abs-max qoefficient.

1. Part 1/2: High Order of Vanishing (Orv)

Notation: $P_{\vec{i}} = \frac{1}{\vec{i}!} \partial^{\vec{i}} P \coloneqq \frac{1}{i_1! \cdots i_n!} \frac{\partial^{i_1}}{\partial^{i_1} x_1} \cdots \frac{\partial^{i_n}}{\partial^{i_n} x_n} P$.

- $P \in \mathbb{Z}[\vec{x}] \implies P_{\vec{i}} \in \mathbb{Z}[\vec{x}]$ (Gaurish 3.4.1)
- $\|P_{\vec{i}}\|_\infty \leqslant 2^{|\vec{r}|} \|P\|_\infty$ (Gaurish 3.4.2)

Gaurish3.4.2 actually says something more precise (important for when we apply Siegel lemma). Each term of $P_{\vec{i}}$ has 3 prodands: q part, factorial part, $\vec{x}$ part. More specifically, Gaurish3.4.1 computes $P_{\vec{i}}(\vec{x}) = \sum_{\vec{i} \leqslant \vec{j} \leqslant \vec{r}} q(\vec{j}) \cdot \prod \binom{\vec{j}}{\vec{i}} \cdot \vec{x}^{\vec{j} - \vec{i}}$, but not important to know the specifics. Gaurish3.4.2 shows that all the factorial parts $\leqslant 2^{|\vec{r}|}$. In summary,

$$\text{"each term of } P_{\vec{i}} \text{ has 3 prodands: } \underbrace{q \text{ part}}_{\leqslant \|P\|_\infty}, \underbrace{\text{factorial part}}_{\leqslant 2^{|\vec{r}|}}, \underbrace{\vec{x} \text{ part}}_{\leqslant (|\alpha| \vee 1)^{|\vec{r}|}}\text{"}$$

$\mathfrak{a}, \mathfrak{a}, \mathfrak{a}, \mathfrak{a}, \mathfrak{a}$

Gaurish 3.4.4: α alg. int. with min-poly $\mathfrak{a}(x) = x^d + \ldots$.

Then $\forall \ell \geqslant 1$, α^ℓ written in basis $\alpha^{<d} \coloneqq \{\alpha^{d-1}, \ldots, \alpha^0\}$ in bdd way:

$$\alpha^\ell = \langle \underbrace{\text{basis weights}}_{\ell^\infty \text{ bdd by } (\|\mathfrak{a}\|_\infty + 1)^\ell}, \{\alpha^{<d}\} \rangle$$

1.1. Central Estimate (introduce $\phi^{\vec{r}}(\vec{i})$)

Central Estimate: for above poly $P(\vec{x})$, Taylor expand centered at $\vec{\alpha} \coloneqq (\underbrace{\alpha, \ldots, \alpha}_{n \text{ times}})$, eval at point $(\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}) =: \frac{\vec{a}}{\vec{b}}$, get values in $\in \mathbb{Z}[\prod_{\vec{}} \frac{1}{\vec{b}^{\vec{r}}}]$ (discrete lattice!!!). Assuming poly $\neq 0$ at this point,

$$\prod_{\vec{}} \frac{1}{\vec{b}^{\vec{r}}} \leqslant \left| P\left(\frac{\vec{a}}{\vec{b}}\right) \right| \leqslant \underbrace{\sum_{0 \leqslant \vec{i} \leqslant \vec{r}}}_{\text{"}L^1\text{"}} \left(\underbrace{|P_{\vec{i}}(\vec{\alpha})| \cdot \prod_{\vec{}} \left| \frac{\vec{a}}{\vec{b}} - \vec{\alpha} \right|^{\vec{i}}}_{\text{"}L^\infty\text{"}} \right) \leqslant (\#\text{terms}) \cdot (\text{size of largest term})$$

Control lime:

$$\underbrace{|P_{\vec{i}}(\vec{\alpha})|}_{\text{"}L^\infty\text{"}} \leqslant (\#\text{terms}) \cdot (\text{size of largest term})$$

$$\leqslant \prod_{\vec{}} (\vec{r} + 1) \cdot \|P_{\vec{i}}\|_\infty (|\alpha| \vee 1)^{|\vec{r}|}$$

$$\leqslant 2^{|\vec{r}|} \cdot 2^{|\vec{r}|} \|P\|_\infty (|\alpha| \vee 1)^{|\vec{r}|}$$

(the blue bound by $2^{|\vec{r}|}$ is super over estimate, but oh well)

Control orange: recall that we assume $\frac{a}{b}$ s.t. $\left|\frac{a}{b} - \alpha\right| < \frac{1}{b^{2+\delta}}$ so

$$\underbrace{\prod_{\vec{}} \left| \frac{\vec{a}}{\vec{b}} - \vec{\alpha} \right|^{\vec{i}}}_{\text{"}L^\infty\text{"}} \leqslant \max_{0 \leqslant \vec{i} \leqslant \vec{r}} \prod_{\vec{}} \left| \frac{\vec{a}}{\vec{b}} - \vec{\alpha} \right|^{\vec{i}} < \max_{0 \leqslant \vec{i} \leqslant \vec{r}} \prod_{\vec{}} \left(\frac{1}{\vec{b}} \right)^{(2+\delta)\vec{i}}$$

And so "Central Estimate" rearranges to

$$\frac{1}{2^{|\vec r|}\cdot \text{lime}} \leqslant \text{orange}\cdot \frac{1}{\prod_{\vec{}}(\frac{1}{\vec b})^{\vec r}} \leqslant \max_{0\leqslant \vec i\leqslant \vec r} \prod_{\vec{}}\left(\frac{1}{\vec b}\right)^{2.01\vec i-\vec r} \leqslant \max_{0\leqslant \vec i\leqslant \vec r} \frac{1}{\prod_{\vec{}} \vec b^{\vec r\cdot(2.01\frac{\vec i}{\vec r}-1)}};$$

supposing (all entries of) $\vec b^{\vec r}\geqslant b_1^{r_1}$, then

$$\frac{1}{2^{|\vec r|}\cdot (4|\alpha|\vee 4)^{|\vec r|}\|P\|_\infty} \leqslant \ldots \leqslant \max_{0\leqslant \vec i\leqslant \vec r} \frac{1}{\left(b_1^{r_1}\right)^{2.01(\sum_{\vec{}}\frac{\vec i}{\vec r})-n}} = \frac{b_1^{r_1 n}}{\left(b_1^{r_1}\right)^{2.01\cdot \min_{0\leqslant \vec i\leqslant \vec r}(\sum_{\vec{}}\frac{\vec i}{\vec r})}} \qquad (❦)$$

Above, whenever I wrote "$0\leqslant \vec i\leqslant \vec r$", I really meant $\vec i\in \mathfrak{I}_{\vec a/\vec b}(P)$, defined as:

$\mathfrak{I},\mathfrak{F},\mathfrak{J},\mathfrak{I},\mathfrak{J},\mathfrak{I},I,\mathcal{I},\mathcal{J}$

Define $\mathfrak{I}_{\vec p}(P):=\mathfrak{I}(P;\vec p):=\{\vec i(\leqslant \vec r)\in\mathbb{N}^n : P_{\vec i}(\vec p)\neq 0\}$. We quantify "**order of vanishing** of P at point $\vec p$ (w.r.t. "measurement" corresponding to $\vec r$)" by measuring this set of indices $\mathfrak{I}$ using a linear functional $\phi=\phi^{\vec r}$ ("project set to a line, and look at closest point to origin")

$$\phi^{\vec r}(\vec i):=\sum_{\vec{}}\frac{\vec i}{\vec r}=\langle \vec i,\frac{1}{\vec r}\rangle. \qquad \mathrm{ind}_{(\vec p,\vec r)}(P):=\mathrm{orv}^{\vec r}_{\vec p}(P):=\min \phi^{\vec r}(\mathfrak{I}_{\vec p}(P))$$

See "Appendix on Orv" later.

1.2. Chernoff — indices concentrated at $\frac{n}{2}$

The "2" in $\frac{n}{2}$ is the cause of the "2" in Roth's $b^{(2+\delta)}$ — cancellation $\frac{n}{2}\cdot(2+\delta)$ in next section! Hoeffding inequality applies: consider independent $X_j\sim \mathrm{Unif}[0..r_j]$, so $U_j:=\frac{1}{r_j}X_j$ are $[0,1]$-valued, mean $\frac{1}{2}$; then

$$\#\left\{\text{integer tuples } \vec i\in[0..\vec r] : \underbrace{\left|\sum_{j=1}^n U_j-\frac{n}{2}\right|}_{\phi^{\vec r}(\vec i)}>\varepsilon\cdot\frac{n}{2}\right\}\leqslant 2e^{-\varepsilon^2 n/2}\prod_{j=1}^n(r_j+1)$$

We actually only need lower-deviation half of Hoeffding:

$$\#\left\{\text{integer tuples } \vec i\in[0..\vec r] : \phi^{\vec r}(\vec i)<(1-\varepsilon)\cdot\frac{n}{2}\right\}\leqslant e^{-\varepsilon^2 n/2}\prod_{j=1}^n(r_j+1)$$

1.3. Cook up P with high Orv at $\vec\alpha$ (using Chernoff and Siegel lemma)

$\mathfrak{P},\mathfrak{P},\mathfrak{P},\mathfrak{P},\mathfrak{P},\mathcal{P},\mathcal{P},\mathcal{P}$

(Gaurish3.4.1) We use Siegel's lemma (lin. alg. + pigeonhole) and above Chernoff bound to cook up a polynomial $\mathfrak{P}(\vec x)=\sum_{0\leqslant \vec i\leqslant \vec r} q(\vec i)\vec x^{\vec i}\in\mathbb{Z}[\vec x]$ with high order-of-vanishing at $\vec\alpha$ (we get to control n, the dimensionality of the poly $\mathfrak{P}$; the higher it is, the easier this construction. We pick n large enough, then enemy can pick $\vec r$).

We must determine $N:=\prod(\vec r+1)=\prod_{j=1}^n(r_j+1)$ many qoefficients $\in\mathbb{Z}$.

More specifically, we want $\mathrm{orv}^{\vec r}_{\vec\alpha}(\mathfrak{P})\geqslant\frac{n}{2}(1-\varepsilon)$, which is equivalent to

$$[\mathfrak{P}_{\vec i}(\vec\alpha)=0 \text{ for all } 0\leqslant\vec i\leqslant\vec r \text{ with } \phi^{\vec r}(\vec i)<\tfrac{n}{2}(1-\varepsilon)]$$

By Chernoff above, these amount to $\leqslant(e^{-\varepsilon^2 n/2})\cdot N$ many linear constraints $\sum_{\vec i}[\mathbb{Z}[\alpha]\text{-coeffs}]\cdot q(\vec i)$ on the qoefficients $\{q(\vec i)\}_{0\leqslant\vec i\leqslant\vec r}$.

I.e. the linear constraint equations have variables $\{q(\vec{i})\}_{0\leqslant\vec{i}\leqslant\vec{r}}$, and $\mathbb{Z}[\alpha]$-coefficients bounded by $2^{|\vec{r}|}(|\alpha|\vee 1)^{|\vec{r}|}$ — cf. above discussion on Gaurish3.4.2,

$$\text{“each term of } \mathfrak{P}_{\vec{i}} \text{ has 3 prodands: } \underbrace{q \text{ part}}_{\leqslant\|\mathfrak{P}\|_\infty}, \underbrace{\text{factorial part}}_{\leqslant 2^{|\vec{r}|}}, \underbrace{\vec{x} \text{ part}''}_{\leqslant(|\alpha|\vee 1)^{|\vec{r}|}} .$$

By expanding in basis $\{\alpha^{d-1},\dots,\alpha^0\}$ (of d-dim $\mathbb{Q}$-VS $\mathbb{Q}[\alpha]$), and Gaurish3.4.4 (from which recall $\mathfrak{a}\in\mathbb{Z}[x]$ is min-poly for α) we get

$\leqslant M := (e^{-\varepsilon^2 n/2})\cdot N\cdot d$ many linear constraints in N $(\leqslant 2^{|\vec{r}|})$ many variables $\{q(\vec{i})\}_{0\leqslant\vec{i}\leqslant\vec{r}}$ with $\mathbb{Z}$-coeffs bounded by $Q := 2^{|\vec{r}|}(\|\mathfrak{a}\|_\infty+1)^{|\vec{r}|}$.

Siegel's lemma (Gaurish3.4.3 — pretty easy lin. alg. + pigeonhole) says: IF
given M linear constraints
in $N > M$ variables
and $\mathbb{Z}$-coeffs abs. bdd. $\leqslant Q$,
THEN there is an integer solution $\vec{z}$ with $\|\vec{z}\|_\infty \leqslant (NQ)^{M/(N-M)}$. In particular, if $N > 2M$, then $\|\vec{z}\|_\infty \leqslant NQ$.

In conclusion, for $\boxed{n \text{ s.t. } e^{-\varepsilon^2 n/2} d \leqslant \frac{1}{2}}$ and any $\vec{r}\in\mathbb{N}_+^n$, Siegel's lemma produces for us qoefficients $q(\vec{i})$ for a poly $\mathfrak{P}(\vec{x})$ s.t. $\|\mathfrak{P}\|_\infty \leqslant NQ \leqslant (2\cdot 2(\|\mathfrak{a}\|_\infty+1))^{|\vec{r}|}$ and $\text{orv}^{\vec{r}}_{\vec{\alpha}}(\mathfrak{P}) \geqslant \frac{n}{2}(1-\varepsilon)$, as desired.

1.4. Trickle-down vanishing (high vanishing spreads to nearby)

(Gaurish3.4.2) This section, we show that for $\frac{\vec{a}}{\vec{b}}$ “close enough” to $\vec{\alpha}$, the poly $\mathfrak{P}$ we just constructed (Siegel lemma) has $\geqslant n\varepsilon$ “order-of-vanishing (w.r.t. $\vec{r}$)” at $\frac{\vec{a}}{\vec{b}}$,

We do this by showing that for $T := \mathfrak{P}_{\vec{i}}$ for $\vec{i}$ s.t. $\phi^{\vec{r}}(\vec{i}) < n\varepsilon$, $T(\frac{\vec{a}}{\vec{b}})$ must be 0 (i.e. we want to force vanishing).

Gaurish3.4.6=Dalmau4.5 item (a) tells us $\text{orv}^{\vec{r}}_{\vec{\alpha}}(T) \geqslant \frac{n}{2}(1-3\varepsilon)$.

To force vanishing, by the Central Estimate section above, it suffices to prove the opposite inequality to (❦), which recall was supposing (all entries of) $\vec{b}^{\vec{r}} \geqslant b_1^{r_1}$,

$$\frac{1}{2^{|\vec{r}|}\cdot(4|\alpha|\vee 4)^{|\vec{r}|}\|\mathfrak{P}\|_\infty} \overset{\text{want}}{>} \frac{b_1^{r_1 n}}{\left(b_1^{r_1}\right)^{2.01\cdot\text{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(T)}}$$

Well, we do have

$$\left(b_1^{r_1}\right)^{2.01\cdot\text{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(T)-n} \geqslant \left(b_1^{r_1}\right)^{(\frac{2+\delta}{2})n(1-3\varepsilon)-n} = \left(b_1^{r_1}\right)^{n\left((1+\frac{\delta}{2})(1-3\varepsilon)-1\right)}$$

Fact: $(1+\vartheta)(1-\vartheta) = 1-\vartheta^2 < 1$, but $(1+2\vartheta)(1-\vartheta) = 1+\vartheta-\vartheta^2 > 1+\frac{\vartheta}{2}$ for $\vartheta\in[0,\frac{1}{2})$, so for $\boxed{\delta \text{ s.t. } \delta<1}$, $\boxed{\varepsilon \text{ s.t. } \varepsilon<\frac{\delta}{12}}$, we can continue the previous display-mode line to

$$\left(b_1^{r_1}\right)^{2.01\cdot\text{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(T)-n} \geqslant \ldots \geqslant (b_1^{r_1})^{n\cdot\delta/8}.$$

So the goal is now

$$\left(2\cdot 4(|\alpha|\vee 1)\cdot 4(\|\mathfrak{a}\|_\infty+1)\right)^{|\vec{r}|} \overset{\text{want}}{<} (b_1^{r_1})^{n\cdot\frac{\delta}{8}}$$

In conclusion, for $\boxed{\varepsilon \text{ s.t. } \varepsilon<\frac{\delta}{12}}$ and $\boxed{\vec{r},\vec{b} \text{ s.t. } \vec{b}^{\vec{r}} \geqslant b_1^{r_1} \text{ and } |\vec{r}| \leqslant n\cdot r_1 \text{ and } b_1^{\delta/8} > D := (32(|\alpha|\vee 1)(\|\mathfrak{a}\|_\infty+1))}$ then we win: $\text{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(\mathfrak{P}) \geqslant n\varepsilon$.

1.4.1. Weirder conditions on $\vec{r}$.— Suppose further that $\vec{b}^{\vec{r}} \leqslant b_1^{(1+\varepsilon)r_1}$, and $\vec{b}^{\delta/16} > D$, and $D \coloneqq 32(|\alpha| \vee 1)(\|\mathfrak{a}\|_\infty + 1)$. Then the above conditions are satisfied: indeed,

$$D^{|\vec{r}|} = \prod_{j=1}^{n} D^{r_j} < \prod_{j=1}^{n} b_j^{r_j \cdot \frac{\delta}{16}} \leqslant \prod_{j=1}^{n} b_1^{r_1(1+\varepsilon)\cdot\frac{\delta}{16}} \leqslant \prod_{j=1}^{n} b_1^{r_1 \cdot \frac{\delta}{8}} \quad \checkmark$$

2. Part 2/2: Low Order of Vanishing

Theorem 2.1: Roth's low vanishing theorem (1955)

IF we have the following:
Let $n \in \mathbb{N}^+$.
Let $\varepsilon < \frac{\delta}{12}$.
Define double exponentially small $⓪ = ⓪(n,\varepsilon) \coloneqq \frac{24}{2^n}(\frac{\varepsilon}{12})^{2^{n-1}}$ (if ε is $0.[p$ zeroes$]$, then think $⓪$ is $0.[$like $(p+1)\cdot 2^{n-1} + \frac{n}{3}$ zeroes$]$).
Note that for $\varepsilon' \coloneqq \frac{1}{12}\varepsilon^2$, $⓪' \coloneqq ⓪(n-1,\varepsilon') = 2 \cdot ⓪(n,\varepsilon)$. Quadratically smaller ε, but $n-1$ instead of n more than compensates to make $⓪'$ "weaker" than $⓪$ ("half as punishingly small").

Consider sequence $\vec{r}$,

- (I) ($\vec{r} \searrow$ Cond.) starting off with r_1 HUGE and decreasing by factor of $⓪$ each time: $r_{j+1} \leqslant ⓪ \cdot r_j$.

Suppose $\theta \in [0,1]$ (though when we apply low vanishing thm later §2.4, we will take $\theta = 1$) and $\frac{a_j}{b_j}$ reduced fractions with denominators $\vec{b} > 0$, in fact

- (II) ($\vec{b} \geqslant 日$ Cond.) $\vec{b}^{\theta⓪} \geqslant 2^{3n}$ (in ptlcr, $\implies b_1^{\theta⓪} \geqslant 2^{3n}$). Notation $日 \coloneqq 2^{3n\cdot\frac{1}{\theta⓪}}$ chosen b/c $2^3 = 8$ looks like $日$.
- (III) ($\vec{b}^{\vec{r}}$ 1st-term Cond.) $\vec{b}^{\vec{r}} \geqslant (b_1^{r_1})^\theta$

Let $P(\vec{x}) \in \mathbb{Z}[\vec{x}]$ non-zero with degree $\leqslant \vec{r}$ s.t. $|\vec{r}| \leqslant n \cdot r_1$ (I think this is easily implied by (I)?) and

- (IV) (Height Bound Cond.) $\|P\|_\infty \leqslant (b_1^{r_1})^{\theta⓪}$.

THEN, we have extremely small order of vanishing (measured w.r.t. $\phi^{\vec{r}}$) of P at $\frac{\vec{a}}{\vec{b}}$:

$$\mu \coloneqq \mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P) \leqslant \varepsilon.$$

2.1. Preliminaries on Wronskian

2.1.1. 1-variable z.— $d_i \coloneqq \frac{1}{i!}\partial^i$
Given functions $g_0, \ldots, g_{l-1}$ in 1 variable z.

$$W(z) \coloneqq \det\left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{l-1} \end{pmatrix}(g_0, g_1, \ldots, g_{l-1})\right]$$

Key Fact: $g_0, \ldots, g_{l-1}$ are $\mathbb{Q}$-LI iff $W \not\equiv 0$:

For functions, LD $\implies$ $[W \equiv 0]$ is easy.

Claim: For polys $g_j \in \mathbb{Q}[z]$, $[W \equiv 0] \implies$ LD. Slightly more work — but just linear algebra. See Appendix.

2.1.2. Multi-variable $z_1, \dots, z_h$.— $\mathfrak{D}, \mathcal{D}, \mathcal{D}$

"some/any $\mathcal{D}_r$" means "some/any differential operator $\mathcal{D}_{\vec{i}} := \frac{1}{\vec{i}!}\partial^{\vec{i}}$ for $|\vec{i}| \leqslant r$".

Given functions $g_0, \dots, g_{l-1}$ in variables $z_1, \dots, z_h$. (Later when we apply this Wronskian theory, $h = n-1$ or $h = 1$.)

For any $\mathcal{D}_0, \dots, \mathcal{D}_{l-1}$ (which again stand for $\mathcal{D}_{\vec{i}_0}, \dots, \mathcal{D}_{\vec{i}_{l-1}}$ where $|\vec{i}_0| \leqslant 0, \dots, |\vec{i}_{l-1}| \leqslant l$) we can form the **generalized Wrońskian** (which does depend on the specific diff ops — i.e. the specific indices $|\vec{i}| \leqslant r$ — chosen)

$$G(z_1, \dots, z_h) := \det\left[\begin{pmatrix} \mathcal{D}_0 \\ \mathcal{D}_1 \\ \vdots \\ \mathcal{D}_{l-1} \end{pmatrix}(g_0, g_1, \dots, g_{l-1})\right]$$

gen Wronsk G depend on 2 choices not explicitly written in notation for G: "col-vec" of choice of diff ops, and "row-vec" for choice of (usually LI — see below) polys.

Main theorem: LI $\leftrightarrow$ Wronskian (nonvanishing) connection:

Gaurish3.5.2 says: $g_0, \dots, g_{l-1} \in \mathbb{Q}[z_1, \dots, z_h]$ (say multi-degree $\leqslant \vec{d} \leqslant d$) are $\mathbb{Q}$-LI implies at-least-one gen. Wronskian $G(z_1, \dots, z_h) \not\equiv 0$.

> *Proof idea (Gaurish3.5.2):* consider single-var polys $g_0(\vec{\gamma}(t)), \dots, g_{l-1}(\vec{\gamma}(t)) \in \mathbb{Q}[t]$ where $\vec{\gamma}(t) := (t, t^d, \dots, t^{d^{h-1}})$ is a kind of "moment curve", which must still be $\mathbb{Q}$-LI (moment curve exponents chosen so we can use uniqueness of base-d repr.); then use 1-var Wronskian result to get W (w.r.t. $g_0 \circ \vec{\gamma}, \dots, g_{l-1} \circ \vec{\gamma}$ in $\mathbb{Q}[t]$) is $\not\equiv 0$, which does eventually force some gen. Wronskian $G \not\equiv 0$ — keep in mind fact (Gaurish3.5.1)
>
> $$\left[d_r[g_j \circ \vec{\gamma}]\right](t) = \sum \underbrace{[\text{poly}]}_{\in \mathbb{Q}[t]} \cdot [\mathcal{D}_r g_j](t).$$

Bostan-Dumas https://arxiv.org/pdf/1301.6598 (mentioned above) prove multi-var case without reducing to 1-var case, using Vandermonde matrix theory. See Appendix.

2.2. Roth low-vansishing thm. proof — Base case $n = 1$

Base case $n = 1$ *(Gaurish pg.55):* Note $\circledcirc(1, \varepsilon) = \varepsilon =: \circledcirc$. We have $P(x) = [\text{leading qoeff}]x^{r_1} + \ldots \in \mathbb{Z}[x]$ of deg r_1, which vanishes to order/multiplicity $m \in \mathbb{N}$: (Taylor expansion:)

$$P(x) = (x - \tfrac{a_1}{b_1})^m \cdot \underbrace{Q(x)}_{\in \mathbb{Q}[x], Q(\frac{a_1}{b_1}) \neq 0}$$

The polynomial $b_1 x - a_1 \in \mathbb{Z}[x]$ is irreducible (by assumption of reduced frac $\frac{a_1}{b_1}$), so $(b_1 x - a)^m$ dividing $[\text{const}(= b_1^m)] \cdot P(x)$ in the UFD $\mathbb{Z}[x]$ forces $(b_1 x - a)^m$ to divide $P(x)$ in $\mathbb{Z}[x]$:

$$P(x) = (b_1 x - a_1)^m \cdot \underbrace{Z(x)}_{\in \mathbb{Z}[x]}.$$

Therefore, b_1^m divides leading qoeff of P, so

$$b_1^m \leqslant \|P\|_\infty \leqslant (b_1^{r_1})^{\theta^{\circledcirc}} \leqslant (b_1^{r_1})^{\circledcirc} \implies m \leqslant r_1{}^{\circledcirc} \qquad \implies \text{orv}^{(r_1)}_{(a_1/b_1)}(P) := \frac{m}{r_1} \leqslant \varepsilon. \quad \boxtimes$$

2.3. Roth low-vanishing thm. proof — Induction step $n-1$ to n

Induction step $n-1$ to n (Gaurish pg.56): we start with $P(\vec{x}) \in \mathbb{Z}[\vec{x}]$ of degree $\leqslant \vec{r}$ satisfying preconditions of Roth low-vanishing thm.

Let $k \in \mathbb{N}^+$ be minimal pos. integer satisfying

$$P(\vec{x}) = \sum_{j=1}^{k} \underbrace{\varphi_j(x_1,\dots,x_{n-1})}_{\substack{\in \mathbb{Z}[\vec{x}']:=\mathbb{Z}[x_1,\dots,x_{n-1}] \\ \text{multi-deg} \leqslant \vec{r}'}} \cdot \underbrace{\psi_j(x_n)}_{\substack{\in \mathbb{Z}[x_n] \\ \deg \leqslant r_n}} = (\psi_1(x_n),\dots,\psi_k(x_n)) \begin{pmatrix} \varphi_1(\vec{x}') \\ \vdots \\ \varphi_k(\vec{x}') \end{pmatrix}.$$

Minimality of k forces (trivial rearranging algebra argument) $\varphi_1(\vec{x}'),\dots,\varphi_k(\vec{x}')$ are $\mathbb{Q}$-LI, and similarly $\psi_1(x_n),\dots,\psi_k(x_n)$ also $\mathbb{Q}$-LI. (Gaurish pg.56 Claim1,2).

From Wronskian theory (see above),

$$\mathbb{Z}[x_n] \ni \quad W(x_n) \coloneqq \det \left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \end{pmatrix} (\psi_1(x_n),\dots,\psi_k(x_n)) \right] \not\equiv 0$$

[Note: $W \not\equiv 0$ implies $k-1 \leqslant r_n$, because $\deg \psi_j(x_n) \leqslant r_n$ for all $j = [1..k]$, though we only use this fact in next subsection §2.3.1] and (for some choice of diff ops $\mathscr{D}'_0,\dots,\mathscr{D}'_{k-1}$, prime-superscript to emphasize they only act on variables $\vec{x}' \coloneqq (x_1,\dots,x_{n-1})$)

$$\mathbb{Z}[\vec{x}'] \ni \quad G(\vec{x}') \coloneqq \det \left[\begin{pmatrix} \mathscr{D}'_0 \\ \mathscr{D}'_1 \\ \vdots \\ \mathscr{D}'_{k-1} \end{pmatrix} (\varphi_1(\vec{x}'),\dots,\varphi_{k-1}(\vec{x}')) \right] \not\equiv 0$$

I claim (proof below) that

$$\begin{aligned} \left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \end{pmatrix} (\mathscr{D}'_0,\dots,\mathscr{D}'_{k-1})P \right] &= \left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \end{pmatrix} (\psi_1(x_n),\dots,\psi_k(x_n)) \right] \left[\mathscr{D}'_0 \begin{pmatrix} \varphi_1(\vec{x}') \\ \vdots \\ \varphi_k(\vec{x}') \end{pmatrix}, \dots, \mathscr{D}'_{k-1} \begin{pmatrix} \varphi_1(\vec{x}') \\ \vdots \\ \varphi_k(\vec{x}') \end{pmatrix} \right] \\ &= \left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \end{pmatrix} (\psi_1(x_n),\dots,\psi_k(x_n)) \right] \left[\begin{pmatrix} \mathscr{D}'_0 \\ \mathscr{D}'_1 \\ \vdots \\ \mathscr{D}'_{k-1} \end{pmatrix} (\varphi_1(\vec{x}'),\dots,\varphi_{k-1}(\vec{x}')) \right]^\top \end{aligned}$$

The 1st equality is true because

$$(\mathscr{D}'_0,\dots,\mathscr{D}'_{k-1})P = (\psi_1(x_n),\dots,\psi_k(x_n))(\mathscr{D}'_0 \begin{pmatrix} \varphi_1(\vec{x}') \\ \vdots \\ \varphi_k(\vec{x}') \end{pmatrix}, \dots, \mathscr{D}'_{k-1} \begin{pmatrix} \varphi_1(\vec{x}') \\ \vdots \\ \varphi_k(\vec{x}') \end{pmatrix}).$$

Taking determinants of the big matrix equalities above we get

$$U(\vec{x}) \coloneqq \det \left[\begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \end{pmatrix} (\mathscr{D}'_0,\dots,\mathscr{D}'_{k-1})P \right] = G(\vec{x}') \cdot W(x_n).$$

This is very good for induction.

For the sake of grappling with the def to understand it better, and also because we use this fact later §2.3.1, note that all entries of the $(k \times k)$-matrix are of the form $P_{\vec{i}}$ where $\vec{i} \coloneqq (\vec{i}', i_n)$ satisfies

$|\vec{i'}| \leqslant k-1$ ($\leqslant r_n$) and $i_n \leqslant k-1$ ($\leqslant r_n$), so the determinant is a sum of k-fold products of such $P_{(\vec{i'},i_n)}$.

We will eventually relate the Orv of P and U at $\frac{\vec{a}}{\vec{b}}$. (Recall we WTS $\mu := \mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P) \leqslant \varepsilon$.)

But first, we use induction to show $\mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U) \leqslant \frac{k}{6}\varepsilon^2$. By the valuation property of Orv, it suffices to show both

- $\mathrm{orv}^{\vec{r'}}_{\vec{a'}/\vec{b'}}(G) \leqslant \frac{k}{12}\varepsilon^2$ ($\Leftrightarrow \mathrm{orv}^{k\cdot\vec{r'}}_{\vec{a'}/\vec{b'}}(G) \leqslant \frac{1}{12}\varepsilon^2$)
- $\mathrm{orv}^{r_n}_{a_n/b_n}(W) \leqslant \frac{k}{12}\varepsilon^2$ ($\Leftrightarrow \mathrm{orv}^{k\cdot r_n}_{a_n/b_n}(W) \leqslant \frac{1}{12}\varepsilon^2$)

Want to apply induction hypothesis to $G(\vec{x'})$ (which has (multi-)degree $\leqslant k\vec{r'}$) with parameters $\varepsilon' := \frac{1}{12}\varepsilon^2$ and $\textcircled{\scriptsize D}' := \textcircled{\scriptsize D}(n-1,\varepsilon') := 2\cdot\textcircled{\scriptsize D}(n,\varepsilon)$

(I) we started with $\vec{r}$ satisfying ($\vec{r} \searrow$ Cond.) w.r.t. factor $\textcircled{\scriptsize D}$, and Cond. is monotone, continues to hold for $\textcircled{\scriptsize D}' > \textcircled{\scriptsize D}$ (more explicitly, $r_{j+1} \leqslant \textcircled{\scriptsize D}\cdot r_j \leqslant \textcircled{\scriptsize D}'\cdot r_j$)

(II) ($\vec{b} \geqslant \exists$ Cond.) also monotone: $\vec{b}^{\theta\textcircled{\scriptsize D}'} > \vec{b}^{\theta\textcircled{\scriptsize D}} \geqslant 2^{3n} > 2^{3(n-1)}$.

(III) ($\vec{b}^{\vec{r}}$ 1st-term Cond.) raise both sides to k: $\vec{b}^{\vec{r}} \geqslant (b_1^{r_1})^\theta \implies \vec{b}^{k\cdot\vec{r}} \geqslant (b_1^{k\cdot r_1})^\theta$.

(IV) (Height Bound Cond.) want $\|G\|_\infty \overset{\text{want}}{\leqslant} (b_1^{k\cdot r_1})^{\theta\textcircled{\scriptsize D}'}$. Suffice to prove (s.t.p.)

$$\|G\|_\infty \overset{\text{want}}{\leqslant} \|P\|_\infty^k \cdot (b_1^{r_1})^{\theta\textcircled{\scriptsize D}\cdot k}$$

See below box.

Note $U(\vec{x}) = W(\vec{x'})G(x_n) \implies \|W\|_\infty, \|G\|_\infty \leqslant \|W\|_\infty \cdot \|G\|_\infty = \|U\|_\infty$ (last equality follows because W, G in different variables, so monomials of U come from exactly 1 pair of monomials of W, G — no cancellation can occur)

Observe $U(\vec{x})$ is $(k\times k)$-det of polys of form $P_{\vec{i}}$, so $\|U\|_\infty \leqslant \|P_\bullet\|_\infty^k \leqslant (2^{|\vec{r}|}\|P\|_\infty)^k$.

Roth assumptions include $|\vec{r}| \leqslant n\cdot r_1$ (of course $\leqslant 3n\cdot r_1$), so

$$2^{|\vec{r}|} \leqslant (2^{3n})^{r_1} \underset{\text{(II)}}{\leqslant} (b_1^{\theta\textcircled{\scriptsize D}})^{r_1}$$

and so all together we have $\|W\|_\infty, \|G\|_\infty \leqslant \|U\|_\infty \leqslant \ldots \leqslant \|P\|_\infty^k \cdot (b_1^{r_1})^{\theta\textcircled{\scriptsize D}\cdot k} \leqslant (b_1^{r_1})^{\theta\textcircled{\scriptsize D}'}$, as desired for (IV) (see above box).

For $W(x_n)$, same $\varepsilon'' := \varepsilon' := \frac{1}{12}\varepsilon^2$, and $\textcircled{\scriptsize D}'' := \textcircled{\scriptsize D}(1,\varepsilon'') \geqslant= \textcircled{\scriptsize D}'$ (really understating the $\geqslant$ there lol). We want to apply induction hyp. to $W(x_n)$ (which has deg $\leqslant kr_n$) with params $\varepsilon'', \textcircled{\scriptsize D}''$.

Conditions (I), (II), (III) hold again trivially (monotonicity, monotonicity, raise-both-sides-to-k). And want

$$\text{(IV)}\quad \|W\|_\infty \overset{\text{want}}{\leqslant} (b_1^{k\cdot r_1})^{\theta\textcircled{\scriptsize D}''}.$$

This 'want' is implied (very *a fortiori*) by the bound we proved in the previous box $\|W\|_\infty \leqslant (b_1^{k\cdot r_1})^{\theta\textcircled{\scriptsize D}'}$ (because $\textcircled{\scriptsize D}'' > \textcircled{\scriptsize D}'$, in fact muchmuchmuch $>$)

2.3.1. Relate Orv of P and U at $\frac{\vec{a}}{\vec{b}}$; get sandwich bound on Orv P.— We just proved upper bound $\mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U) \leqslant \frac{k}{6}\varepsilon^2$. Now we prove lower bound $\mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U) \geqslant$[something involving $\mu := \mathrm{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P)$].

Consider any $\vec{i'} \in \mathbb{N}^{n-1}$ s.t. $|\vec{i'}| \leqslant k-1 \leqslant r_n$ (the fact that $k-1 \leqslant r_n$ was remarked immediately after defining $W \not\equiv 0$ above), and $i_n \in [0..k-1]$ — these conditions produce $\vec{i} := (\vec{i'}, i_n)$ that are exactly the multi-indices seen in the entries of the matrix defining U, above.

Because $r_1 \geqslant \ldots \geqslant r_n$, we have $\phi^{\vec{r}} \leqslant \phi^{\vec{r}_n}$ (pointwise as linear functionals $\mathbb{N}^n \to [0,\infty)$)

$$\begin{aligned}\operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P_{\vec{i}}) &\geqslant \underbrace{\operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P)}_{=:\mu} - \phi^{\vec{r}}(\vec{i}) \geqslant \mu - \underbrace{\phi^{(\vec{r_{n-1}})'}(\vec{i'})}_{=\frac{i_1}{r_{n-1}}+\ldots+\frac{i_{n-1}}{r_{n-1}}} - \frac{i_n}{r_n} \\ &= \mu - \frac{|\vec{i'}|}{r_{n-1}} - \frac{i_n}{r_n} \\ &\geqslant \mu - \frac{r_n}{r_{n-1}} - \frac{i_n}{r_n} \geqslant \mu - \textcircled{n} - \frac{i_n}{r_n}\end{aligned}$$

For $n \geqslant 2$, $\textcircled{n}(n,\varepsilon) \leqslant \frac{1}{24}\varepsilon^2$. So we conclude (also Orv always $\geqslant 0$), for all $\vec{i'} \in \mathbb{N}^{n-1}$ s.t. $|\vec{i'}| \leqslant k-1 \leqslant r_n$, and $i_n \in [0..k-1]$

$$\operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P_{\vec{i}}) \geqslant \max\left(\mu - \frac{1}{24}\varepsilon^2 - \frac{i_n}{r_n}, 0\right) \geqslant -\frac{1}{24}\varepsilon^2 + \max\left(\mu - \frac{i_n}{r_n}, 0\right)$$

But the determinant defining U is

- a sum (with some coeffs) of
  - k-fold products, selecting 1 prodand from each row,col of a $(k \times k)$-matrix; where the matrix-entries in row $0, \ldots, k-1$ (resp.) are derivatives $[\partial_{i_n} \mathcal{D}_{\vec{i'}} P]$ with i_n equal $0, \ldots, k-1$ (resp.).

So by valuation properties: $\operatorname{orv} \sum \geqslant \min \operatorname{orv}$, and $\operatorname{orv} \prod = \sum \operatorname{orv}$, we get for some $(\vec{i'})$'s of the above form ($\vec{i'}$ may be different in each summand; the notation just stands for *some* "ever changing/flickering" len=$(n-1)$ multi-index satisfying above ineqs),

$$\operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U) \geqslant \sum_{i_n=0}^{k-1} \operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(P_{(\vec{i'},i_n)}) \geqslant -\frac{k}{24}\varepsilon^2 + \sum_{i_n=0}^{k-1} \max\left(\mu - \frac{i_n}{r_n}, 0\right)$$

The sum on the RHS simplifies to a shorter sum with NO max (so then can use baby Gauss's sum formula!):

$$\sum_{i_n=0}^{k'-1}\left(\mu - \frac{i_n}{r_n}\right) = k'\mu - \frac{1}{r_n}\sum_{i_n=0}^{k'-1} i_n = \left(k'\mu - \frac{1}{r_n}\frac{k'(k'-1)}{2}\right)$$

for some $k' \leqslant k \leqslant r_n + 1 \leqslant 2r_n$ (namely k' s.t. $\mu - \frac{k'-1}{r_n} > 0$, but terms past that point are $\leqslant 0$ and hence get killed by max). Note in particular k' satisfies

$$\boxed{\mu \leqslant \tfrac{k'}{r_n} \leqslant \tfrac{k'}{\frac{1}{2}k}}.$$

2.3.2. Finish Roth low-vanishing thm proof.— Putting two halves of sandwich ineq for $\operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U)$, and then promptly forgetting about U, we get (Gaurish pg.60 "Claim3")

$$-\frac{k}{24}\varepsilon^2 + \left(k'\mu - \frac{1}{r_n}\frac{k'(k'-1)}{2}\right) \leqslant \operatorname{orv}^{\vec{r}}_{\vec{a}/\vec{b}}(U) \leqslant \frac{k}{6}\varepsilon^2.$$

Rearranging (and using that $\frac{1}{6} + \frac{1}{24} = \frac{5}{24} \leqslant \frac{1}{4}$), we get (recall $\mu - \frac{k'-1}{r_n} > 0$)

$$k'\mu - \frac{1}{r_n}\frac{k'(k'-1)}{2} \leqslant \frac{k}{4}\varepsilon^2 \qquad \Longrightarrow \frac{k}{k'}\cdot\tfrac{1}{4}\varepsilon^2 \geqslant \mu - \frac{1}{r_n}\frac{k'-1}{2} > \mu - \frac{\mu}{2}.$$

(Think about the case $k' = k$, it is easier.) But $\boxed{\mu \leqslant \tfrac{k'}{r_n} \leqslant \tfrac{k'}{\frac{1}{2}k}}$, so we get

$$\frac{2}{\mu}\cdot\tfrac{1}{4}\varepsilon^2 > \frac{\mu}{2},$$

so $\mu < \varepsilon$ as desired.

2.4. Final contradiction (thus proving Roth Thm.)

Recall in §1 we proved:

> In conclusion, for $\boxed{n \text{ s.t. } e^{-\varepsilon^2 n/2} d \leqslant \frac{1}{2}}$ and any $\vec{r} \in \mathbb{N}_+^n$, Siegel's lemma produces for us qoefficients $q(\vec{i})$ for a poly $\mathfrak{P}(\vec{x})$ s.t. $\|\mathfrak{P}\|_\infty \leqslant NQ \leqslant (2 \cdot 2(\|\mathfrak{a}\|_\infty + 1))^{|\vec{r}|}$ and $\mathrm{orv}_{\vec{\alpha}}^{\vec{r}}(\mathfrak{P}) \geqslant \frac{n}{2}(1-\varepsilon)$
>
> In conclusion, for $\boxed{\varepsilon \text{ s.t. } \varepsilon < \frac{\delta}{12}}$ and $\boxed{\vec{r}, \vec{b} \text{ s.t. } \vec{b}^{\vec{r}} \geqslant b_1^{r_1} \text{ and } |\vec{r}| \leqslant n \cdot r_1 \text{ and } b_1^{\delta/8} > D \coloneqq \big(32(|\alpha| \vee 1)(\|\mathfrak{a}\|_\infty + 1)\big)}$ then we win: $\mathrm{orv}_{\vec{a}/\vec{b}}^{\vec{r}}(\mathfrak{P}) \geqslant n\varepsilon$.

We make our landing:

- Fix α alg. int. with min-poly $\mathfrak{a}$ with deg d.
- Fix $\delta > 0$ (we will assume f.s.o.c. there are lots of $\frac{a}{b}$ s.t. $|\frac{a}{b} - \alpha| < \frac{1}{b^{2+\delta}}$ — "δ-good rational approx (to α)").
- Fix $\boxed{\varepsilon < \frac{\delta}{12}}$.
- Fix n s.t. $\boxed{e^{-\varepsilon^2 n/2} d \leqslant \frac{1}{2}}$.
  - Later when we settle on $\vec{r}$, we will invoke §1 (Siegel lemma) to get $\mathfrak{P}(\vec{x}) \in \mathbb{Z}[\vec{x}]^n$ of multi-deg $\leqslant \vec{r}$ with $\|\mathfrak{P}\|_\infty \leqslant (4(\|\mathfrak{a}\|_\infty + 1))^{|\vec{r}|}$.
  - Also set $\textcircled{\scriptsize 1} = \textcircled{\scriptsize 1}(n, \varepsilon) \coloneqq \frac{24}{2^n}(\frac{\varepsilon}{12})^{2^{n-1}}$
- Fix (reduced frac) $\frac{a_1}{b_1}$ a δ-good rational approx to α with $b_1 > 0$ large enough to satisfy the lower bounds

$$b_1^{\textcircled{\scriptsize 1}} > (4(\|\mathfrak{a}\|_\infty + 1))^n \vee 2^{3n}, \qquad \boxed{b_1^{\delta/8} > D \coloneqq (32(|\alpha| \vee 1)(\|\mathfrak{a}\|_\infty + 1))}$$

- Fix (reduced fracs) $\frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}$ all δ-good rational approx to α, s.t.

$$b_j \leqslant b_{j+1}^{\textcircled{\scriptsize 1}/2},$$

  so in particular $b_1 < \ldots < b_n$, which implies $\vec{b}^{\textcircled{\scriptsize 1}} \geqslant b_1^{\textcircled{\scriptsize 1}} \geqslant 2^{3n}$ (II).
- Fix $r_1 \in \mathbb{N}^+$ so large that … (read the next phrase after reading the below bullet points) … for all $j \in [1..n]$, the smallest (pos) integer r_j s.t. $b_j^{r_j} \geqslant b_1^{r_1}$ satisfies $\textcircled{\scriptsize 1} \cdot r_j \geqslant 1$.
- Set r_j to be the smallest (pos) integer s.t. $b_j^{r_j} \geqslant b_1^{r_1}$. So $\boxed{\vec{b}^{\vec{r}} \geqslant b_1^{r_1}}$ (III).
- By above ineq $b_j \leqslant b_{j+1}^{\textcircled{\scriptsize 1}/2}$, we have $b_1^{r_1} \leqslant b_j^{r_j} \leqslant b_{j+1}^{\textcircled{\scriptsize 1} r_j/2}$, which implies (by "smallest integer"-ness)

$$r_{j+1} \leqslant \frac{\textcircled{\scriptsize 1} r_j}{2} + 1 \leqslant \textcircled{\scriptsize 1} \cdot r_j \text{ (I)}$$

  where the 2nd inequality holds because $\textcircled{\scriptsize 1} \cdot r_j \geqslant 1$ (see above "Choose $r_1 \in \mathbb{N}^+ \ldots$" bullet).

  Because $\textcircled{\scriptsize 1} \leqslant 1$, this easily implies $\boxed{|\vec{r}| \leqslant n \cdot r_1}$.
- Above dealt with (I), (II), (III). We are finally ready to invoke §1 (Siegel lemma), meet condition (IV), and apply Roth's low vanishing theorem.
- §1 (Siegel lemma) produces $\mathfrak{P}(\vec{x}) \in \mathbb{Z}[\vec{x}]^n$ of multi-deg $\leqslant \vec{r}$ with $\|\mathfrak{P}\|_\infty \leqslant (4(\|\mathfrak{a}\|_\infty + 1))^{|\vec{r}|} < (b_1^{\textcircled{\scriptsize 1}})^{r_1}$, which is precisely (IV) (Height Bound Cond.).
- The contradiction is between §1 $\mathrm{orv}_{\vec{a}/\vec{b}}^{\vec{r}}(\mathfrak{P}) \geqslant n\varepsilon$ and §2 Roth's conclusion $\mathrm{orv}_{\vec{a}/\vec{b}}^{\vec{r}}(\mathfrak{P}) < \varepsilon$.

3. Appendices

3.1. Appx. on "Orv of multi-var Poly (at point) (w.r.t. measuring stick ϕ)"

Analogous to order of vanishing for meromorphic functions in complex analysis $\mathrm{ord}_{z_0}(f) \coloneqq \mathrm{orv}_{z_0}(f) = \min\{i : \partial^i f(z_0) \neq 0\}$. [Draw integer lattice points in $r_1 \times r_2$ rectangle in $\mathbb{R}^2$, circle some lattice points $\vec{i}$ signifying those s.t. $P_{\vec{i}}(\vec{p}) \neq 0$, look at those that could be "closest to the origin" in the eyes of some linear projection]

Example 3.1: Some concrete calculations to practice notation

$n = 2$.
$P(x,y) \coloneqq x - y, \quad Q(x,y) \coloneqq x + y + 2$.

$$(PQ)_{\vec{i}=(1,1)} = P_{(1,1)}\,Q + P_{(1,0)}Q_{(0,1)} + P_{(0,1)}Q_{(1,0)} + P\,Q_{(1,1)}$$

P, Q were cooked up so that $P_xQ_y \coloneqq P_{(1,0)}Q_{(0,1)}, P_yQ_x \coloneqq P_{(0,1)}Q_{(1,0)}$ are each $\neq 0$, but sum to 0.
Also, choosing $\vec{p} \coloneqq (-1,-1)$ (originally I had $\vec{p} = (1,1)$ and $Q(x,y) = x + y - 2$, but that $\vec{p}$ looks too much like an index $\vec{i}$), have $P(\vec{p}), Q(\vec{p}) = 0$ but all derivatives P_x, P_y, Q_x, Q_y at $\vec{p}$ are $\neq 0$.

$$\mathrm{I}_{\vec{p}=(-1,-1)}(P) = \{(0,1),(1,0)\} \subseteq \mathbb{N}^2 \quad \text{(same for } Q\text{)}$$

I claim that $\mathrm{I}_{\vec{p}}(PQ) \subseteq \mathrm{I}_{\vec{p}}(P) + \mathrm{I}_{\vec{p}}(Q)$ (proved below). In this example we have extra cancellation so this $\subseteq$ is actually $\subsetneq$:

$$\mathrm{I}_{\vec{p}=(-1,-1)}(PQ) = \{(0,2),(2,0)\} \subseteq \mathbb{N}^2$$

Can take $\vec{r} = (1,1)$ (or $(2,1)$, etc. — as long as our indices $0 \leqslant \vec{i} \leqslant \vec{r}$).

$$\mathrm{orv}^{\vec{r}}(PQ;\vec{p}) = \mathrm{orv}^{(1,1)}(PQ;(-1,-1)) = \mathrm{ind}_{(-1,-1)(1,1)}(PQ) = 2$$

I claim that $\mathrm{I}_{\vec{p}}(PQ) \subseteq \mathrm{I}_{\vec{p}}(P) + \mathrm{I}_{\vec{p}}(Q)$ (sumset). This is because $\partial^{\vec{i}}(PQ) = \sum_{\vec{i}_1+\vec{i}_2=\vec{i}} [\text{coeff}](\partial^{\vec{i}_1}P)(\partial^{\vec{i}_1}Q)$, so $\vec{i} \notin \mathrm{I}_{\vec{p}}(P) + \mathrm{I}_{\vec{p}}(Q)$ implies the summand corresponding to any sum $\vec{i}_1 + \vec{i}_2 = \vec{i}$ will have at least one of the prodands $\partial^{\vec{i}_1}P, \partial^{\vec{i}_2}Q$ equaling 0.

$$\implies \phi(\mathrm{I}_{\vec{p}}(PQ)) \subseteq \phi(\mathrm{I}_{\vec{p}}(P)) + \phi(\mathrm{I}_{\vec{p}}(Q))) \implies \min\phi(\mathrm{I}_{\vec{p}}(PQ)) \geqslant \min\phi(\mathrm{I}_{\vec{p}}(P)) + \min\phi(\mathrm{I}_{\vec{p}}(Q)))$$

i.e.

$$\mathrm{orv}^{\vec{r}}_{\vec{p}}(PQ) \geqslant \mathrm{orv}^{\vec{r}}_{\vec{p}}(P) + \mathrm{orv}^{\vec{r}}_{\vec{p}}(Q).$$

Dalmau4.5abc, "valuation" properties (cf. discrete valuation ring DVR)

a) $\mathrm{I}_{\vec{p}}(P_{\vec{i}}) \subseteq \mathrm{I}_{\vec{p}}(P) + \{-\vec{i}\}$ (obvious). Implies $\phi(\mathrm{I}_{\vec{p}}(P_{\vec{i}})) \subseteq \phi(\mathrm{I}_{\vec{p}}(P)) + \{-\phi(\vec{i})\}$, implies

$$\mathrm{orv}^{\phi}_{\vec{p}}(P_{\vec{i}}) \geqslant \mathrm{orv}^{\phi}_{\vec{p}}(P) - \phi(\vec{i})$$

b) $\mathrm{I}_{\vec{p}}(P+Q) \subseteq \mathrm{I}_{\vec{p}}(P) \cup \mathrm{I}_{\vec{p}}(Q)$ (obvious). Implies (by applying ϕ, take min)

$$\mathrm{orv}^{\phi}_{\vec{p}}(P+Q) \geqslant \min\left\{\mathrm{orv}^{\phi}_{\vec{p}}(P), \mathrm{orv}^{\phi}_{\vec{p}}(Q)\right\}$$

c) Above, got sumset containment and deduced $\mathrm{orv}^{\phi}_{\vec{p}}(PQ) \geqslant \mathrm{orv}^{\phi}_{\vec{p}}(P) + \mathrm{orv}^{\phi}_{\vec{p}}(Q)$.
The opposite inequality is also true, i.e. "log property"

$$\mathrm{orv}^{\phi}_{\vec{p}}(PQ) = \mathrm{orv}^{\phi}_{\vec{p}}(P) + \mathrm{orv}^{\phi}_{\vec{p}}(Q)$$

$$\left(\quad \min\phi(\mathrm{I}_{\vec{p}}(PQ)) = \min\phi(\mathrm{I}_{\vec{p}}(P)) + \min\phi(\mathrm{I}_{\vec{p}}(Q)) \quad\right)$$

Take the lex-minimal $\vec{i}_1 \in \mathrm{I}_{\vec{p}}(P)$ (attaining $\min \phi(\mathrm{I}_{\vec{p}}(P))$) and the lex-minimal $\vec{i}_2 \in \mathrm{I}_{\vec{p}}(Q)$ (attaining $\min \phi(\mathrm{I}_{\vec{p}}(Q))$); then defining $\vec{i} := \vec{i}_1 + \vec{i}_2$, I **Claim** the sum

$$\partial^{\vec{i}}(PQ) = \sum_{\vec{i}'_1 + \vec{i}'_2 = \vec{i}} [\text{coeff}](\partial^{\vec{i}'_1} P)(\partial^{\vec{i}'_1} Q)$$

has only the non-zero term $(\partial^{\vec{i}_1} P)(\partial^{\vec{i}_1} Q)$ (so no cancellation can happen, and so $\vec{i}_1 + \vec{i}_2 \in \mathrm{I}_{\vec{p}}(PQ)$, which proves the desired inequality).

Reason for Claim: first, $\vec{i}'_1 + \vec{i}'_2 = \vec{i} \implies \phi(\vec{i}'_1) + \phi(\vec{i}'_2) = \phi(\vec{i})$, so $\vec{i}'_1, \vec{i}'_2$ resp. must attain $\min \phi(\mathrm{I}_{\vec{p}}(P))$, $\min \phi(\mathrm{I}_{\vec{p}}(Q))$. Next, lex-order on $\mathbb{N}^n$ is total, so $\vec{i}'_1$ either $<, =, > \vec{i}_1$.

- If $<$, then that term must be 0 because $\vec{i}_1$ was chosen lex-minimal (amongst the subset of index-vectors $\mathrm{I}_{\vec{p}}(P)$ attaining $\min \phi(\mathrm{I}_{\vec{p}}(P))$ — a subset in which $\vec{i}_1$ also resides).
- If $>$, then must be that $\vec{i}'_2 < \vec{i}_2$ — o.w. because (weak and strong inequalities for) lex-$<$ on $\mathbb{N}^n$ preserved by addition in $\mathbb{N}^n$, we would get
$$\left.\begin{array}{l} \vec{i}'_1 > \vec{i}_1 \\ \vec{i}'_2 \geqslant \vec{i}_2 \end{array}\right\} \implies \vec{i}'_1 + \vec{i}'_2 > \vec{i}_1 + \vec{i}_2, \lightning$$
- The only case left is $=$, so the Claim is proven.

3.2. Appx. on Wronskians

3.2.1. 1-var.— Claim: For polys $g_j \in \mathbb{Q}[z]$, $[W \equiv 0] \implies$ LD. Slightly more work — but just linear algebra:

Reason: Fix $g_0, \ldots, g_{l-1} \in \mathbb{Q}[z]$. Let d be max degree, so $g_0, \ldots, g_{l-1}$ in $(d+1)$-dim $\mathbb{Q}$-VS V^{d+1} of degree $\leqslant d$ polys $\in \mathbb{Q}[z]$.

[Det=0] implies [rows of above matrix LD] implies there are $c_0, \ldots, c_{l-1} \in \mathbb{Q}(z)$ s.t. all g_j solve the ODE $[c_0(z)\partial_0 + \ldots + c_{l-1}(z)\partial_{l-1}]g = 0$. Can multiply through denoms to get the same conclusion but for some $c_0, \ldots, c_{l-1} \in \mathbb{Z}[z]$. I.e. all g_j are in kernel of the linear transf $\langle \vec{c}(z), \vec{\partial} \rangle : V^{d+1} \to \mathbb{Q}[z]$.

Choose z_0 s.t. $c_{l-1}(z_0) \neq 0$. Suppose polynomial $g \in \mathbb{Q}[z]$ solves the ODE, with " ODE initial data" $\partial_0 g(z_0), \ldots, \partial_{l-2} g(z_0)$ pre-determined.

Then all higher derivatives of g at z_0 can be determined, e.g. divide the following equation by $c_{l-1}(z_0)$

$$c_{l-1}(z_0) \cdot \partial_{l-1} g(z_0) := -[c_0(z_0)\partial_0 + \ldots + c_{l-2}(z_0)\partial_{l-2}](g)(z_0)$$

and so on differentiating both sides (divide by $c_{l-1}(z_0)$ only *after* differentiating).

By Taylor expansion centered at z_0, this determines $g(z) \in \mathbb{Q}[z]$.

So dimension of kernel of above $\mathbb{Q}$-lin. transf. is $\leqslant l-1$ (determined by " ODE initial data"). Thus showing $g_0, \ldots, g_{l-1}$ are $\mathbb{Q}$-LD.

This argument is my own, could be wrong. Google query "wronskian polynomial linear independence" $\rightsquigarrow$ Bostan-Dumas https://arxiv.org/pdf/1301.6598, good resource for this Wronskian stuff. Pottmeyer2022 writeup of Roth's thm. uses ODE theory, and Boston-Dumas cite Hurewicz1958 textbook Chp.3§7 which also relies on ODE theory; which is what inspired my above argument, essentially uniqueness half (only for poly solutions to the ODE) of existence-uniqueness for solutions to general linear homog. ODEs. Gaurish pg.50 uses more elementary but calculational arguments.

3.2.2. Multi-variable.— Let me try to explain Bostan-Dumas Thm3 idea. First, $g_0, \ldots, g_{l-1} \in \mathbb{Q}[\vec{z}]$ are $\mathbb{Q}$-LI implies can do procedure:

- start with g_0
- if g_1 has same lowest-order-monomial (LOM) $\vec{z}^{\vec{e}}$ (lowest w.r.t. lex order by $\vec{e} \in \mathbb{N}^h$) as g_0, then replace g_1 by $g_1 - [\mathbb{Q}\text{-const}]g_0$ so g_1 has strictly larger LOM than g_0.
- continue this process; guaranteed the replaced $g_0, \tilde{g}_1, \dots, \tilde{g}_{l-1}$ still $\mathbb{Q}$-LI.

These correspond to (invertible) col operations on G-matrix, which don't change whether or not det $= 0$.

Thus, it suffices to prove: for distinct-LOM $g_0, \dots, g_{l-1} \in \mathbb{Q}[\vec{z}]$, [$\mathbb{Q}$-LI] $\implies$ [some gen. Wronsk. $G \not\equiv 0$].

I claim it suffices to prove (s.t.p.) that for (distinct) monomials $g_0 := \vec{z}^{\vec{e}_0}, \dots, g_{l-1} := \vec{z}^{\vec{e}_{l-1}} \in \mathbb{Q}[\vec{z}]$, [some generalized Wronskian G — which (cf. Bostan-Dumas Lemma1) must itself be a monomial $\vec{z}^{\vec{e}_0} \dots \vec{z}^{\vec{e}_{l-1}} \cdot \vec{z}^{(-1)\vec{i}_0} \dots \vec{z}^{(-1)\vec{i}_{l-1}} =: \vec{z}^{\clubsuit}$ (though if a coord of $\clubsuit$ is $\leqslant 0$, interpreted as 0) — has non-zero coeff].

Reason for this "s.t.p." is that using multilinearity of det,

$$\det\left[\begin{pmatrix}\mathcal{D}_0\\ \mathcal{D}_1\\ \vdots\\ \mathcal{D}_{l-1}\end{pmatrix}(\underbrace{\qquad g_0 \qquad}_{\vec{z}^{\vec{e}_0}+[\text{higherOrd}]}, \dots, g_{l-1})\right] = \det\begin{bmatrix}\text{matrix coming}\\ \text{from row-vec}\\ \text{of only LOMs}\\ (\vec{z}^{\vec{e}_0}, \dots, \vec{z}^{\vec{e}_{l-1}})\end{bmatrix} + \sum_{\{0,1\}^l \setminus \vec{0}} \det\begin{bmatrix}\text{matrix coming from}\\ \text{row-vec with higherOrd}\\ \text{terms in place of LOM}\\ \text{as indicated by positions}\\ \text{of 1's in current corner}\\ \text{of the hypercube } \{0,1\}^l\end{bmatrix}$$

where LOM of RHS comes from the 1stSummand=[non-zero const]$\vec{z}^{\clubsuit}$ (one way to see this is break into even larger sum of dets-of-monomials, where the resulting dets are (perhaps coeff=0) monomials $\vec{z}^{\spadesuit}$ with $\spadesuit > \clubsuit$).

Ok, finally we can just work with monomials. Assume f.s.o.c. ALL the dets of the following form are 0:

$$\det\left[\begin{pmatrix}\mathcal{D}_{\vec{i}_0}\\ \mathcal{D}_{\vec{i}_1}\\ \vdots\\ \mathcal{D}_{\vec{i}_{l-1}}\end{pmatrix}(\vec{z}^{\vec{e}_0}, \dots, \vec{z}^{\vec{e}_{l-1}})\right] = \det\left[\begin{pmatrix}!_{\vec{i}_0}\\ !_{\vec{i}_1}\\ \vdots\\ !_{\vec{i}_{l-1}}\end{pmatrix}(\vec{e}_0, \dots, \vec{e}_{l-1})\right] \cdot [\text{if need, const}\neq 0]\cdot\text{monomial } (z^{\clubsuit})$$

where $!_{\vec{i}}$ is multi-index falling factorial operator, e.g. $!_{(2,1)}$ applied to $(7, 13)$ is $(7 \cdot 6) \cdot 13$.

(Generalized/multi-index) Vandermonde matrix is same by instead of $!_{\vec{i}}$, it's power operator $^{\backslash \vec{i}}$, e.g. $^{\backslash (2,1)}$ applied to $(7, 13)$ is $(7, 13)^{(2,1)} := 7^2 13^1$ (classical multi-index notation).

The final piece of Bostan-Dumas Thm3 is to consider Vandermonde determinant ($\vec{u}$ is (len l) new indeterminates/variables)

$$V(\langle\vec{u}, \vec{e}_0\rangle, \dots, \langle\vec{u}, \vec{e}_{l-1}\rangle) := \det\left[\begin{pmatrix}^{\backslash 0}\\ ^{\backslash 1}\\ \vdots\\ ^{\backslash l-1}\end{pmatrix}(\langle\vec{u}, \vec{e}_0\rangle, \dots, \langle\vec{u}, \vec{e}_{l-1}\rangle)\right] = \det\left[\begin{pmatrix}\langle\vec{u}, __\rangle^0\\ \langle\vec{u}, __\rangle^1\\ \vdots\\ \langle\vec{u}, _\rangle^{l-1}\end{pmatrix}(\vec{e}_0, \dots, \vec{e}_{l-1})\right].$$

Example calculation: $\langle(u_1, u_2), (7, 13)\rangle^3 = u_1^3(7^3) + 3u_1^2u_2(7^2 13^1) + 3u_1u_2^2(7^1 13^2) + u_2^3(13^3)$. In general, we have equality of operators

$$\langle\vec{u}, __\rangle^r = \sum_{|\vec{i}|=r} [\mathbb{N}\text{-coeff}] \cdot \vec{u}^{\vec{i}} \cdot {}^{\backslash \vec{i}}.$$

The conclusion is the above Vandermonde det is homog poly in $\vec{u}$ with coefficients that are $\mathbb{N}$-linear combination of determinants

$$\det\left[\begin{pmatrix} {}^{\wedge}\vec{i}_0 \\ {}^{\wedge}\vec{i}_1 \\ \vdots \\ {}^{\wedge}\vec{i}_{l-1} \end{pmatrix}(\vec{e}_0,\dots,\vec{e}_{l-1})\right].$$

And these pow/hat determinants are themselves $\mathbb{Z}$-linear combinations of the dets with falling factorials. Example:

$$!_{(2,1)}+!_{(1,1)} = {}^{\wedge}{}_{(2,1)} \quad (\text{e.g. } (7\cdot 6)\cdot 13 + (7)\cdot 13 = (7\cdot 7)\cdot 13), \qquad !_{(3,1)} + 3{\cdot}!_{(2,1)}+!_{(1,1)} = {}^{\wedge}{}_{(3,1)}$$

(the latter because $t(t-1)(t-2)+3t(t-1)+t = t^3$ — can always make such identities by starting with e.g. $t(t-1)(t-2)$, the finding the correct $\mathbb{Z}$-coeff for $t(t-1)$ to kill the t^2 term, etc.). Also 2nd-coord "1" can be replaced by any other nat $\in \mathbb{N}$; all the combinations we did were about the 1st-coord.

Another example (combining the previous two hehe)

$$!_{(3,2)} + 3{\cdot}!_{(2,2)}+!_{(1,2)} = {}^{\wedge}{}_{(3,0)}\circ !_{(0,2)} \quad (=!_{(0,2)}\circ {}^{\wedge}{}_{(3,0)}), \qquad {}^{\wedge}{}_{(3,2)} = {}^{\wedge}{}_{(3,0)}\circ !_{(0,2)} + \underbrace{{}^{\wedge}{}_{(3,0)}\circ !_{(0,1)}}_{={}^{\wedge}{}_{(3,1)}}$$

And one more to hammer in the point:

$$\begin{aligned} {}^{\wedge}{}_{(3,3)} &= {}^{\wedge}{}_{(3,0)}\circ !_{(0,3)} + {}^{\wedge}{}_{(3,0)} \circ 3{\cdot}!_{(0,2)} + {}^{\wedge}{}_{(3,0)}\circ !_{(0,1)}\cdot \\ &= (!_{(3,3)} + 3{\cdot}!_{(2,3)}+!_{(1,3)}) + 3(!_{(3,2)} + 3{\cdot}!_{(2,2)}+!_{(1,2)}) + (!_{(3,1)} + 3{\cdot}!_{(2,1)}+!_{(1,1)}) \end{aligned}$$

So finale: can linear combinate to go from dets with $!_{\vec{i}}$, to dets with ${}^{\wedge}\vec{i}$, to the Vandermonde det, V, above. With the above assumption f.s.o.c., get that $V \equiv 0$. Vandermonde theory says at least 2 of the $\vec{e}_0,\dots,\vec{e}_{l-1}$ are equal; ※.