



Sharp estimates for trilinear oscillatory integrals and an algorithm of two-dimensional resolution of singularities

Lechao Xiao

Abstract. We obtain sharp estimates for certain trilinear oscillatory integrals, extending Phong and Stein’s seminal result to a trilinear setting. This result partially answers a question raised by Christ, Li, Tao and Thiele, concerning sharp estimates for certain multilinear oscillatory integrals. The method in this paper relies on a self-contained algorithm of resolution of singularities in \mathbb{R}^2 , which may be of independent interest.

1. Introduction

The purpose of this paper is to study sharp decay estimates of the following trilinear oscillatory integrals:

$$\Lambda_S(f_1, f_2, f_3) = \iint e^{i\lambda S(x,y)} f_1(x) f_2(y) f_3(x+y) a(x,y) dx dy,$$

where $a \in C_0^\infty(\mathbb{R}^2)$ is a smooth cut-off function supported in a small neighborhood of the origin, and the phase S is real analytic.

1.1. Background

Consider the following oscillatory integral operator:

$$(1.1) \quad T(f)(x) = \int e^{i\lambda S(x,y)} f(y) a(x,y) dy,$$

where $S(x,y)$ is a smooth real-valued function in $\mathbb{R}^n \times \mathbb{R}^n$ and $a(x,y)$ is a smooth cut-off function.

Mathematics Subject Classification (2010): Primary 42B20; Secondary 14E15.

Keywords: Multilinear operators, oscillatory integrals, resolution of singularities, Newton polyhedra.

One of the central topics in oscillatory integrals is the study of the asymptotic behavior of $\|T\|_{2 \rightarrow 2}$ as $|\lambda| \rightarrow \infty$. Equivalently, from a dual point of view, can one find the optimal decay $C(\lambda)$ such that

$$|\Lambda_2(f, g)| \leq C(\lambda) \|f\|_2 \|g\|_2,$$

where the bilinear form $\Lambda_2(f, g)$ is given by

$$(1.2) \quad \Lambda_2(f, g) = \langle T(f), g \rangle = \iint e^{i\lambda S(x, y)} f(y) \overline{g(x)} a(x, y) dx dy?$$

During the last decades, this problem and related topics have been extensively studied by many authors. Fruitful results have been obtained via rich techniques. We begin with a classical result from Hörmander [12], concerning sharp L^2 estimates of (1.1) when S is non-degenerate.

Theorem 1.1 (Hörmander, [12]). *Assume $a(x, y)$ is a smooth cut-off function supported in a neighborhood of 0 and $S(x, y)$ is a smooth function such that*

$$(1.3) \quad \left| \det \frac{\partial^2 S}{\partial x \partial y} \right| \geq 1, \quad \text{for all } (x, y) \in \text{supp } (a).$$

Then one has

$$(1.4) \quad \|T(f)\|_2 \leq C |\lambda|^{-n/2} \|f\|_2.$$

Establishing sharp estimates in a more general setting, in particular when $S(x, y)$ is degenerate, was significantly more difficult. Until the early '90s, by the seminal works of Phong and Stein [14], [15], [16], a full understanding of (1.1) was obtained when S is a real analytic function of two variables. In their works, a systematic treatment was introduced to deal with the degenerate setting. The key ingredient to characterize the sharp decay rate is the geometric concept: Newton polyhedra.

Definition 1.2. Assume the Taylor series expansion of $S(x, y)$ is given by

$$(1.5) \quad S(x, y) = \sum_{p, q \in \mathbb{N}} c_{p, q} x^p y^q.$$

The Newton polyhedron of S is defined as

$$\mathcal{N}(S) = \text{Conv}(\cup_{p, q} \{(u, v) \in \mathbb{R}^2 : u \geq p, v \geq q \text{ and } c_{p, q} \neq 0\}),$$

where $\text{Conv}(X)$ denotes the convex hull of a set $X \subset \mathbb{R}^2$. The Newton distance of $\mathcal{N}(S)$ is defined to be

$$\delta_S = \inf\{t \in \mathbb{R} : (t, t) \in \mathcal{N}(S)\}.$$

Theorem 1.3 (Phong–Stein, [16]). *Let $S(x, y)$ be real-analytic and assume the support of $a(x, y)$ is contained in a sufficiently small neighborhood of 0 in \mathbb{R}^2 . Then*

$$\|T(f)\|_2 \leq C |\lambda|^{-\frac{1}{2(1+\delta)}} \|f\|_2,$$

where δ is the Newton distance of $\mathcal{N}(\partial_x \partial_y S)$.

It has been shown by the classic work of Varchenko, confirming earlier hypotheses of Arnold, that the Newton polyhedron is the appropriate notion to characterize the decay rate of scalar oscillatory integrals [20].

1.2. Motivation

In this paper, we study some trilinear analogues of the above problems. Set

$$(1.6) \quad \Lambda_S(f_1, f_2, f_3) = \iint e^{i\lambda S(x,y)} f_1(x) f_2(y) f_3(x+y) a(x,y) dx dy.$$

We want to characterize the optimal constant ϵ such that the following is true:

$$|\Lambda_S(f_1, f_2, f_3)| \leq C |\lambda|^{-\epsilon} \|f_1\|_2 \|f_2\|_2 \|f_3\|_2,$$

for some constant C independent of f_1, f_2, f_3 and λ .

The study of the trilinear form (1.6) is also motivated by the work of Christ, Li, Tao and Thiele [3], where certain multilinear oscillatory integrals were studied in a very general setting.

To formulate the questions posed in [3], we need some preliminary notations. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)$, where each $\pi_j: \mathbb{R}^n \rightarrow \mathbb{R}^{n_j} \subset \mathbb{R}^n$ is a surjective linear projection. Let $S: \mathbb{R}^n \rightarrow \mathbb{R}$ be a polynomial and $a(x)$ be a smooth cut-off function supported in a small neighborhood of $0 \in \mathbb{R}^n$. For each j , let $f_j: \mathbb{R}^{n_j} \rightarrow \mathbb{C}$ be a measurable function. Consider the following multilinear form:

$$(1.7) \quad \Lambda_{S,\boldsymbol{\pi}}(f_1, f_2, \dots, f_J) = \int e^{i\lambda S(x)} a(x) \prod_{j=1}^J f_j \circ \pi_j(x) dx.$$

Q1: For what kind of input $(S, \boldsymbol{\pi})$, the following is true:

$$(1.8) \quad |\Lambda_{S,\boldsymbol{\pi}}(f_1, f_2, \dots, f_J)| \leq C |\lambda|^{-\delta} \prod_{j=1}^J \|f_j\|_{p_j}$$

for some $\delta > 0$, some $\mathbf{p} = (p_1, \dots, p_J) \in [1, \infty]^J$ and all $f_j \in L^{p_j}(\mathbb{R}^{n_j})$?

Q2: If **Q1** could be answered affirmatively, what is the optimal exponent δ ?

Giving a complete answer to **Q1** for a most general input $(S, \boldsymbol{\pi})$ is quite challenging. Still, an affirmative answer to **Q1** was given in [3] under certain dimension assumptions on $\boldsymbol{\pi}$. For further progress of **Q1**, we refer the reader to [1], [2], [4], and [8].

For **Q2**, some results were already known. For instance, when $J = 2$, $n_1 = n_2 = n/2$ (assume n is even) and S is smooth, Theorem 1.1 provides a sufficient characterization when the best possible decay can be obtained. Theorem 1.3 settled the case $n = J = 2$ and S is an arbitrary analytic function; see [7], [18], [19] for $S \in C^\infty(\mathbb{R}^2)$. For $n = J \geq 2$ and S is a polynomial, almost sharp estimates (probably up to a power of $\log |\lambda|$) were known, by the work of Phong, Stein and Sturm [17].

In Christ, Li, Tao and Thiele's attempt to answer **Q1**, an important step is a reduction to the trilinear setting. Thus, it is helpful to fully understand the trilinear case, in particular to determine the optimal exponent in (1.8) in this setting. This motivates us to study the sharp estimate of the trilinear form (1.6), which corresponds to the case $n = 2$, $J = 3$, S is analytic and $\boldsymbol{\pi} = \boldsymbol{\pi}_0$, where

$$(1.9) \quad \boldsymbol{\pi}_0(x, y) = (\pi_{01}(x, y), \pi_{02}(x, y), \pi_{03}(x, y)) = (x, y, x + y).$$

Indeed, a more general setting

$$(1.10) \quad \Lambda_{S, \boldsymbol{\pi}}(f_1, f_2, f_3) = \iint e^{i\lambda S(x, y)} a(x, y) \prod_{j=1}^3 f_j \circ \pi_j(x, y) dx dy,$$

can be reduced to (1.6) via an invertible linear transformation in \mathbb{R}^2 (see Section 2), where $\pi_j : \mathbb{R}^2 \rightarrow \mathbb{R}$ are pairwise linearly independent projections for $j = 1, 2, 3$.

One necessary condition for (1.10) to possess a decay bound is that S should be non-degenerate relative to $\boldsymbol{\pi}$, in the sense that $S(x, y)$ cannot be represented as a sum of functions of $\{\pi_j(x, y)\}$. Otherwise, if

$$S(x, y) = \sum_{1 \leq j \leq 3} S_j \circ \pi_j(x, y),$$

then we can incorporate each $e^{i\lambda S_j \circ \pi_j(x, y)}$ into $f_j \circ \pi_j(x, y)$ by setting

$$\tilde{f}_j \circ \pi_j(x, y) = e^{i\lambda S_j \circ \pi_j(x, y)} f_j \circ \pi_j(x, y).$$

Since $\|\tilde{f}_j\|_{p_j} = \|f_j\|_{p_j}$, one cannot expect any decay as in (1.8).

Let $\pi_j^\perp : \mathbb{R}^2 \rightarrow \mathbb{R} \subset \mathbb{R}^2$ be linear projections such that $\pi_j \circ \pi_j^\perp = 0$ and $\|\pi_j^\perp\|_2 = 1$. Set $\boldsymbol{\pi}^\perp = (\pi_1^\perp, \pi_2^\perp, \pi_3^\perp)$ and $D_{\boldsymbol{\pi}^\perp} = \prod_{j=1}^3 \pi_j^\perp \cdot \nabla$. Then S is called simply degenerate relative to $\boldsymbol{\pi}$ if $D_{\boldsymbol{\pi}^\perp} S \equiv 0$ [3], otherwise S is called simply non-degenerate relative to $\boldsymbol{\pi}$. In addition, S is simply degenerate at a point (x_0, y_0) if $D_{\boldsymbol{\pi}^\perp} S(x_0, y_0) = 0$. Simple non-degeneracy implies non-degeneracy and the converse is not true in general. But in our case, they are equivalent; see Proposition 3.1 in [3].

1.3. Results

The following theorem, extending Theorem 1.1 to the trilinear setting when $n = 1$, states that if S is simply non-degenerate everywhere in $\text{Conv}(\text{supp}(a))$, then one can obtain the optimal bound of (1.10).

Theorem 1.4. *Assume $a(x, y)$ is a smooth cut-off function supported in a neighborhood of $0 \in \mathbb{R}^2$ and $S(x, y)$ is smooth such that*

$$(1.11) \quad |D_{\boldsymbol{\pi}^\perp} S(x, y)| \geq 1 \quad \text{for all } (x, y) \in \text{Conv}(\text{supp}(a)).$$

Then

$$(1.12) \quad |\Lambda_{S, \boldsymbol{\pi}}(f_1, f_2, f_3)| \leq C |\lambda|^{-1/6} \prod_{j=1}^3 \|f_j\|_2.$$

This theorem is indeed implicitly proved by Li [13]. We also extend Theorem 1.3 to the trilinear form (1.10). Different to what was expected, the characterization of the sharp exponent in this case is distinct from the ones in Varchenko's or Phong–Stein's results. Instead, it is described algebraically by the relative multiplicity of S . Write S as a sum of homogeneous polynomials $S(x, y) = \sum_i S_i(x, y)$, where each S_i is a homogeneous polynomial of degree i . The multiplicity of S , i.e., the order of the zero of the function S at the origin, is

$$\text{mult}(S) = \min\{i : S_i \neq 0\}.$$

We also adopt the convention that $\text{mult}(S) = -\infty$ if $S = 0$. The multiplicity of S relative to π is defined as

$$\text{mult}_\pi(S) = \min\{i : D_{\pi^\perp} S_i \neq 0\} = \text{mult}(D_{\pi^\perp} S) + 3,$$

which is the multiplicity of the quotient of S by the class of degenerate analytic functions. Notice that if S is simply degenerate, then $\text{mult}_\pi(S) = -\infty$. One of the two main results of this paper is:

Theorem 1.5. *Assume $S(x, y)$ is a real analytic function and the support of $a(x, y)$ is sufficiently small. Then*

$$(1.13) \quad |\Lambda_{S,\pi}(f_1, f_2, f_3)| \leq C |\lambda|^{-1/(2\text{mult}_\pi(S))} \prod_{j=1}^3 \|f_j\|_2.$$

The result (1.13) is sharp in the sense that if $a(0, 0) \neq 0$, then

$$(1.14) \quad |\Lambda_{S,\pi}(f_1, f_2, f_3)| \geq C' |\lambda|^{-1/(2\text{mult}_\pi(S))} \prod_{j=1}^3 \|f_j\|_2,$$

as $|\lambda| \rightarrow \infty$, for some $C' > 0$ and some $\{f_j\}_{1 \leq j \leq 3}$.

Existence of a (non-sharp) decay rate in the bound of (1.13) was treated in [3]. What is new and interesting here is its sharpness and its connection to the given phase. It is intriguing to compare the above statement with Theorem 1.3:

1. Where do the exponents in both theorems come from?
2. Why do the exponents look so different?
3. Show that both exponents are sharp.

We provide our answers here since they are quite accessible. We begin with interpreting both exponents geometrically via the Newton polyhedron of P , where

$$(1.15) \quad P(x, y) = \partial_x \partial_y (\partial_x - \partial_y) S(x, y)$$

in the trilinear setting, and

$$(1.16) \quad P(x, y) = \partial_x \partial_y S(x, y)$$

in the bilinear setting. Index the vertices of $\mathcal{N}(P)$ from left to right by $V_1 = (p_1, q_1)$, $V_2 = (p_2, q_2)$, \dots , $V_k = (p_k, q_k)$, the compact edges by $E_1 = V_1V_2$, $E_2 = V_2V_3$, \dots , $E_{k-1} = V_{k-1}V_k$, and the vertical and horizontal edges by E_0 and E_k . The set of faces, denoted by $\mathcal{F}(P)$, is the union of the set of vertices and the set of edges. The set of supporting lines of $\mathcal{N}(P)$, denoted by $\mathcal{SL}(P)$ are the lines that intersect the boundary of $\mathcal{N}(P)$ and do not intersect any other points of $\mathcal{N}(P)$. There is a one-to-one correspondence \mathcal{M} between $\mathcal{SL}(P)$ and $[0, \infty]$, given by defining $\mathcal{M}(L)$ to be the negative reciprocal of the slope of L for each $L \in \mathcal{SL}(P)$. We often use the notation $L_m \in \mathcal{SL}(P)$ to refer the supporting line $\mathcal{M}(L_m) = m$. Since each E_j lies in exactly one supporting line, we also use E_j to denote that line. Let $m_j = \mathcal{M}(E_j)$, then

$$0 = m_0 < m_1 < m_2 < \dots < m_{k-1} < m_k = \infty.$$

We assign two decay constants d_L and δ_L to each $L \in \mathcal{SL}(P)$ as follows:

- (1) $d_L = \min\{d_{L,x}, d_{L,y}\}$, the minimum of the x -intercept and the y -intercept of L ;
- (2) (δ_L, δ_L) is the intersection of L and the bisecting line $p = q$.

We associate each $L_m \in \mathcal{SL}(P)$ in the (p, q) -plane with a region $|y| \sim |x|^m$ in the (x, y) -plane, which will be known as the ‘good’ region defined by L_m ; see Section 4. The two decay constants assigned to $L \in \mathcal{SL}(P)$ are then associated to the decay rates of the trilinear and bilinear forms when the support of the cut-off function $a(x, y)$ is localized to the associated region of L .

Next we assign the corresponding decay constants to each face F as follows:

$$d_F = \sup d_L \quad \text{and} \quad \delta_F = \sup \delta_L,$$

where both supremums range over all $L \in \mathcal{SL}(P)$ containing F . Then

$$\text{mult}(P) = \max\{d_F : F \in \mathcal{F}(P)\} = \sup\{d_L : L \in \mathcal{SL}(P)\}$$

and

$$\delta_P = \max\{\delta_F : F \in \mathcal{F}(P)\} = \sup\{\delta_L : L \in \mathcal{SL}(P)\}.$$

For simplicity, we restrict our attention to the non-vertical edges. Let $E = E_j$, $m = m_j$ and $(p, q) = (p_j, q_j)$ for some j . In the trilinear setting, there are subtle differences between $m \leq 1$ and $m \geq 1$. In the case $m \leq 1$, one has $d_E = d_{E,x} = p + mq$. The value $|\lambda|^{-\frac{1}{2(3+d_E)}}$ corresponds to the bound of Λ_S when restricting $\text{supp}(a)$ to the region associated to E . If $m \geq 1$, then $d_E = d_{E,y} \geq p/m + q$. When restricting $\text{supp}(a)$ to the regions associated to E , the corresponding bound of Λ_S is $|\lambda|^{-\frac{1}{2(3+d_E)}}$; see Proposition 5.2 and Proposition 5.4. In general, the bound of Λ_S is $|\lambda|^{-\frac{1}{2(3+d_F)}}$, when $\text{supp}(a)$ is restricted to the region associated to F . However, such difference does not arise in the bilinear setting and we only need one expression for δ_E which is $\delta_E = (p + mq)/(1 + m)$. In general, the bound for the bilinear form is $|\lambda|^{-\frac{1}{2(1+\delta_F)}}$, when $\text{supp}(a)$ is restricted to the region associated to F .

For the second question, the noticeable difference between the operators is the extra term $f_3(x+y)$ in the trilinear form. If x and y vary in intervals of length δ_1 and δ_2 respectively, then $(x+y)$ varies in an interval of length about $\max\{\delta_1, \delta_2\}$. If f_3 is a characteristic function supported in this interval, then $\|f_3\|_2 \sim \max\{\delta_1^{1/2}, \delta_2^{1/2}\}$. Heuristically, this freezes the ratio $\log|\delta_1|/\log|\delta_2|$ to be 1, if one wants to optimize the bound of the trilinear form. However, without the extra term $f_3(x+y)$, the ratio $\log|\delta_1|/\log|\delta_2|$ is totally free. The effect of such difference between the operators is realized by the difference between the following two Schur's type lemmas:

Lemma 1.6. *Assume $a(x, y)$ is a measurable function supported in a strip of x -width no more than δ_1 and y -width no more than δ_2 . Assume $\|a\|_\infty \leq 1$. Then*

$$(1.17) \quad \left| \iint f_1(x)f_2(y)a(x, y) dx dy \right| \leq C(\delta_1\delta_2)^{1/2} \|f_1\|_2 \|f_2\|_2.$$

Lemma 1.7. *The assumptions on $a(x, y)$ are the same as in Lemma 1.6. Then*

$$(1.18) \quad \left| \iint f_1(x)f_2(y)f_3(x+y)a(x, y) dx dy \right| \leq C \min\{\delta_1^{1/2}, \delta_2^{1/2}\} \|f_1\|_2 \|f_2\|_2 \|f_3\|_2.$$

Lemma 1.6 is employed in [16] to control the norm of the operator in Theorem 1.3 when the phase fails to provide sufficient decay. Lemma 1.7 plays the same role in our proof; see Section 2 for its proof.

In Lemma 1.6, the two parameters δ_1 and δ_2 are symmetric, which illustrates why there is no constraint on the ratio $\log|\delta_1|/\log|\delta_2|$. Indeed, the sharpness of Theorem 1.3 can be verified by setting $f_1 = \mathbf{1}_{[0, \delta_1]}$ and $f_2 = \mathbf{1}_{[0, \delta_1^m]}$ with an appropriate choice of δ_1 and with m being the negative reciprocal of the slope of any supporting lines containing the so-called main face, i.e., the edge that intersects the bisecting line or the vertex that lies on it.

However, such symmetry breaks down in the trilinear setting, even though the parameters x and y appear symmetrically in Λ_S . One should think of (1.18) as two different estimates: the first one with the bound $C\delta_1^{1/2}$ and the second one with $C\delta_2^{1/2}$. Consequently, they will lead to two different bounds for (1.6) accordingly. The first (second) bound comes from employing the first (second) estimate of (1.18) and Theorem 2.1, with decay exponents represented in terms of x -intercepts $d_{E_j, x}$ (y -intercepts $d_{E_j, y}$). This explains why we need to split the range of m into two cases: $m \leq 1$ and $m \geq 1$. However, these two bounds coincide when the x -intercept and the y -intercept are equal, i.e., $m = 1$. In the picture of the Newton polyhedron, this corresponds to the supporting line of slope -1 , which is given by the equation $p + q = \text{mult}(P)$ with $|y| \sim |x|$ as the associated region. This illustrates how the relative multiplicity comes into play and why the sharpness of Λ_S is obtained when $|y| \sim |x|$.

Now we come to the sharpness of the trilinear form. Write S as a sum of homogeneous polynomials:

$$S(x, y) = \sum_{n=n_0}^{\infty} S_n(x, y).$$

Without loss of generality, we assume n_0 is the relative multiplicity of S , i.e.,

$$(1.19) \quad \partial_x \partial_y (\partial_x - \partial_y) S_{n_0} \neq 0;$$

see Section 1.2.

Let $f_1 = f_2$ be characteristic functions of $I_A = [-\lambda^{-1/n_0}/A, \lambda^{-1/n_0}/A]$, and let f_3 be the characteristic function of $[-2\lambda^{-1/n_0}/A, 2\lambda^{-1/n_0}/A]$, where $A > 0$ is a number independent of λ such that

$$|\lambda S(x, y)| \leq 2^{-100}, \quad \text{for all } x, y \in I_A.$$

Then

$$\left| \iint e^{i\lambda S(x, y)} a(x, y) f_1(x) f_2(y) f_3(x + y) dx dy \right| \sim |I_A| \times |I_A| \sim \lambda^{-2/n_0}.$$

Notice that

$$\|f_1\|_2 \sim \|f_2\|_2 \sim \|f_3\|_2 \sim \lambda^{-1/(2n_0)}.$$

Hence, if

$$(1.20) \quad \left| \iint e^{i\lambda S(x, y)} a(x, y) f_1(x) f_2(y) f_3(x + y) dx dy \right| \lesssim C(\lambda) \prod_{j=1}^3 \|f_j\|_2,$$

then $C(\lambda) \gtrsim \lambda^{-1/(2n_0)} = \lambda^{-1/(2\text{mult}_{\pi_0}(S))}$ as desired.

1.4. Methods

Like Phong and Stein's proof of Theorem 1.3, the proof of Theorem 1.5 requires elaborate analysis. There are two main ingredients in their proof:

- (1) the operator version of the van der Corput lemma [15]; see Theorem 2.2, and
- (2) the Weierstrass preparation theorem (WPT) and the Puiseux expansion.

In order to extend Phong and Stein's framework to the trilinear setting, we first establish the trilinear analogue of (1):

- (1') Theorem 2.1: trilinear version of Phong–Stein's operator van der Corput lemma.

In addition, we develop:

- (2') a self-contained algorithm of resolution of singularities in \mathbb{R}^2 , which is our second main result. We also use the WPT in the proof of the algorithm, but its use is not essential. One can use the implicit function theorem instead.

Theorem 1.8. *For each analytic function P defined in a neighborhood of $0 \in \mathbb{R}^2$, there is an open set U containing the origin, such that up to a set of measure zero, one can partition U into a finite collection of regions $\{V_k\}_{1 \leq k \leq K}$, such that P*

behaves almost like a monomial in each V_k in the following sense. There is an integer $M \in \mathbb{N}$, and for each k there is a one-to-one function

$$\begin{aligned} \rho_k : V_k &\rightarrow \rho_k(V_k) \\ (x, y) &\mapsto (x_k, y_k) \end{aligned}$$

satisfying the following properties:

$$(1.21) \quad P(x, y) = P_k(x_k, y_k) = x_k^{p_k} y_k^{q_k} \cdot Q_k(x_k, y_k) \quad \text{for all } (x, y) \in V_k,$$

where

(i) $(x_k, y_k) = \rho_k(x, y)$ is given by

$$(1.22) \quad \begin{cases} x = x_k \\ y = \gamma_k(|x_k|^{1/M}) + |x_k|^{M_k/M} y_k, \end{cases}$$

where $M_k \in \mathbb{N}$ and γ_k is a polynomial, unless $P(x, \gamma_k(|x|^{1/M})) = 0$, then γ_k is a convergent power series.

(ii) $P_k = P \circ \rho_k^{-1}$ and (p_k, q_k) is a vertex of the Newton polyhedron of P_k ;

(iii) The function Q_k is smooth and nonvanishing near 0 in $\rho_k(V_k)$, i.e.,

$$\lim_{(x_k, y_k) \rightarrow (0,0)} Q_k(x_k, y_k) \neq 0 \quad \text{inside } \rho_k(V_k);$$

(iv) $\rho_k(V_k)$ (as well as V_k) is a curved triangular region, whose upper and lower boundaries are given by $y_k = C_k |x_k|^{m_k}$ and $y_k = C'_k |x_k|^{m'_k}$, for some $0 \leq m_k, m'_k \leq \infty$ with $m_k M, m'_k M \in \mathbb{N} \cup \{\infty\}$, and for some constants C_k, C'_k .

Moreover, the constants $m_k, m'_k, (p_k, q_k), M_k/M$ and the function¹ γ_k can be computed explicitly via the Newton polyhedra of $\{P_k\}_{1 \leq k \leq K}$.

Remark 1.9. See Theorem 4.11 in Section 4 for a complete version.

The idea of employing resolution of singularities to investigate oscillatory integrals appeared in Varchenko's work [20], where deep results from Hironaka [11] played a crucial role. In [16], to control the lower bound of $|\partial_x \partial_y S|$, one resolves the singularities of $\partial_x \partial_y S$ by means of Puiseux series expansions of roots. More recently, a direct algorithm of resolution of singularities in \mathbb{R}^2 was introduced by Greenblatt [6], where an elegant proof of Theorem 1.3 was presented based on this algorithm.

Our proofs of Theorem 1.5 and the algorithm here are both influenced by [6], as well as its predecessor [16]. Many of the ideas inside the algorithm here are very elementary and have even been known for centuries, which may come back to Newton's method, known as the Newton–Puiseux algorithm for solving $P(x, y) = 0$ by a fractional power series $y = y(x^{1/M})$, i.e., the Puiseux series; see [5]. The

¹In the case γ_k is an infinite series, we can compute any partial sum of γ_k .

philosophy of the algorithm here is similar to that of the one in [6]: one wants to decompose some neighborhood of a singular point of a given function into a finite collection of ‘good’ regions, in each of which the function behaves like a monomial. However, the algorithm here is distinct from the one in [6] in many aspects. One noticeable difference is that we do not need to employ the implicit function theorem, while it is a key for the termination of the algorithm in [6]. We refer the reader to Section 3 for the main ideas of the algorithm, comparisons to the one in [6] and some examples that implement the algorithm.

Concluding remarks

It is possible to generalize the simply non-degenerate case, i.e., Theorem 1.4 to higher dimensions, however such generalization for the degenerate case, i.e., Theorem 1.5 is extremely challenging. First, the equivalence between degeneracy and simple-degeneracy breaks down completely, and indeed the class of non-degenerate functions is significantly larger than that of simply non-degenerate functions. Even for proving the existence of decays, the case for higher dimensions is substantially more complicated; see [4]. Second, progress on degenerate oscillatory integrals in higher dimensions are slow and arduous, and even the higher dimension analogue of Theorem 1.3 is still far from being fully understood; see [9]. Perhaps a more approachable problem to consider is the n -linear analogue of Theorem 1.5. However, starting from $n = 4$, there won’t be any decay if one attempts to bound the n -linear form by the product of functions in $L^2(\mathbb{R})$. Thus, one needs to consider the functions in general $L^{p_j}(\mathbb{R})$ space. As a starting point, one should address **Q2** above for (1.6) for general p, q and r . Certain symmetry, which holds in the setting of n functions with n variables (see [17]), breaks down completely in this trilinear setting. Consequently, finding optimal decays for all triples (p, q, r) is substantially more complicated than that in [17]; see [10] for recent progress. Finally, we refer the reader to [21] for another interesting application of the resolution algorithm, where we obtain a complete characterization for the $L^p \rightarrow L^p$ mapping properties for one-dimensional oscillatory integral operators.

The organization of this paper is as follows. In Section 2, we reduce the trilinear form $\Lambda_{S,\pi}(f_1, f_2, f_3)$ to the special case $\Lambda_S(f_1, f_2, f_3)$ as in (1.6) and then prove Theorem 1.4 and Theorem 2.1. The latter one is the trilinear version of Phong–Stein’s van der Corput Lemma. In Section 3, we outline some main ideas of the algorithm. Model examples are also provided to illustrate how the algorithm is implemented. Details for the proof of the algorithm will appear in Section 4. The method is purely analytic. In Section 5, we apply the algorithm and Theorem 2.1 to prove Theorem 1.5.

Acknowledgements. The author wishes to thank his advisor Xiaochun Li for suggesting this problem. Without his valuable and insightful suggestions and enthusiastic guidance this work can not be done. The author would also like to thank Philip T. Gressman, the anonymous referees for many constructive comments, and Maxim Gilula for his help in mathematical writing.

2. Preliminaries

In this section, we employ the basic strategy from [13] to prove Theorem 1.4 and its local analogue, Theorem 2.1. At the end, we will prove Lemma 1.7.

Theorem 2.1. *Assume $a(x, y)$ is a smooth function supported in a strip of x -width no more than δ_1 and y -width no more than δ_2 , satisfying the following derivative conditions:*

$$(2.1) \quad |\partial_y a(x, y)| \lesssim \delta_2^{-1} \quad \text{and} \quad |\partial_y^2 a(x, y)| \lesssim \delta_2^{-2}.$$

Let $\mu > 0$ and $S(x, y)$ be a smooth function such that, for all $(x, y) \in \text{Conv}(\text{supp}(a))$,

$$(2.2) \quad |D_{\pi_0^\perp} S(x, y)| \gtrsim \mu \quad \text{and} \quad |\partial_y^\beta D_{\pi_0^\perp} S(x, y)| \lesssim \frac{\mu}{\delta_2^\beta} \quad \text{for } \beta = 1, 2.$$

Then for Λ_S defined as in (1.6), one has

$$|\Lambda_S(f_1, f_2, f_3)| \lesssim |\lambda\mu|^{-1/6} \prod_{j=1}^3 \|f_j\|_2.$$

The above theorem can be viewed as a trilinear analogue of Phong–Stein’s operator version of van der Corput lemma [15]:

Theorem 2.2. *Assume $a(x, y)$ is a smooth function supported in a strip of x -width no more than δ_1 and y -width no more than δ_2 , satisfying the same derivative conditions as in (2.1). Suppose $\mu > 0$ and $S(x, y)$ is a smooth function in \mathbb{R}^2 such that the following holds for all $(x, y) \in \text{supp}(a)$:*

$$(2.3) \quad |\partial_x \partial_y S(x, y)| \gtrsim \mu \quad \text{and} \quad |\partial_x \partial_y^{\beta+1} S(x, y)| \lesssim \frac{\mu}{\delta_2^\beta} \quad \text{for } \beta = 1, 2.$$

Then, for T defined as in (1.1),

$$\|T(f)\|_2 \lesssim (\lambda\mu)^{-1/2} \|f\|_2.$$

In both theorems above, we have adopted the notation $X \lesssim Y$ to denote $|X| \leq CY$, where C can depend on a and S , but is independent of δ_1 , δ_2 , μ and λ . It is also worth pointing out that theorem 2.2 is not exactly the same as the one employed by Phong–Stein in [15], we have adopted a slightly more general version from Greenblatt [6]. For the proof of Theorem 2.2, we also refer the reader to [6].

Now we turn to the technical details in proving Theorem 1.4 and Theorem 2.1. First of all, (1.10) can be reduced to (1.6). Set

$$(2.4) \quad \|\Lambda_{S,\pi}\| = \sup\{|\Lambda_{S,\pi}(f_1, f_2, f_3)|, \|f_j\|_2 \leq 1 \quad \text{for } j = 1, 2, 3\}$$

and let $\|\Lambda_S\|$ be defined similarly. We may assume $\pi_1(x, y) = x$, $\pi_2(x, y) = y$ and $\pi_3(x, y) = Ax + By$, where $A \neq 0$ and $B \neq 0$. Change variables $u = Ax$

and $v = By$ to obtain

$$\begin{aligned}\Lambda_{S,\pi}(f_1, f_2, f_3) &= \iint e^{i\lambda S(x,y)} f_1(x) f_2(y) f_3(Ax + By) dx dy \\ &= \frac{1}{AB} \iint e^{i\lambda S(u/A, v/B)} f_1(u/A) f_2(v/B) f_3(u+v) du dv \\ &= \frac{1}{AB} \iint e^{i\lambda S_{A,B}(u,v)} f_{1,A}(u) f_{2,B}(v) f_3(u+v) du dv,\end{aligned}$$

where $S_{A,B}(u, v) = S(u/A, v/B)$, $f_{1,A}(u) = f_1(u/A)$ and $f_{2,B}(v) = f_2(v/B)$. Notice that

$$D_{\pi_0^\perp} S_{A,B}(u, v) = \frac{1}{AB} ((\partial_u/A - \partial_v/B) \partial_u \partial_v S)(u/A, v/B).$$

Thus $D_{\pi_0^\perp} S_{A,B} = CD_{\pi^\perp} S$ for an appropriate constant C . In addition $\|f_{1,A}\|_2 = \sqrt{A}\|f_1\|_2$ and $\|f_{2,B}\|_2 = \sqrt{B}\|f_2\|_2$. Finally, notice that the convexity is preserved under linear mappings. Therefore, for an appropriate constant C , one has

$$\|\Lambda_{S,\pi}\| = C\|\Lambda_{S_{A,B}}\|.$$

Now we turn to the proofs of Theorem 1.4 and Theorem 2.1, and only need to consider Λ_S . For simplicity, we assume $\|f_1\|_2 = \|f_2\|_2 = \|f_3\|_2 = 1$. Applying change of variables $(u, v) = (x + y, y)$ and duality, one has

$$\|\Lambda_S(f_1, f_2, f_3)\| \leq \|B(f_1, f_2)\|_2 \|f_3\|_2 = \|B(f_1, f_2)\|_2,$$

where

$$B(f_1, f_2)(u) = \int e^{i\lambda S(u-v, v)} f_1(u-v) f_2(v) a(u-v, v) dv.$$

Employing the method of TT^* , one obtains

$$\begin{aligned}\|B(f_1, f_2)\|_2^2 &= \iiint e^{i(\lambda S(u-v_1, v_1) - \lambda S(u-v_2, v_2))} f_1(u-v_1) \bar{f}_1(u-v_2) f_2(v_1) \bar{f}_2(v_2) \\ &\quad \times a(u-v_1, v_1) a(u-v_2, v_2) dv_1 dv_2 du.\end{aligned}$$

Change variables: $x = u - v_1$, $y = v_1$ and $\tau = v_2 - v_1$ and set

$$\begin{aligned}(2.5) \quad S_\tau(x, y) &= S(x, y) - S(x - \tau, y + \tau), \\ F_\tau(x) &= f_1(x) \bar{f}_1(x - \tau), \\ G_\tau(y) &= f_2(y) \bar{f}_2(y + \tau), \\ a_\tau(x, y) &= a(x, y) a(x - \tau, y + \tau).\end{aligned}$$

This yields

$$(2.6) \quad \|B(f_1, f_2)\|_2^2 = \iiint e^{i\lambda S_\tau(x,y)} F_\tau(x) G_\tau(y) a_\tau(x, y) dx dy d\tau.$$

The proofs of Theorem 1.4 and Theorem 2.1 slightly diverge now and are presented in two separated subsections.

2.1. Proof of Theorem 1.4

Split $\|B(f_1, f_2)\|_2^2$ into $B_1 + B_2$ according to the value of $|\tau|$ below:

- Case 1. $|\tau| \leq |\lambda|^{-1/3}$.
- Case 2. $|\tau| \geq |\lambda|^{-1/3}$.

In Case 1, we simply move the absolute value into the integrals, which yields

$$(2.7) \quad B_1 \leq \int_{|\tau| \leq |\lambda|^{-1/3}} \|F_\tau\|_1 \|G_\tau\|_1 d\tau \leq |\lambda|^{-1/3} \|f_1\|_2^2 \|f_2\|_2^2 = |\lambda|^{-1/3}.$$

In Case 2, in order to employ Theorem 1.1 to the inner double-integral, we assume for a moment in the support of a_τ , the following holds for some positive constant C :

$$(2.8) \quad |\partial_x \partial_y S_\tau(x, y)| \geq C |\tau|.$$

By Theorem 1.1 and the Cauchy–Schwarz inequality, B_2 is dominated by

$$(2.9) \quad \begin{aligned} & \int_{|\tau| \geq |\lambda|^{-1/3}} C |\lambda \tau|^{-1/2} \|F_\tau\|_2 \|G_\tau\|_2 d\tau \\ & \leq C |\lambda|^{-1/3} \int \|F_\tau\|_2 \|G_\tau\|_2 d\tau \leq C |\lambda|^{-1/3} \left(\int \|F_\tau\|_2^2 d\tau \cdot \int \|G_\tau\|_2^2 d\tau \right)^{1/2} \\ & = C |\lambda|^{-1/3} \|f\|_2^2 \|g\|_2^2 \leq C |\lambda|^{-1/3}. \end{aligned}$$

Thus

$$\|B(f, g)\|_2^2 = B_1 + B_2 \leq C |\lambda|^{-1/3}.$$

It remains to verify (2.8) on the support of (2.5). Set

$$F(t) = S_{xy}(x - t, y + t).$$

Then,

$$(2.10) \quad |F'(t)| = |(\partial_x - \partial_y) \partial_x \partial_y S(x - t, y + t)|.$$

By the mean value theorem, there is a t_0 between 0 and τ such that

$$(2.11) \quad |\partial_x \partial_y S_\tau(x, y)| = |F(0) - F(\tau)| = \left| \int_0^\tau F'(t) dt \right| = |\tau| |F'(t_0)|.$$

Notice that $(x, y) \in \text{supp}(a)$ and $(x - \tau, y + \tau) \in \text{supp}(a)$, then by convexity $(x - t_0, y + t_0) \in \text{Conv}(\text{supp}(a))$. Therefore, (1.11), (2.11) and (2.10) yield (2.8).

2.2. Proof of Theorem 2.1

Similarly, we split $\|B(f_1, f_2)\|_2^2$ into $B_1 + B_2$ according to the value of $|\tau|$ as below:

- Case 1. $|\tau| \leq |\lambda \mu|^{-1/3}$.
- Case 2. $|\tau| \geq |\lambda \mu|^{-1/3}$.

In Case 1, we simply move the absolute value into the integrals and thus

$$B_1 \leq \int_{|\tau| \leq |\lambda\mu|^{-1/3}} \|F_\tau\|_1 \|G_\tau\|_1 d\tau \leq |\lambda\mu|^{-1/3} \|f_1\|_2^2 \|f_2\|_2^2 = |\lambda\mu|^{-1/3}.$$

In Case 2, assume at a moment that (2.1) is true for a_τ and (2.3) are true for S_τ with μ replaced by $|\lambda\mu|$. Then Theorem 2.2 implies

$$(2.12) \quad B_2 \leq C \int_{|\tau| \geq |\lambda\mu|^{-1/3}} |\lambda\mu\tau|^{-1/2} \|F_\tau\|_2 \|G_\tau\|_2 d\tau \leq C |\lambda\mu|^{-1/3}.$$

It remains to verify the conditions mentioned above. Indeed (2.1) follows by $a_\tau(x, y) = a(x, y)a(x - \tau, y + \tau)$. S_τ satisfies the first part of (2.3) with μ replaced by $|\lambda\mu|$ due to (2.2), (2.10), (2.11) and the convexity assumption in Theorem 2.1. If we set

$$F_1(t) = \partial_x \partial_y^2 S(x - t, y + t) \quad \text{and} \quad F_2(t) = \partial_x \partial_y^3 S(x - t, y + t),$$

then the second part of (2.3) (with μ replaced by $|\lambda\mu|$) follows from (2.2), (2.10), (2.11) (with F replaced by F_1 and F_2) and convexity.

2.3. Proof of Lemma 1.7

Proof. By the Cauchy–Schwarz inequality, Hölder’s inequality and the Fubini theorem, one has

$$\begin{aligned} & \left| \iint f_1(x) f_2(y) f_3(x + y) a(x, y) dx dy \right| \\ &= \left| \int f_1(x) \left(\int f_2(y) f_3(x + y) a(x, y) dy \right) dx \right| \\ &\leq \left| \int \left(\int f_2(y) f_3(x + y) a(x, y) dy \right)^2 dx \right|^{1/2} \cdot \|f_1\|_2 \\ &\leq \left| \int \left(\int |a(x, y)|^2 dy \right) \left(\int |f_2(y) f_3(x + y)|^2 dy \right) dx \right|^{1/2} \cdot \|f_1\|_2 \\ &\leq C \delta_2^{1/2} \|f\|_1 \|f_2\|_2 \|f_3\|_2. \end{aligned}$$

The other bound can be obtained similarly. \square

The estimates in both Lemma 1.7 and Lemma 1.6 are sharp, which can be seen by taking $f_1 = \mathbf{1}_{[0, \delta_1]}$, $f_2 = \mathbf{1}_{[0, \delta_2]}$ and $f_3 = \mathbf{1}_{[0, \delta_1 + \delta_2]}$.

3. Ideas of the algorithm

To employ Theorem 2.1 to attack Theorem 1.5, one needs to decompose $\text{supp}(a)$ into regions such that $P(x, y)$ is well-behaved, where $P = \partial_x \partial_y (\partial_x - \partial_y) S$. Ideally, one hopes $P(x, y)$ to behave like a monomial with a negligible error term. The

algorithm is driven by this idea. In each stage of iteration, ‘good’ regions (with the desired property) are obtained via vertices and edges of $\mathcal{N}(P)$ when $P(x, y)$ is ‘nonvanishing’, and ‘bad’ regions are obtained when $P(x, y)$ ‘vanishes’ on these edges. In each such ‘good’ region, $P(x, y)$ behaves like a monomial and no further treatment is required. Each of those ‘bad’ regions is then carried to the next stage of iteration. A branch of iterations is created for each ‘bad’ region and one hopes all the ‘bad’ regions will eventually go away as the iterations go deeper. To see how it works, drop all zero coefficients and write the Taylor expansion of $P(x, y)$ as

$$P(x, y) = \sum_{p, q \in \mathbb{N}} c_{p, q} x^p y^q$$

The Newton diagram is the boundary of $\mathcal{N}(P)$, which consists of two non-compact edges, a finite collection of compact edges $\mathcal{E}(P)$ (may be empty) and a finite collection of vertices $\mathcal{V}(P)$. The Euler formula gives $\#\mathcal{V}(P) - \#\mathcal{E}(P) = 1$. For each face $F \in \mathcal{F}(P)$, define

$$(3.1) \quad P_F(x, y) = \sum_{(p, q) \in F} c_{p, q} x^p y^q.$$

Choose one $L_m \in \mathcal{SL}(P)$ and consider the region $|y| \sim |x|^m$ associated to it. Notice that L_m goes through at least one $V \in \mathcal{V}(P)$, say $V = (p_v, q_v)$. Let E_l and E_r be the edges left to and right to V with $\mathcal{M}(E_l) = m_l$ and $\mathcal{M}(E_r) = m_r$, respectively. Then $0 \leq m_l \leq m \leq m_r \leq \infty$. Consider the following three possible cases:

Case (1) $m_l < m < m_r$, **Case (2)** $m = m_l$ and **Case (3)** $m = m_r$, which correspond to: (1) the vertex V ‘dominates’ $P(x, y)$, (2) the edge E_l ‘dominates’ $P(x, y)$ and (3) the edge E_r ‘dominates’ $P(x, y)$ respectively.

In **Case (1)**, $p_v + mq_v < p + mq$ for any other (p, q) with $c_{p, q} \neq 0$. Then in the region $|y| \sim |x|^m$,

$$P(x, y) = P_V(x, y) + O(|x|^{p_v + mq_v + \nu}), \quad \text{for some } \nu > 0.$$

Given $|x|$ sufficiently small, $P_V(x, y)$ is the dominant term in $P(x, y)$, i.e.,

$$P(x, y) \sim P_V(x, y) = c_{p_v, q_v} x^{p_v} y^{q_v}.$$

Case (2) and **Case (3)** are exactly the same, so we only discuss **Case (2)** $m = m_l$ here. In addition, we focus on the right half plane $x > 0$ and assume $y = rx^m$, where $r \in \mathbb{R}$ is a parameter. Notice that $p_v + mq_v = p + mq$ for all $(p, q) \in E_l$ and $p_v + mq_v < p + mq$ for all $(p, q) \notin E_l$. Thus

$$\begin{cases} P_{E_l}(x, rx^m) = P_{E_l}(1, r)x^{p_v + mq_v} \\ P(x, y) = P_{E_l}(1, r)x^{p_v + mq_v} + O(x^{p_v + mq_v + \nu}), \quad \text{for some } \nu > 0. \end{cases}$$

As long as r is away from the non-zero real roots of $P_{E_l}(1, r)$, the origin and the infinity, $P_{E_l}(x, y)$ is the dominant term of $P(x, y)$. We refer this as the case that the edge E_l ‘dominates’ $P(x, y)$.

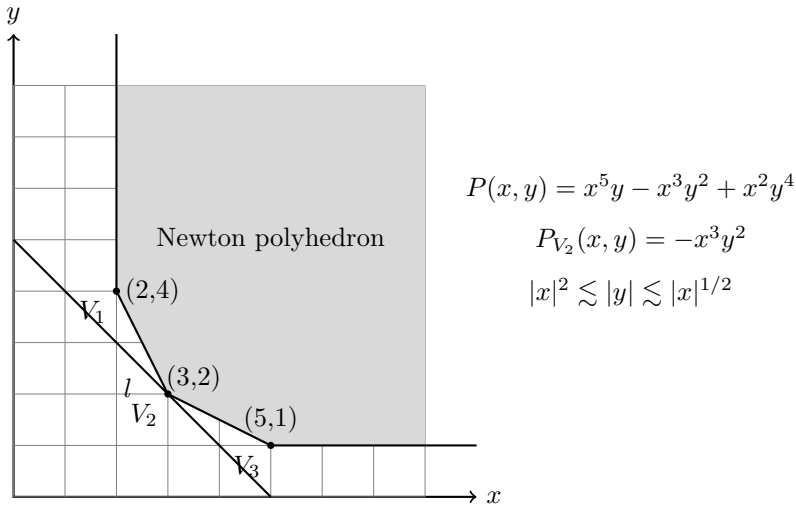


Figure 1. Case (1). The vertex V_2 is dominant, where $1/2 < m < 2$.

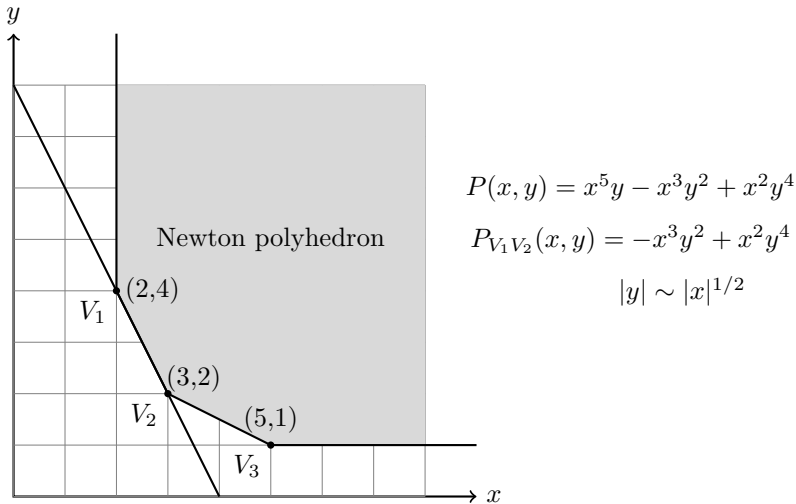


Figure 2. Case (2). The edge V_1V_2 is dominant, where $m = 1/2$.

What remains are the ‘bad’ regions: when r is in some neighborhood of the non-zero roots of $P_{E_i}(1, r)$, which demands most of the work. The traditional way of further analyzing these ‘bad’ regions would be to do a coordinate change. Choose a non-zero root r_0 of $P_{E_i}(1, r)$, set $(x_1, y_1) = (x, y - r_0x^m)$ and consider the function $P(x, y)$ in the new coordinates, namely $P_1(x_1, y_1) = P(x_1, y_1 + r_0x_1^m)$. We can then apply the above arguments to the function $P_1(x_1, y_1)$, which will result in a further partition of a ‘bad’ region into ‘good’ regions where vertices or edges dominate, and ‘bad’ regions. Iterating such procedures would end up with an infinite collection of ‘good’ regions. As a by-product, one can find solutions for $P(x, \gamma(x)) = 0$, where

each $\gamma(x)$ is a Puiseux series. However, for analytic purposes it would be of significant advantage and even crucial for the algorithm to converge. By saying an algorithm converges, we always mean that it terminates after finitely many steps. Indeed, if the algorithm fails to converge,

1) we may fail to find a neighborhood for the desired partition. Notice that in each stage of iteration, in order for the functions $P(x, y), P_1(x_1, y_1), \dots$ to behave like monomials, we need the support of $|x|, |x_1|, \dots$ to be sufficiently small. In the case the algorithm fails to converge, it is unclear whether there is a uniform upper bound for $|x| = |x_1| = \dots$, that works for all stage of iterations. Without such an upper bound, one may not partition the support of the cut-off function $a(x, y)$ into ‘good’ regions. However, this issue does not arise if the algorithm converges. Another possible way to fix this issue is to investigate on how the upper bound of $|x|$ relies on the function $P(x, y)$. This direction seems quite challenging but is of great interest, for it is related to the stability of oscillatory integrals/operators.

2) Even assuming we can find a neighborhood to do the partition, it is significantly simpler to work on a finite collection of ‘good’ regions rather than an infinite collection. For the former, we essentially work in single ‘good’ region, since the desired estimates are allowed to rely on the cardinality of ‘good’ regions, unless one seeks for a uniform bound for a class of phases. For the latter, one has to keep track of all constants appearing in all stages of iterations, and hope to be able to absolutely sum all the resulting estimates together. Heuristically, a convergent algorithm allows us to divide a problem concerning an arbitrary real analytic function into finitely many subproblems, each of which is concerning a ‘monomial’.

Unfortunately, only performing the change of coordinates $(x_1, y_1) = (x, y - r_0 x^m)$ is not sufficient for the algorithm to converge. For example, if we apply the above algorithm to the function

$$P(x, y) = \left(y - \left(\sum_{j=0}^{\infty} r_j x^{j+1} \right) \right)^n,$$

we will end up with repeating the following change of variables:

$$y_{j+1} = y_j - r_j x, \quad \text{for } j \geq 0,$$

and the iterations will never stop.

How can one modify the above ideas to ensure the convergence of the algorithm? Greenblatt [6] had some very nice observations in achieving this goal. The key is to complexify the change of coordinates in each stage of iterations by means of the implicit function theorem (IFT). Roughly speaking, each such change of coordinates helps to decrease certain ‘order’ of $P(x, y)$ by at least 1,

To see how this works, choose a root r_0 of $P_{E_l}(1, r)$ and assume its order is s_0 . Applying the IFT to $\partial_y^{s_0-1} P(x, y) = 0$ yields that there is a unique $h(x)$ such that

1. $\partial_y^{s_0-1} P(x, h(x)) = 0$,
2. $h(x)$ is a real analytic function of $x^{1/M}$ whose leading term is given by $r_0 x^m$, where M is some positive integer.

Instead of doing coordinate change of the form $y_1 = y - r_0 x^m$, one must do coordinate change of the form $y_1 = y - h(x)$. Consequently, each root r_1 , an analogue of r_0 from the next stage of iteration, has order s_1 at most $s_0 - 1$. Then the algorithm must converge. As a trade-off, the function $h(x)$ is given implicitly as a Puiseux series, which is not easy to compute in general. For instance, consider the function

$$P(x, y) = (y - x)^n + x^n y^{2n}.$$

One can see that $\mathcal{N}(P)$ has only one compact edge E and $P_E(1, r)$ has only one root r_0 of order $s_0 = n$. To apply the algorithm in [6], one first uses the IFT to solve $\partial_y^{n-1} P(x, y) = 0$, i.e., $y - x + c x^n y^{n+1} = 0$ where $c = \frac{(2n)!}{n!(n+1)!}$. The solution is then given by $y = h(x) = x + O(x^{1+\mu})$ for some $\mu > 0$. Then change the coordinate $y = y_1 + h(x_1)$ and plug it into $P(x, y)$. In general, to do iterations in later stages, one needs to compute the Puiseux expansion of $h(x)$ up to a certain number of terms. One then needs to run the Newton–Puiseux algorithm to do such computation.

Can one retain the simplicity and explicitness (avoid using the IFT) of the change of variables/coordinates, and also ensure the convergence of the algorithm? The answer is affirmative. To do so, we perform the change of variables in each stage of iterations as follows:

$$(3.2) \quad (x, y) = (x_1, r_0 x_1^m + y_1 x_1^m).$$

The x_1^m term in front of y_1 plays the role in rescaling ‘curved’ regions back into ‘uncurved’ regions, allowing us to iterate in the same kind of regions. Now we assume the algorithm does not stop, resulting in an infinite chain

$$(3.3) \quad [U, P] = [U_0, P_0] \rightarrow [U_1, P_1] \rightarrow [U_2, P_2] \rightarrow \cdots \rightarrow [U_n, P_n] \rightarrow \cdots.$$

Each $U_n = (0, \epsilon) \times (-\epsilon, \epsilon)$ is the domain for $P_n(x_n, y_n)$, obtained inductively via change of variables of the form (3.2), for instance $P_1(x_1, y_1) = P(x_1, r_0 x_1^m + x_1^m y_1)$. We then search for certain patterns inside $\mathcal{N}(P_n)$ as $n \rightarrow \infty$. Amazingly, the number of compact edges of $\mathcal{N}(P_n)$ converges to 1. More precisely, there exists an $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, $\mathcal{N}(P_n)$ has only one compact edge E_n and the restriction of P_n to it is $c_n (y_n - r_n x_n^{m_n})^{s_{n_0}}$. With this observation at hand, we can now turn off the iterations. At the n_0 -th stage of iteration, instead of repeating the change of variables of the form (3.2), we do all of them together at a time

$$y_{n_0} = y_{n_0+1} x_{n_0}^{m_{n_0}} + \sum_{k=n_0}^{\infty} r_k x_{n_0}^{m_{n_0} + m_{n_0+1} + \cdots + m_k},$$

then the algorithm stops immediately; see Lemma 4.9.

The benefits of the explicitness and simplicity of the change of variables can help keep very good track of variables in different stages. For any given n , one can write down the relation between (x_n, y_n) and (x, y) explicitly; see (4.42). One is also able to compute the Newton polyhedron of P_n , assuming we know how to find roots of one-variable polynomials. It is also quite convenient to estimate lower bounds of $|P(x, y)|$ in ‘good’ regions from later stages of iterations inductively via these Newton polyhedra; see (4.47) and (4.48).

3.1. Two model examples

The function in the first example is quite simple, we aim to implement the algorithm in details. In particular, we provide details to compute the size of the original domain, i.e., $\epsilon = \epsilon_0$ below. However, we would recommend the reader to focus on the change of variables rather than the values of ϵ_k and ρ_k below. The second example is slightly more complicated, with primary focus on the convergence of the algorithm. In these two examples, we only handle the right half plane $x > 0$, for the other half can be handled similarly.

Example 3.1. The first example is

$$P(x, y) = (y - x)(y - x - x^2) = (y^2 - 2xy + x^2) - x^2y + x^3,$$

and we shall decompose the domain $U_0 = U = (0, \epsilon) \times (-\epsilon, \epsilon)$ into a finite collection of ‘good’ regions, where $\epsilon = \epsilon_0$ is sufficiently small that will be determined at the end of the algorithm. The pair $[U_0, P_0] = [U, P]$ is the input and the algorithm runs as follows.

- (0) Notice $\mathcal{N}(P)$ has two vertices $V_1 = (2, 0)$, $V_2 = (0, 2)$ and one compact edge $E = V_1V_2$, whose slope is -1 . Let $\rho_0 > 0$ be sufficiently small, but much greater than ϵ_0 , say $\rho_0 > 2^{10}\epsilon_0$. The polynomial $P_E(1, y)$ has only one root $r_0 = 1$. Then in the ‘good’ region

$$U_{0,g,V_1} = \{(x, y) \in U : (1 + \rho_0)x \leq y < \epsilon_0\}$$

one has

$$P(x, y) \sim_{\rho_0} y^2 = P_{V_1}(x, y),$$

and in the ‘good’ region

$$U_{0,g,V_2} = \{(x, y) \in U : -\epsilon_0 < y \leq (1 - \rho_0)x\}$$

one has

$$P(x, y) \sim_{\rho_0} x^2 = P_{V_2}(x, y).$$

The remaining ‘bad’ region is $y = (1 + y_1)x$, where $|y_1| < \rho_0$. Set $x_1 = x$, $y = x_1 + x_1y_1$ and

$$P_1(x_1, y_1) = P(x_1, x_1(1 + y_1)) = x_1^2y_1^2 - x_1^3y_1.$$

Choose ϵ_1 sufficiently small but much bigger than ρ_0 , say $\epsilon_1 > 2^{10}\rho_0$, and set $U_1 = (0, \epsilon_1) \times (-\epsilon_1, \epsilon_1)$.

- (1) Now we are in the 1-st stage of iteration, whose input is $[U_1, P_1]$. Notice $\mathcal{N}(P_1)$ has two vertices $V'_1 = (2, 2)$, $V'_2 = (3, 1)$ and only one compact edge $E' = V'_1V'_2$. Let $\rho_1 > 2^{10}\epsilon_1$. Similarly, in the region $x_1 < \epsilon_1$, $(1 + \rho_1)x_1 \leq y_1 < \epsilon_1$,

$$P_1(x_1, y_1) \sim_{\rho_1} y_1^2x_1^2 = P_{1,V'_1}(x_1, y_1)$$

and in the region $x_1 < \epsilon_1$, $-\epsilon_1 < y_1 \leq (1 - \epsilon_1)x_1$,

$$P_1(x_1, y_1) \sim_{\rho_1} -y_1 x_1^3 = P_{1, V'_2}(x_1, y_1).$$

The ‘bad’ region is $y_1 = (1 + y_2)x_2$ with $|y_2| < \rho_1$. Change variables $x_2 = x_1$, $y_1 = (1 + y_2)x_2$ and set

$$P_2(x_2, y_2) = P_1(x_2, (1 + y_2)x_2) = x_2^3(1 + y_2)y_2.$$

Choose $\epsilon_2 > 2^{10}\rho_1$ and set $U_2 = (0, \epsilon_2) \times (-\epsilon_2, \epsilon_2)$.

- (2) In the 2-nd stage, the input is $[U_2, P_2]$. Now $\mathcal{N}(P_2)$ has only one vertex $V'' = (3, 1)$ and $P_2(x_2, y_2) \sim x_2^3 y_2$, given ϵ_2 sufficiently small. The algorithm stops here.

One can now determine the value of ϵ_0 , whose choice relies on how small one wants the error term to be: by choosing $\epsilon_0 > 0$ sufficiently small, we can let the ratio between the ‘dominant term’ and the ‘error term’ be sufficiently large. In general, we need to compute $\{\epsilon_k\}$ and $\{\rho_k\}$ backward. In this case, one can safely take $2^{-10} > \epsilon_2 > 2^{10}\rho_1 > 2^{20}\epsilon_1 > 2^{30}\rho_0 > 2^{40}\epsilon_0$ and in particular any value of ϵ_0 less than 2^{-50} .

The corresponding partition of $U = (0, \epsilon) \times (-\epsilon, \epsilon)$ will be

$$\begin{cases} U_{0,g,V_1} = \{(x, y) \in U : y \geq (1 + \rho_0)x\}, & \text{where } P(x, y) \sim y^2, \\ U_{0,g,V_2} = \{(x, y) \in U : y \leq (1 - \rho_0)x\}, & \text{where } P(x, y) \sim x^2, \\ U_{1,g,V'_1} = \{(x, y) \in U : y_1 \geq (1 + \rho_1)x_1\}, & \text{where } P(x, y) \sim x_1^2 y_1^2, \\ U_{1,g,V'_2} = \{(x, y) \in U : y_1 \leq (1 - \rho_1)x_1\}, & \text{where } P(x, y) \sim -x_1^3 y_1, \\ U_{2,g} = \{(x, y) \in U : -\rho_1 < y_2 < \rho_1\}, & \text{where } P(x, y) \sim x_2^3 y_2, \end{cases}$$

where

$$\begin{cases} x = x_1 = x_2, \\ y = x + x y_1 = x + x^2 + x^2 y_2. \end{cases}$$

To help the reader better understand how the two algorithms work differently, we also implement the one in [6] for this example.

First, notice $s_0 = 2$, we then need to solve $\partial_y P(x, y) = 0$. This gives $y = x + x^2/2$. Set $y = x + x^2/2 + y_1$, then

$$\tilde{P}_1(x, y_1) = P(x, x + x^2/2 + y_1) = (y_1 - x^2/2)(y_1 + x^2/2).$$

The only compact edge of $\mathcal{N}(\tilde{P}_1)$ has two roots, $r_1 = 1/2$ and $r'_1 = -1/2$. Set $y_1 = x^2/2 + y_2$ and assume $|y_2| < x^2/2$, then

$$\tilde{P}_2(x, y_2) = \tilde{P}_1(x, x^2/2 + y_2) = y_2(y_2 + x^2) \sim x^2 y_2.$$

Similarly set $y_1 = -x^2/2 + y'_2$ and assume $|y'_2| < x^2/2$, one has

$$\tilde{P}'_2(x, y'_2) = \tilde{P}_1(x, -x^2/2 + y'_2) = y'_2(y'_2 - x^2) \sim -x^2 y'_2.$$

The algorithm stops here.

Example 3.2. Since the major focus is the termination of the algorithm, we shall be brief and not worry about technical details. Consider

$$P(x, y) = xy(y^2 - x)(y - x^2) \left(y - \sum_{n=1}^{\infty} x^n \right)^2.$$

In the 0-th stage of iteration, the ‘bad’ regions are those near the curves $y^2 - x = 0$, $y - x^2 = 0$ and $y - x = 0$. Thus, away from the neighborhoods of the following curves: $y = -\sqrt{x}$, $y = \sqrt{x}$, $y = x^2$ and $y = x$, the function $P(x, y)$ behaves like a monomial. Each of the ‘bad’ regions near $y = -\sqrt{x}$, $y = \sqrt{x}$ and $y = x^2$ will go away via one step of change of variables – all the resulting new Newton polyhedra have only one vertex; see Example 4.7 in Section 4.

To address the ‘bad’ region near $y = x$, we do change of variables $x = x_1$ and $y = x_1 + x_1 y_1$, which yields

$$P_1(x_1, y_1) = P(x_1, x_1 + x_1 y_1) = Q_1(x_1, y_1) x_1^6 \left(y_1 - \sum_{n=1}^{\infty} x_1^n \right)^2,$$

where

$$Q_1(x_1, y_1) = (1 + y_1)(1 + y_1 - x_1)(-1 + x_1(1 + y_1)^2).$$

One can ignore the function $Q_1(x_1, y_1)$ since it is non-vanishing near the origin. Doing change of variables of the form $y_{k-1} = x_k + x_k y_k$ does not help to resolve the singularities, and the Newton polyhedron of P_{k-1} only shifts to the right by 2 units. But we want the Newton polyhedron to end with only one vertex, which corresponds to the case where a single vertex is dominant. This indeed can be done as follows. Set

$$\begin{cases} x_1 = x_2, \\ y_1 = x_2 y_2 + \sum_{n=1}^{\infty} x_2^n; \end{cases}$$

then

$$P_2(x_2, y_2) = P_1 \left(x_2, x_2 y_2 + \sum_{n=1}^{\infty} x_2^n \right) = Q_2(x_2, y_2) x_2^8 y_2^2.$$

The function $Q_2(x_2, y_2)$ is non-vanishing and $P_2(x_2, y_2)$ behaves like the monomial $x_2^8 y_2^2$. The algorithm stops here.

4. An algorithm for resolution of singularities in \mathbb{R}^2

Let U be a sufficiently small neighborhood of 0 in \mathbb{R}^2 . For simplicity, we restrict our discussion to the right half-plane $x > 0$, since the left half plane can be reduced to this case through the change of variables $(x, y) \rightarrow (-x, y)$ and the y -axis can be ignored first. For the rest of this section, we take $U = \{(x, y) : 0 < x < \epsilon, -\epsilon < y < \epsilon\}$ and choose ϵ to be sufficiently small. The exact value of ϵ depends on later stages of iterations and will be clear at the end of the algorithm. An inductive

resolution procedure will be performed on the pair $[U, P]$, where P is any analytic function defined in an open neighborhood containing U . Let $M \in \mathbb{N}$ be a large integer.

Definition 4.1. A region W is standard if

$$W = \{(X, Y) : 0 < X < \epsilon, -\epsilon < Y < \epsilon\}$$

for some $\epsilon > 0$, which is referred to as the size of W . A standard pair $[W, Q]$ is a standard region W together with a power series $Q(X, Y)$ of $(X^{1/M}, Y)$, which converges absolutely in some neighborhood of W .

The iterations will always be performed on standard pairs, with the regions and functions of $(x^{1/M}, y)$ varying.

4.1. Part I. Single step of partition

Set $[U_0, P_0] = [U, P]$, $\epsilon_0 = \epsilon$ and $(x_0, y_0) = (x, y)$ to indicate the procedure is in the 0-th stage. Consider

$$P_E(x, y) = \sum_{(p,q) \in E} c_{p,q} x^p y^q, \quad \text{for } E \in \mathcal{E}(P).$$

Let $V_{E,l} = (p_{E,l}, q_{E,l})$ and $V_{E,r} = (p_{E,r}, q_{E,r}) \in \mathcal{V}(P)$ be the left and right vertices of E . Set $m_E = \frac{p_{E,l} - p_{E,r}}{q_{E,r} - q_{E,l}}$, then the slope of E is $-1/m_E$. The constant m_E is the most important constant associated to each edge E . Set $e_E = p_{E,l} + m_E q_{E,l}$, then $e_E = p' + m_E q'$ for all $(p', q') \in E$ and there is a $\nu > 0$ such that $p'' + m_E q'' \geq e_E + \nu$ for all $(p'', q'') \notin E$. On the curve $y = rx^{m_E}$, where $r \in \mathbb{R} \setminus \{0\}$,

$$P_E(x, y) = x^{e_E} \sum_{(p,q) \in E} c_{p,q} r^q =: x^{e_E} P_E(r)$$

and

$$P(x, y) - P_E(x, y) = O(x^{e_E + \nu})$$

has a higher degree. Thus, given $|x|$ sufficiently small, $P_E(x, y)$ dominates $P(x, y)$, unless

$$P_E(r) = \sum_{(p,q) \in E} c_{p,q} r^q \rightarrow 0.$$

Assume $\{r_{E,j}\}_{1 \leq j \leq J_E}$, labeled in an increasing order, is the set of non-zero roots of $P_E(r) = 0$ of orders $\{s_{E,j}\}_{1 \leq j \leq J_E}$, respectively. Then

$$J_E \leq \sum_{1 \leq j \leq J_E} s_j \leq q_{E,l} - q_{E,r},$$

since $P_E(r) = r^{q_{E,r}} \sum_{(p,q) \in E} c_{p,q} r^{q - q_{E,r}}$. For simplicity, we say $r_{E,j}$ is a root of E instead of saying $r_{E,j}$ is a root of $P_E(r)$. Assign two constants c_E and C_E to each edge E . They are independent from ϵ_0 and satisfy

$$(4.1) \quad 0 < c_E < 2^{-10} |r_{E,j}| < 2^{10} |r_{E,j}| < C_E, \quad \text{for all } 1 \leq j \leq J_E.$$

The numbers 2^{-10} and 2^{10} above (same as below) are just a random choice for small and large numbers. Let $I_j^{\rho_0}(E) = (r_{E,j} - \rho_0, r_{E,j} + \rho_0)$, where $\rho_0 > 0$ is much greater than ϵ_0 , but much smaller than any c_E (and hence any $r_{E,j}$), and

$$(4.2) \quad 0 < \rho_0 < 2^{-10} \cdot \min_E \{c_E\}$$

is to be determined at the end of iterations.

Set

$$(4.3) \quad \begin{cases} I(E) = [c_E, C_E] \cup [-C_E, -c_E], \\ I_b(E) = \cup_{1 \leq j \leq J_E} I_j^{\rho_0}(E), \\ I_g(E) = I(E) \setminus I_b(E). \end{cases}$$

Here $I_b(E)$ are the neighborhood of the roots $\{r_{E,j}\}$ and $I_g(E)$ are the points away from the non-zero roots, 0 and ∞ . Then

$$(4.4) \quad |P_E(r)| \gtrsim_{\rho_0} 1 \quad \text{for } r \in I_g(E).$$

Thus $P_E(x, y)$ dominates $P(x, y)$ if $y = rx^{m_E}$ and $r \in I_g(E)$. Let

$$U_{0,g}(E) = \{(x, y) \in U_0 : y = rx^{m_E}, r \in I_g(E)\}$$

be the ‘good’ regions generated by the edge E , which is a disjoint union of $(J_E + 2)$ ‘good’ regions: $U_{0,g}(E, j)$. Each ‘good’ region $U_{0,g}(E, j)$ is a curved triangular region of the form

$$(4.5) \quad U_{0,g}(E, j) = \{(x, y) \in U_0 : b_j x^{m_E} \leq y \leq B_j x^{m_E}\},$$

where $[b_j, B_j] := I_g(E, j)$ is just a connected component of $I_g(E)$ and $(J_E + 2)$ comes from the number of connected components of $I_g(E)$.

In the above notation, the subindex 0 in $U_{0,g}(E, j)$ indicates the algorithm is in the 0-th stage, g indicates the region is ‘good’. The ‘bad’ regions are defined as

$$(4.6) \quad \begin{aligned} U_{0,b}(E, j) &= \{(x, y) \in U_0 : y = rx^{m_E}, r \in I_j^{\epsilon}(E)\} \\ &= \{(x, y) \in U_0 : (r_j - \rho_0)x^{m_E} < y < (r_j + \rho_0)x^{m_E}\}, \end{aligned}$$

for $1 \leq j \leq J_E$. If $\{r_{E,j}\}$ is empty, then there is no ‘bad’ region generated by this edge and the only two ‘good’ regions are

$$U_{0,g}(E) = \{(x, y) \in U_0 : c_E x^{m_E} < |y| < C_E x^{m_E}\}.$$

The following lemma states that P behaves almost like a monomial in each ‘good’ region $U_{0,g}(E, j)$.

Lemma 4.2. *Let $N > 0$ and $L > 0$ be given. For any choices of c_E 's, C_E 's and ρ_0 above, one can choose ϵ_0 sufficiently small, such that for all $E \in \mathcal{N}(P)$ and $(x, y) \in U_{0,g}(E, j)$, one has*

$$(4.7) \quad |x^{p_{E,l}} y^{q_{E,l}}| \sim_{\rho_0} |P_E(x, y)| \geq 2^N |P(x, y) - P_E(x, y)|.$$

Here $(p_{E,l}, q_{E,l})$ is the left vertex of the edge E . In addition,

$$(4.8) \quad |\partial_x^\alpha \partial_y^\beta P(x, y)| < C \min\{1, |x^{p_{E,l}-\alpha} y^{q_{E,l}-\beta}|\}$$

for $0 \leq \alpha, \beta \leq L$.

Proof. In the region $y = rx^{m_E}$ where $r \in I(E)$,

$$|P(x, y) - P_E(x, y)| < C x^{e_E+\nu},$$

where ν is a positive fraction (can be computed but not necessary). By (4.4), one has $|P_E(r)| \geq C$ for $r \in I_g(E)$, where $C = C(c_{E_E}, C_E, \rho_0, P) > 0$. Thus if $\epsilon_0 = \epsilon_0(\rho_0, \nu, C)$ is sufficiently small, then for all $(x, y) \in U_{0,g}(E, j)$ we have

$$|P_E(x, y)| \sim_{\rho_0} |x^{p_{E,l}-\alpha} y^{q_{E,l}-\beta}| \sim x^{e_E} > 2^N \cdot O(x^{e_E+\nu}) > 2^N |P(x, y) - P_E(x, y)|,$$

which proves (4.7).

Now we turn to (4.8). The bound $|\partial_x^\alpha \partial_y^\beta P(x, y)| \lesssim 1$ is trivial. In the region $y = rx^{m_E}$ where $r \in I(E)$, for $0 \leq \alpha, \beta \leq L$ and every $(p'', q'') \in E$, one has

$$|x^{p_{E,l}-\alpha} y^{q_{E,l}-\beta}| \sim |x^{p''-\alpha} y^{q''-\beta}| \sim |x|^{e_E-\alpha-m_E\beta},$$

even for $p'' - \alpha < 0$ or/and $q'' - \beta < 0$. Notice that $|y| \sim |x|^{m_E}$; then

$$|\partial_x^\alpha \partial_y^\beta (P(x, y) - P_E(x, y))| \lesssim |x|^{e_E+\nu-\alpha-m_E\beta}.$$

Thus given $|x|$ sufficiently small, one has

$$|\partial_x^\alpha \partial_y^\beta P(x, y)| \lesssim |x^{p_{E,l}-\alpha} y^{q_{E,l}-\beta}|.$$

This completes the the proof of (4.8). \square

The above lemma handled the case when an edge E is dominant. Another easy case is when a vertex $V = (p_v, q_v)$ plays the dominant role. In this case, let E_l and E_r be the edges left and right to V , with slopes $-1/m_{E_l}$ and $-1/m_{E_r}$ respectively. Then $0 \leq m_{E_l} < m_{E_r} \leq \infty$. Consider the following region:

$$(4.9) \quad U_{0,g}(V) = \{(x, y) \in U_0 : C_{E_r} x^{m_{E_r}} < |y| < c_{E_l} x^{m_{E_l}}\},$$

where C_{E_r} and c_{E_l} are the constants chosen in (4.1). One can always choose ϵ_0 small enough such that the two curves $y = C_{E_r} x^{m_{E_r}}$ and $y = c_{E_l} x^{m_{E_l}}$ do not meet inside U_0 . In the case $m_{E_r} = \infty$, i.e., V is the rightmost vertex, set

$$(4.10) \quad U_{0,g}(V) = \{(x, y) \in U_0 : |y| < c_{E_l} x^{m_{E_l}}\},$$

which also contains the portion of the x -axis inside U_0 . Similarly, when $m_{E_l} = 0$, i.e., V is the leftmost vertex, we replace $c_{E_l} x^{m_{E_l}}$ by ϵ_0 in (4.9).

The following lemma is the vertex analogue of Lemma 4.2, whose proof is similar to that of Lemma 4.2.

Lemma 4.3. *Given preselected numbers N and L , one can choose ϵ_0 sufficiently small (depending on N , L , C_E 's, c_E 's and P), such that for $(x, y) \in U_{0,g}(V)$ one has, for $0 \leq \alpha, \beta \leq L$,*

$$(4.11) \quad |x^{p_v} y^{q_v}| \sim |P_V(x, y)| \geq 2^N |P(x, y) - P_V(x, y)|,$$

$$(4.12) \quad |\partial_x^\alpha \partial_y^\beta P(x, y)| < C \min\{1, |x^{p_v - \alpha} y^{q_v - \beta}|\}.$$

Now let

$$\mathcal{G}_0(P_0) = \mathcal{G}_0(P) = \{U_{0,g}(V) : V \in \mathcal{V}(P)\} \cup \{U_{0,g}(E, j) : E \in \mathcal{E}(P) \text{ and all } j\},$$

which are the collection of ‘good’ regions in the 0-th stage. In addition, we say $U_{0,g} \in \mathcal{G}_0(P_0)$ is defined by (E, m_E) if $U_{0,g} = U_{0,g}(E, j)$ for some E and j , where $-1/m_E$ is the slope of E , or defined by an edge E for short. Similarly, $U_{0,g} \in \mathcal{G}_0(P_0)$ is defined by (V, m_l, m_r) represents $U_{0,g} = U_{0,g}(V)$ and $-1/m_l, -1/m_r$ are the slopes of the edges left and right to V , or defined by a vertex V for short.

Now we focus on the ‘bad’ regions

$$(4.13) \quad U_{0,b}(E, j) = \{(x, y) \in U_0 : (r_j - \rho_0)x^{m_E} < y < (r_j + \rho_0)x^{m_E}\}.$$

Set

$$\mathcal{B}_0(P_0) = \mathcal{B}_0(P) = \{U_{0,b}(E, j) : E \in \mathcal{E}(P) \text{ and } 1 \leq j \leq J_E\}$$

which is the collection of ‘bad’ regions in the 0-stage. If $U_{0,b} \in \mathcal{B}_0(P_0)$ has the form of (4.13), we say $U_{0,b}$ is defined by $(E, y = r_j x^{m_E})$ or defined by $y = r_j x^{m_E}$ for short. The following graph demonstrates a partition of U (in the first quadrant) into ‘good’ and ‘bad’ regions, according to the analytic function $P(x, y) = xy(y^2 - x)(y - \sum_{n=1}^{\infty} x^n)^2$. Notice that the function $R(x, y) = xy(y^2 - x)(y - x^2)(y - x)^2$ also has the same partition. Here, we choose $P(x, y)$ for the purpose of illustrating the convergence of the algorithm later.

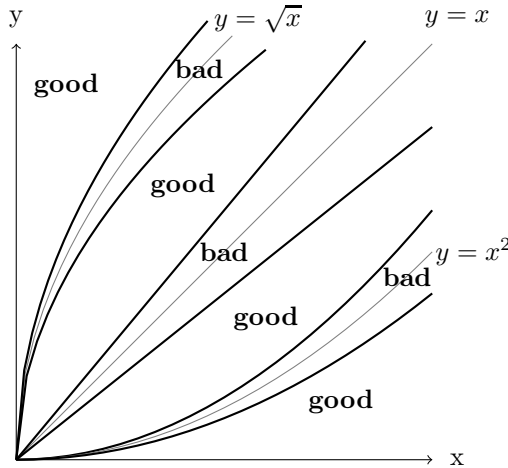


Figure 3. ‘Good’ regions and ‘bad’ regions of $P(x, y) = xy(y^2 - x)(y - x^2)(y - \sum_{n=1}^{\infty} x^n)^2$ in the first quadrant.

We summarize the above discussion as follows:

Proposition 4.4 (Single step of partition). *Let U be a standard region and P be a real analytic function. If the size of U is sufficiently small, then U can be partitioned into two families of curved triangular regions: $\mathcal{G}_0(P)$ and $\mathcal{B}_0(P)$. For each $U_{0,g} \in \mathcal{G}_0(P)$, $U_{0,g}$ is defined by (4.5) or (4.9). The behaviors of P in $U_{0,g}$ are characterized by Lemma 4.2 or Lemma 4.3. Each $U_{0,b} \in \mathcal{B}_0(P)$ is defined by (4.13). Finally, the cardinalities of $\mathcal{G}_0(P)$ and $\mathcal{B}_0(P)$ are finite, depending on P .*

4.2. The resolution algorithm, Part II. Iterations

The next step is to iterate Proposition 4.4. One main problem is that the region $U_{0,b} \in \mathcal{B}_0(P)$ is not standard. Nevertheless, via an appropriate change of variables, we can always turn a $U_{0,b}$ into a subset of a standard region U_1 . Proposition 4.4 can then be applied to $[U_1, P_1]$, where P_1 is obtained from $P_0 = P$ via this change of variables. Notice that the arguments in the previous subsection work equally well for analytic functions of $(x^{1/M}, y)$. The flowchart below illustrates the main ideas of how the algorithm runs. The letter ‘g’ represents a ‘good’ region, while ‘b’ represents a bad region. Each time, we choose one ‘bad’ region and ‘rescale’ it (via change of variable) into a subset of a standard region.

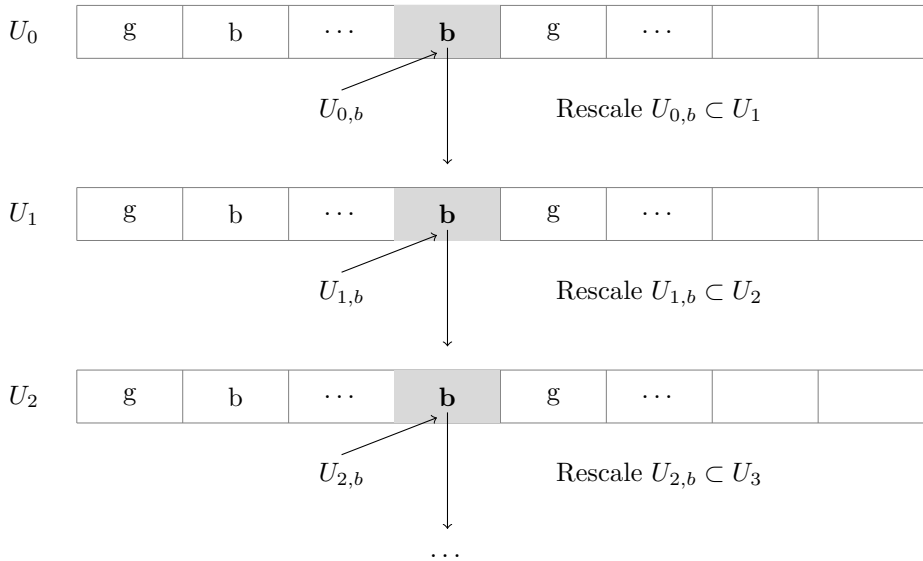


Figure 4.

Before diving into the details, we introduce the following notations which can help keep track on certain invariants as the iterations go deeper.

Definition 4.5. Let (p_l, q_l) and (p_r, q_r) be the leftmost and rightmost vertices of $\mathcal{N}(P)$, the Heights of $\mathcal{N}(P)$ or P are defined as

$$\begin{aligned} \text{Hght}(\mathcal{N}(P)) &= \text{Hght}(P) = q_l - q_r, \\ \text{Hght}^*(\mathcal{N}(P)) &= \text{Hght}^*(P) = q_l. \end{aligned}$$

For an edge $E \in \mathcal{E}(P)$, let $(p_{E,l}, q_{E,l})$ and $(p_{E,r}, q_{E,r})$ be its left and right vertices. Then the height of this edge is defined to be

$$\text{Hght}(E) = q_{E,l} - q_{E,r}.$$

If $\{r_{E,j}\}_{1 \leq j \leq J_E}$ is the set of non-zero roots of $P_E(r)$ of orders $\{s_{E,j}\}_{1 \leq j \leq J_E}$, then we define the order of E as

$$\text{Ord}(E) = \sum_{j=1}^{J_E} s_{E,j},$$

and the order of P to be

$$\text{Ord}(P) = \sum_{E \in \mathcal{E}(P)} \text{Ord}(E) = \sum_{E \in \mathcal{E}(P)} \sum_{j=1}^{J_E} s_{E,j}.$$

Finally, we say r is a root of $P(x, y)$ or $\mathcal{N}(P)$ if $r = r_{E,j}$ for some $E \in \mathcal{E}(P)$ and some $1 \leq j \leq J_E$.

Under the above notation,

$$\text{Ord}(E) \leq \text{Hght}(E) = q_{E,l} - q_{E,r}$$

and

$$\text{Ord}(P) \leq \text{Hght}(P) = q_l - q_r \leq q_l = \text{Hght}^*(P).$$

Notice that

$$\#\mathcal{B}_0(P) = \sum_{E \in \mathcal{E}(P)} J_E \leq \sum_{E \in \mathcal{E}(P)} \text{Ord}(E) = \text{Ord}(P),$$

which gives:

Lemma 4.6.

$$\#\mathcal{B}_0(P) \leq \text{Ord}(P) \leq \text{Hght}^*(P).$$

Choose a $U_{0,b} \in \mathcal{B}_0(P_0)$ and assume it is defined by $y = r_0 x^{m_0}$. The next step is to utilize change of variables to turn $[U_{0,b}, P]$ into a standard pair. Adopt the previous notations $[U_0, P_0] = [U, P]$ and $(x_0, y_0) = (x, y)$ and choose x to be the principal variable which will be unchanged during iterations, i.e., $x = x_n$ for all $n \in \mathbb{N}$. Change variables

$$\begin{cases} x_0 = x_1, \\ y_0 = (r_0 + y_1) x_1^{m_0}. \end{cases}$$

Then the region $U_{0,b}$ under the new coordinates/variables is

$$\begin{cases} 0 < x_1 < \epsilon_0, \\ -\rho_0 < y_1 < \rho_0. \end{cases}$$

Since $\epsilon_0 \neq \rho_0$, this region is not standard, but can be embedded into a larger standard region U_1 . Let ϵ_1 be sufficiently small but greater than ρ_0 . Set

$$\begin{cases} P_1(x_1, y_1) = P(x_1, (r_0 + y_1)x_1^{m_0}), \\ U_1 = \{(x_1, y_1) : 0 < x_1 < \epsilon_1, -\epsilon_1 < y_1 < \epsilon_1\}. \end{cases}$$

Then $[U_1, P_1]$ is a standard pair in the coordinates (x_1, y_1) . By applying Proposition 4.4 to $[U_1, P_1]$, a finite collection $\mathcal{G}_1(P_1)$ of ‘good’ regions $U_{1,g}$ ’s and a finite collection $\mathcal{B}_1(P_1)$ of ‘bad’ regions $U_{1,b}$ ’s are obtained. In a ‘good’ region $U_{1,g}$, the function $P_1(x_1, y_1)$ behaves like a monomial of (x_1, y_1) and no further treatment is required. For the ‘bad’ regions, choose a $U_{1,b} \in \mathcal{B}_1(P_1)$ and assume $U_{1,b}$ is defined by $y_1 = r_1 x_1^{m_1}$, i.e.,

$$U_{1,b} = \{(x_1, y_1) \in U_1 : (r_1 - \rho_1)x_1^{m_1} < y_1 < (r_1 + \rho_1)x_1^{m_1}\}.$$

where ρ_1 is an analogue of ρ_0 in the 0-th stage of iteration. Like what has been done, perform the following change of variables,

$$\begin{cases} x_1 = x_2, \\ y_1 = (r_1 + y_2)x_2^{m_1}, \end{cases}$$

choose ϵ_2 sufficiently small but greater than ρ_1 , and set

$$\begin{cases} P_2(x_2, y_2) = P_1(x_2, (r_1 + y_2)x_2^{m_1}), \\ U_2 = \{(x_2, y_2) : 0 < x_2 < \epsilon_2, -\epsilon_2 < y_2 < \epsilon_2\}. \end{cases}$$

Same procedures are repeated on the standard pair $[U_2, P_2]$ and so on. Consequently, a collection of standard pairs $\{[U_n, P_n]\}$ is obtained from these iterations. Notice that this collection forms a tree structure with $[U_0, P_0] = [U, P]$ being the top, with the parameter n representing the level of iterations. Sometimes we need more information than merely the level of iterations. In those cases, we use the notation $[U_n, P_n] = [U_{n,\alpha}, P_{n,\alpha}]$, where

(\star) α contains the information of the change of variables, i.e., for $0 \leq k \leq n-1$ the following is known if α is given:

$$(4.14) \quad \begin{cases} x_k = x_{k+1}, \\ y_k = (r_k + y_{k+1})x_{k+1}^{m_k}. \end{cases}$$

We also use $U_{n,b,\alpha}$ and $U_{n,g,\alpha}$ to represent any ‘bad’ region and any ‘good’ region in $U_{n,\alpha}$. Since $U_{n,\alpha}$ may have more than one such regions, we list them by $U_{n,b,\alpha,j}$ and $U_{n,g,\alpha,j}$ when necessary. The cardinality in j is uniformly bounded (see Lemma 4.8) and thus there is no need to specify its range. Notice that $U_{n,g,\alpha,j}$ is a leaf of the tree, i.e., it has no child and no further analysis is needed.

Both notations are being used here: with and without the subindex α . To avoid confusion, we follow the rules below:

(1) The one without such subindex is our primary choice. We often use $[U_n, P_n]$ to represent an arbitrary pair from the n -th stage of iteration.

(2) The other is used occasionally. It is employed typically when at least two different pairs from the same stage of iterations are mentioned simultaneously.

These conventions also apply to $U_{n,g}$'s, $U_{n,b}$'s, $U_{n,g,\alpha}$ and $U_{n,b,\alpha}$.

Example 4.7. The following graphs demonstrate the first step of the algorithm, for the given analytic function $P(x, y) = xy(y^2 - x)(y - x^2)(y - \sum_{n=1}^{\infty} x^n)^2$. The Newton polyhedron $\mathcal{N}(P)$ has 3 compact edges: E_1 , E_2 and E_3 , see Figure 5 below.

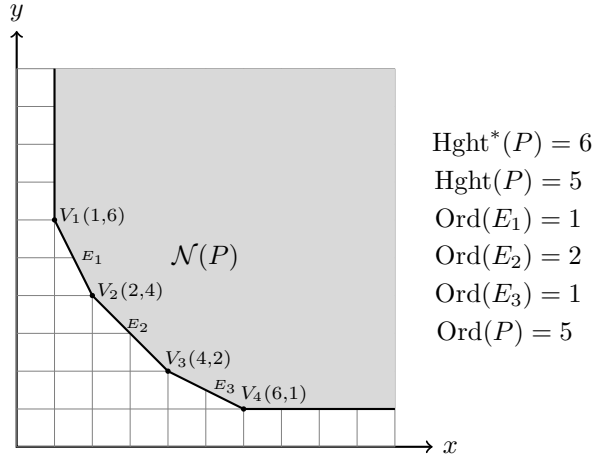


Figure 5.

In the edge E_1 , $P_{E_1} = x(y^2 - x)y^4$ has only two non-zero roots and the corresponding ‘bad’ regions are defined by $y = x^{1/2}$ and $y = -x^{1/2}$. To handle the former, change variables: $(x, y) = (x_1, x_1^{1/2}(1 + y_1))$ yields $P_1(x_1, y_1) = P(x_1, x_1^{1/2}(1 + y_1)) = x_1^{7/2}y_1 \cdot O(1)$. Here $O(1)$ represents a function which is non-vanishing at the origin. In Figure 6, $\mathcal{N}(P_1)$ has only one vertex and the algorithm stops. The latter is similar.

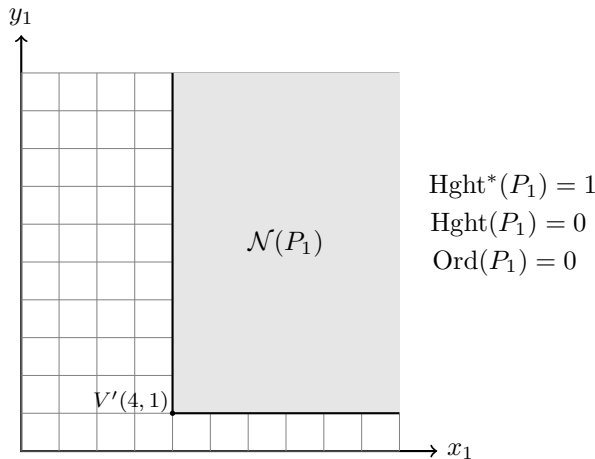


Figure 6.

There is only one root for $P_{E_2} = -x^2y^2(y - x)^2$ and the corresponding ‘bad’ region is defined by $y = x$. Change variables: $(x, y) = (x_1, x_1(1 + y_1))$ yields $P_1(x_1, y_1) = x_1^6(y_1 - \sum_{n=1}^{\infty} x^n)^2 \cdot O(1)$. But $\mathcal{N}(P_1)$ still has one edge which also has one non-zero root; see Figure 7. The algorithm does not stop. Doing the change variables $(x_{k-1}, y_{k-1}) = (x_k, x_k(1 + y_k))$ only gives $\mathcal{N}(P_k) = \mathcal{N}(P_{k-1}) + (2, 0)$ for $k \geq 2$.

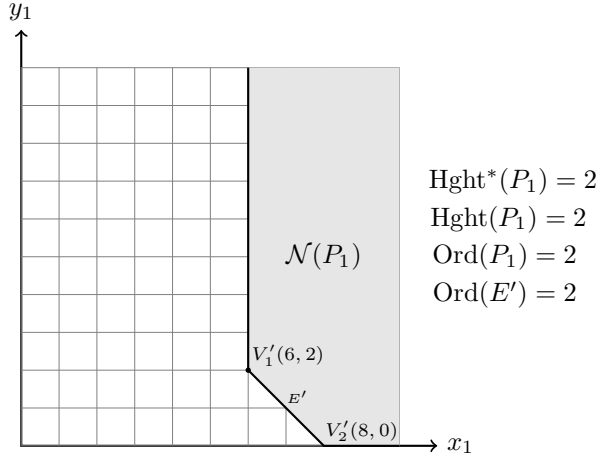


Figure 7.

Finally, $P_{E_3} = -x^4y(y - x^2)$. The only non-zero root is $y = x^2$. Change of variables: $(x, y) = (x_1, x_1^2(1 + y_1))$ yields $P_1(x_1, y_1) = x_1^8y_1 \cdot O(1)$. The algorithm stops; see Figure 8.

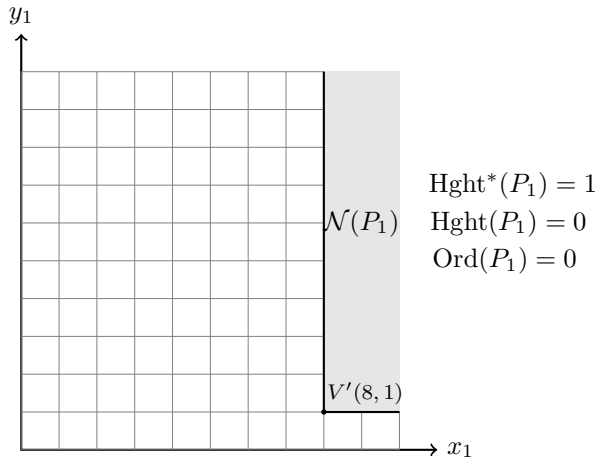


Figure 8.

It is worth mentioning that the change of variables: $(x_n, y_n) \rightarrow (x_{n+1}, y_{n+1})$ maps one-to-one from $U_{n,b}$ to a rectangular subset of U_{n+1} . Thus one can also

embed such subset back into U_n and etc.. In addition, if the change of variables: $(x, y) \rightarrow (x_n, y_n)$ is specified, then we can use both coordinates (x, y) and (x_n, y_n) to identify points in U_n (or $U_{n,b}, U_{n,g}$). The relation between the two coordinates $(x, y) = \rho_n^{-1}(x_n, y_n)$ is given by

$$(4.15) \quad \begin{cases} x = x_n, \\ y = r_0 x^{m_0} + r_1 x^{m_0+m_1} + \cdots + r_{n-1} x^{m_0+\cdots+m_{n-1}} + y_n x^{m_0+\cdots+m_{n-1}}. \end{cases}$$

In the above notation, the $(k+1)$ -th stage of iteration is generated by the ‘bad’ region defined by $y_k = r_k x^{m_k}$ for $0 \leq k \leq n-1$. Under this notation, $P_n = P \circ \rho_n^{-1}$. We should also specify ρ_n^{-1} to $\rho_{n,\alpha}^{-1}$, if U_n is specified to $U_{n,\alpha}$.

To complete the algorithm, we need to settle the following two questions:

(i) In each stage of iteration, is the cardinality of $\{U_{n,b,\alpha,j}\}_{\alpha,j}$ bounded above uniformly?

(ii) Does this procedure terminate after a finite steps?

The answer to the second question is more delicate and extra preparation is needed, but the first can be answered by:

Lemma 4.8. *For each $n \geq 0$, the cardinality of $\{U_{n,b,\alpha,j}\}_{\alpha,j}$ is bounded above by $\text{Ord}(P)$.*

Proof. Indeed, there is a bijection between $\{U_{0,b}\}$ and the non-zero roots of $P_E(r)$ $E \in \mathcal{E}(P)$: each $U_{0,b} = U_{0,b}(E, j)$ is defined by $(E, y = r_{E,j} x^{m_E})$. Assume the order of $r_{E,j}$ is $s_{E,j}$. Then $P_1(x_1, y_1) = P_{1,E,j}(x_1, y_1)$ is obtained by setting $P_1(x_1, y_1) = P_0(x_1, (y_1 + r_{E,j}) x_1^{m_E})$. Here, $P_{1,E,j}$ is used to specify that P_1 is defined by the root $r_{E,j}$.

The following observation serves as an bridge between $\mathcal{N}(P_0)$ and $\mathcal{N}(P_1)$. Let $(p_{E,l}, q_{E,l})$ be the left vertex of E and $(p_{1,l}, q_{1,l})$ be the leftmost vertex of $\mathcal{N}(P_1)$. Then

$$(4.16) \quad \begin{cases} p_{1,l} = p_{E,l} + m_E \cdot q_{E,l}, \\ q_{1,l} = s_{E,j}. \end{cases}$$

This implies the order $s_{E,j}$ (in the 0-stage) is equal to the $\text{Hght}^*(P_{1,E,j})$.

To prove (4.16), notice that

$$P_E(x, y) = P_E(x_1, (y_1 + r_{E,j}) x_1^{m_E}) = x_1^{p_{E,l} + m_E \cdot q_{E,l}} y_1^{s_{E,j}} \cdot O(1).$$

Indeed, the fact that the degree of y_1 is $s_{E,j}$ follows from the fact that $r_{E,j}$ is a root of $P_E(r)$ of order $s_{E,j}$. Moreover, every term in

$$P_1(x_1, y_1) - P_E(x_1, (y_1 + r_{E,j}) x_1^{m_E})$$

has a x_1 -degree strictly greater than $(p_{E,l} + m_E \cdot q_{E,l})$. Consequently, $(p_{E,l} + m_E \cdot q_{E,l}, s_{E,j})$ is the leftmost vertex of $\mathcal{N}(P_1)$.

Thus $\text{Ord}(P_1) \leq \text{Hght}^*(P_1) = s_{E,j}$ and the cardinality of ‘bad’ regions $\{U_{1,b}\}$ from a single P_1 is at most $\text{Ord}(P_1) \leq s_{E,j}$. Counting all possible P_1 (coming from different roots of different edges), the number of all possible $U_{1,b}$ is bounded by

$$(4.17) \quad \sum_{E \in \mathcal{E}(P)} \sum_{1 \leq j \leq J_E} \text{Ord}(P_{1,E,j}) \leq \sum_{E \in \mathcal{E}(P)} \sum_{1 \leq j \leq J_E} s_{E,j} = \text{Ord}(P).$$

The cases when $n \geq 2$ also follow from iterating (4.17). \square

We now turn to the second question, which is the most crucial part of the algorithm. Assume the procedure does not stop. We then obtain an infinite chain of pairs:

$$(4.18) \quad [U_0, P_0] \rightarrow [U_1, P_1] \rightarrow [U_2, P_2] \rightarrow \cdots \rightarrow [U_n, P_n] \rightarrow [U_{n+1}, P_{n+1}] \rightarrow \cdots$$

We shall find certain patterns inside this chain. Specify the change of variables from $[U_n, P_n] \rightarrow [U_{n+1}, P_{n+1}]$ as

$$\begin{cases} x_n = x_{n+1}, \\ y_n = (r_n + y_{n+1}) x_{n+1}^{m_n}. \end{cases}$$

Then r_n is a root of an edge in $\mathcal{N}(P_n)$. Let s_n be the order of r_n and $(p_{n,l}, q_{n,l})$ be the leftmost vertex of $\mathcal{N}(P_n)$ and (p_n, q_n) be the left vertex of the edge in $\mathcal{N}(P_n)$ that defines $[U_{n+1}, P_{n+1}]$. By (4.16) one has

$$(4.19) \quad \begin{cases} p_{n+1} \geq p_{n+1,l} = p_n + m_n \cdot q_n, \\ q_{n+1} \leq q_{n+1,l} = s_n \leq q_n, \end{cases}$$

and thus

$$(4.20) \quad \begin{aligned} \text{Hght}^*(P_0) &\geq \text{Hght}(P_0) \geq s_0 = \text{Hght}^*(P_1) \geq \text{Hght}(P_1) \geq s_1 = \text{Hght}^*(P_2) \\ &\geq \cdots \geq s_{n-1} = \text{Hght}^*(P_n) \geq \text{Hght}(P_n) \geq s_n = \text{Hght}^*(P_{n+1}) \cdots \end{aligned}$$

Notice that for all n , $\text{Hght}(P_n)$ and s_n must be positive integers. Otherwise, if $\text{Hght}(P_n) = 0$ then $\mathcal{N}(P_n)$ has no compact edge and thus no root; if $s_n = 0$, then $\mathcal{N}(P_n)$ has no root. In both situations, the chain ends at the n -stage, which contradicts to our assumption.

Since (4.20) is an infinite sequence and $\text{Hght}^*(P_0)$ is a finite positive number, there is a least integer $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ one has

$$(4.21) \quad \text{Hght}^*(P_n) = \text{Hght}(P_n) = s_n = \text{Hght}^*(P_{n_0}) = \text{Hght}(P_{n_0}) = s_{n_0} > 0.$$

This implies for every $n \geq n_0$:

- (i) $\mathcal{N}(P_n)$ has only one compact edge E_n ,
- (ii) in this edge E_n , $P_n(x_n, y_n)$ has only one root r_n of order $s_n = s_{n_0}$,
- (iii) when P_n is restricted to E_n , $P_{n,E_n}(x_n, y_n) = c_n(y_n - r_n x_n^{m_n})^{s_{n_0}}$, for some non-zero constant c_n .

This is the exact pattern we are looking for, which suggests $P_{n_0}(x_{n_0}, y_{n_0})$ has a factor

$$y_{n_0} - \sum_{n=n_0}^{\infty} r_n x_{n_0}^{m_{n_0}+m_{n_0+1}+\dots+m_n},$$

whose order is equal to s_{n_0} . In Example 4.7, after the change of variable $y = x_1 + x_1 y_1$, the function $P_1(x_1, y_1)$ has only one compact edge, which is factored as $(y_1 - x_1)^2$. From $n_0 = 1$, if we keep doing change of variables $y_{n-1} = x_n + x_n y_n$, we will have

$$P_{n, E_n}(x_n, y_n) = c_n (y_n - x_n)^2.$$

Indeed, in this example $P_1(x_1, y_1)$ has a factor $(y_1 - \sum_{n=1}^{\infty} x_1^n)$ of order exactly 2.

The following lemma shows that the chain (4.18) essentially ends at the (n_0+1) -stage.

Lemma 4.9. *Assume we have an infinite chain (4.18) and n_0 is the constant defined in (4.21). Then there is a positive integer M such that*

$$(4.22) \quad P_{n_0}(x_{n_0}, y_{n_0}) = x_{n_0}^{p_{n_0}} (y_{n_0} - f(x_{n_0}))^{s_{n_0}} Q_{n_0}(x_{n_0}, y_{n_0}),$$

where

$$f(x_{n_0}) = \sum_{n=n_0}^{\infty} r_n x_{n_0}^{m_{n_0}+m_{n_0+1}+\dots+m_n}$$

is an analytic function of $x_{n_0}^{1/M}$, and $Q_{n_0}(x_{n_0}, y_{n_0})$ is an analytic function of $(x_{n_0}^{1/M}, y_{n_0})$ with $Q_{n_0}(0, 0) \neq 0$.

Proof. To obtain $P_{n_0}(x_{n_0}, y_{n_0})$ from $P_0(x, y)$, we have iterated only finitely many steps. Thus $P_{n_0}(x_{n_0}, y_{n_0})$ is a real analytic function of $(x_{n_0}^{1/M}, y_{n_0})$, for some $M \in \mathbb{N}$. For $n \geq n_0$, the change of variables from $[U_n, P_n]$ to $[U_{n+1}, P_{n+1}]$ is $x_n = x_{n+1}$ and $y_n = (y_{n+1} + r_n)x_n^{m_n}$. The only compact edge E_n of P_n is of the form

$$(4.23) \quad P_{n, E_n}(x_n, y_n) = c_n x_n^{p_n} (y_n - r_n x_n^{m_n})^{s_n},$$

where c_n is a nonzero constant and $s_n = s_{n_0}$. Using induction, it is not difficult to prove that $m_n M$ is an integer for all $n \geq n_0$. Thus $P_n(x_n, y_n)$ is a real analytic function of $(x_n^{1/M}, y_n)$ for all $n \in \mathbb{N}$. The Weierstrass preparation theorem implies

$$(4.24) \quad P_{n_0}(x_{n_0}, y_{n_0}) = x_{n_0}^{p_{n_0}} Q_{n_0}(x_{n_0}, y_{n_0}) R(x_{n_0}, y_{n_0}),$$

where $Q_{n_0}(x_{n_0}, y_{n_0})$ is a non-vanishing real analytic function of $(x_{n_0}^{1/M}, y_{n_0})$ and

$$R(x_{n_0}, y_{n_0}) = y_{n_0}^{s_0} + a_{s_0-1}(x_{n_0})y_{n_0}^{s_0-1} + a_{s_0-2}(x_{n_0})y_{n_0}^{s_0-2} + \dots + a_0(x_{n_0})$$

is a Weierstrass polynomial, i.e., a polynomial in y with analytic coefficients in $x_{n_0}^{1/M}$. Our goal is to verify the following factorization:

$$(4.25) \quad R(x_{n_0}, y_{n_0}) = (y_{n_0} - f(x_{n_0}))^{s_{n_0}}$$

in the sense of formal power series. Notice that this also implies that $f(x_{n_0})$ is a real analytic function of $x_{n_0}^{1/M}$ for it equals $-a_{s_0-1}(x_{n_0})/s_{n_0}$. Therefore, (4.25) is true pointwise in some small neighborhood of the origin.

First, notice $(p_n + s_n m_n, 0)$ is the rightmost vertex of $\mathcal{N}(P_n)$. By setting $y_n = 0$, (4.23) yields

$$(4.26) \quad P_n(x_n, 0) = C_n x_n^{p_n + s_n m_n} + O(x_n^{p_n + s_n m_n + \nu}),$$

where $\nu > 0$. Consider the partial sum of $f(x_{n_0})$,

$$(4.27) \quad f_k(x_{n_0}) = \sum_{n=n_0}^k r_n x_{n_0}^{m_{n_0} + m_{n_0+1} + \dots + m_n}, \quad k \geq n_0.$$

Then $y_{n_0} = y_n x_{n_0}^{m_{n_0} + m_{n_0+1} + \dots + m_{n-1}} + f_{n-1}(x_{n_0})$, $n \geq n_0 + 1$, and

$$P_n(x_n, y_n) = P_{n_0}(x_n, y_n x_{n_0}^{m_{n_0} + m_{n_0+1} + \dots + m_{n-1}} + f_{n-1}(x_{n_0})).$$

By (4.26), we have

$$(4.28) \quad P_{n_0}(x_{n_0}, f_{n-1}(x_{n_0})) = P_n(x_n, 0) = C_n x_n^{p_n + s_n m_n} + O(x_n^{p_n + s_n m_n + \nu}).$$

Notice that $m_n \geq 1/M$ and $s_n = s_{n_0}$ is a positive integer, thus

$$p_n + s_n m_n \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

which implies (in the sense of formal power series)

$$(4.29) \quad P_{n_0}(x_{n_0}, f(x_{n_0})) = 0 = R(x_{n_0}, f_{n_0}(x_{n_0})),$$

since $Q_{n_0}(x_{n_0}, y_{n_0})$ is non-vanishing. This yields $R_{n_0}(x_{n_0}, y_{n_0})$ has a factor $(y_{n_0} - f(x_{n_0}))$. To show its order is exactly s_{n_0} , assume

$$(4.30) \quad R(x_{n_0}, y_{n_0}) = (y_{n_0} - f(x_{n_0}))^s \bar{R}(x_{n_0}, y_{n_0}),$$

where $\bar{R}_{n_0}(x_{n_0}, f(x_{n_0})) \neq 0$. Write

$$\bar{R}(x_{n_0}, y_{n_0}) = \bar{R}(x_{n_0}, f(x_{n_0})) + (\bar{R}(x_{n_0}, y_{n_0}) - \bar{R}(x_{n_0}, f(x_{n_0})))$$

and the latter term is divisible by $(y_{n_0} - f(x_{n_0}))$. Assume the leading term of $\bar{R}(x_{n_0}, f(x_{n_0}))$ is $C x_{n_0}^A$ (the nonzero term with lowest degree). Then

$$\bar{R}(x_{n_0}, f_{n-1}(x_{n_0})) = C x_{n_0}^A + O(x_{n_0}^{A+\nu}) \quad \text{as } n \rightarrow \infty$$

for some $\nu > 0$. Combining (4.30) and (4.24), one has

$$(4.31) \quad P_{n_0}(x_{n_0}, f_{n-1}(x_{n_0})) = C x_{n_0}^{p_{n_0} + s(m_{n_0} + \dots + m_n) + A} + O(x_{n_0}^{p_{n_0} + s(m_{n_0} + \dots + m_n) + A + \nu}),$$

as $n \rightarrow \infty$. Notice that, for all $n > n_0$,

$$(4.32) \quad p_n = p_{n-1} + s_{n-1} m_{n-1} = p_{n-1} + s_{n_0} m_{n-1}.$$

Iterating this identity yields

$$(4.33) \quad p_n + s_n m_n = p_{n_0} + s_{n_0} (m_{n_0} + \cdots + m_n).$$

Comparing (4.28) and (4.31) yields

$$A = 0 \quad \text{and} \quad s = s_0,$$

as desired. \square

Remark 4.10. The purpose of employing the Weierstrass preparation theorem (WPT) here is to verify the analyticity of $f(x_{n_0})$, but its use is not essential here. There should be other methods. For example, the referees kindly point out that one can employ the IFT instead. I find the use of the WPT here is quite interesting, for $(y_{n_0} - f(x_{n_0}))^{s_0}$ is indeed the Weierstrass polynomial of $P_{n_0}(x_{n_0}, y_{n_0})$. One may be able to use the algorithm in this paper to compute the Weierstrass polynomial of the original function $P(x, y)$.

Based on Lemma 4.9, in the n_0 -stage, we refine the change of variables as follows:

$$(4.34) \quad x_{n_0} = x_{n_0+1} \quad \text{and} \quad y_{n_0} - f(x_{n_0}) = y_{n_0+1} x_{n_0+1}^{m_{n_0}}.$$

This will help us to eliminate all the nonzero roots of the new function P_{n_0+1} , and thus eliminate the ‘bad’ regions. Indeed, by (4.22) one has

$$(4.35) \quad \begin{aligned} P_{n_0+1}(x_{n_0+1}, y_{n_0+1}) &:= P_{n_0}(x_{n_0+1}, y_{n_0+1} x_{n_0+1}^{m_{n_0}} + f(x_{n_0+1})) \\ &= x_{n_0+1}^{p_{n_0} + s_{n_0} m_{n_0}} y_{n_0+1}^{s_{n_0}} Q_{n_0+1}(x_{n_0+1}, y_{n_0+1}), \end{aligned}$$

where

$$Q_{n_0+1}(x_{n_0+1}, y_{n_0+1}) = Q_{n_0}(x_{n_0+1}, y_{n_0+1} x_{n_0+1}^{m_{n_0}} + f(x_{n_0+1})),$$

is non-vanishing near the origin. One can see that $\mathcal{N}(P_{n_0+1})$ has only one vertex and $P_{n_0+1}(x_{n_0+1}, y_{n_0+1})$ behaves like a monomial. Set

$$U_{n_0+1} = \{(x_{n_0+1}, y_{n_0+1}) : (x_{n_0}, y_{n_0}) \in U_{n_0, b}\},$$

then $U_{n_0+1, g} = U_{n_0+1}$, $U_{n_0+1, b} = \emptyset$ and the procedure ends here. Thus if we take N_P to be the maximum of all such $n_0 + 1$ from all branches of iterations, then the resolution algorithm ends at the N_P -stage. Set

$$(4.36) \quad \mathcal{G}_n = \cup_{\alpha} \cup_j \{U_{n, g, \alpha, j}\},$$

$$(4.37) \quad \mathcal{V}_n = \cup_{\alpha} \mathcal{V}(P_n, \alpha),$$

$$(4.38) \quad \mathcal{E}_n = \cup_{\alpha} \mathcal{E}(P_n, \alpha),$$

which represent the ‘good’ regions, vertices and compact edges in the n -th stage respectively. The followings represent all the ‘good’ regions, vertices and compact edges in all stages:

$$(4.39) \quad \mathcal{G} = \cup_{0 \leq n \leq N_P} \mathcal{G}_n = \cup_n \cup_{\alpha} \cup_j \{U_{n, g, \alpha, j}\},$$

$$(4.40) \quad \mathcal{V} = \cup_{0 \leq n \leq N_P} \mathcal{V}_n = \cup_n \cup_{\alpha} \mathcal{V}(P_n, \alpha),$$

$$(4.41) \quad \mathcal{E} = \cup_{0 \leq n \leq N_P} \mathcal{E}_n = \cup_n \cup_{\alpha} \mathcal{E}(P_n, \alpha).$$

For $n \geq 1$, we often refer to $U_{n,g,\alpha,j}$ as a ‘good’ region defined by an edge of $E \in \mathcal{E}(P)$ in higher stage of iteration. Thus each ‘good’ region is either defined by a vertex $V \in \mathcal{V}(P)$ or by an edge $E \in \mathcal{E}(P)$, or simply say defined by a face $F \in \mathcal{F}(P)$.

We can now determine ϵ_0 , the size of the original region U_0 . One needs to go backward, starting from every leaf of the tree, i.e., every ‘good’ region to the top. Each such path (from a chosen leaf to the top of the tree) will give rise to an upper bound of ϵ_0 and one can just take ϵ_0 to be any number smaller than all those bounds. We briefly describe how to do it. Notice first that all the roots r_E ’s from different edges of different stages of iterations are independent from sizes of the regions U_n ’s, so are all assigned constants c_E ’s and C_E ’s defined in (4.1). For convenience, all the $\{\epsilon_n\}$ and $\{\epsilon'_n\}$ chosen below are less than $2^{-10}c_E$ for all E .

- (1) Take one unchosen leaf from the tree. We use $[U_{n_0+1}, P_{n_0+1}]$ in the above case as our example here. Recall that ϵ_{n_0+1} denotes the size of U_{n_0+1} .
- (2) There is an of upper bound $\epsilon'_{n_0+1} > 0$, such that for every $\epsilon_{n_0+1} < \epsilon'_{n_0+1}$, the function P_{n_0+1} behaves like a monomial in U_{n_0+1} . One can adjust the value of ϵ'_{n_0+1} to control the relative smallness of the error term. Choose one such ϵ_{n_0+1} and move up to $[U_{n_0}, P_{n_0}]$.
- (3) Choose $0 < \rho_{n_0} < \epsilon_{n_0+1}$ and set

$$\xi_{n_0} = \inf\{|P_E(r)| : r \in I_g(E)\}$$

where the infimum runs over all compact edges of $\mathcal{N}(P_{n_0})$ and $I_g(E)$ is given in (4.3) which relies on ρ_{n_0} . Then there is a $0 < \epsilon'_{n_0} < \epsilon_{n_0+1}$ depending on ρ_{n_0}, ξ_{n_0} and P_{n_0} , such that for every $\epsilon_{n_0} < \epsilon'_{n_0}$, the function P_{n_0} behaves like a monomial in all the ‘good’ regions of U_{n_0} . Choose one such ϵ_{n_0} and move up to $[U_{n_0-1}, P_{n_0-1}]$.

- (4) Iterate (3) until we reach the top $[U_0, P_0]$, which will gives rise to one ϵ'_0 .
- (5) Iterate (1), (2), (3) and (4) until there is no unchosen leaf. We then obtain finitely many ϵ'_0 ’s. Choose one $\epsilon = \epsilon_0$ less than any of them.

Theorem 4.11. *For each analytic function P defined in a neighborhood of $0 \in \mathbb{R}^2$, there is a standard region U , which can be partitioned into a finite collection of ‘good’ regions $\{\rho_n^{-1}(U_{n,g}) \cap U, U_{n,g} \in \mathcal{G}\}$, where*

$$\rho_n^{-1}(x_n, y_n) = (x, y) = (x_0, y_0)$$

is given by

$$(4.42) \quad \begin{cases} x_n = x_0 = x, \\ y_n = \gamma_n(x) + y_n x^{m_0 + \dots + m_{n-1}}, \end{cases}$$

where $\gamma_n(x)$ is defined as follows. If $[U_n, P_n]$ lies in an infinite chain

$$(4.43) \quad [U_0, P_0] \rightarrow [U_1, P_1] \rightarrow \dots \rightarrow [U_n, P_n] \rightarrow [U_{n+1}, P_{n+1}] \rightarrow \dots$$

and $n - 1 = n_0$ for some n_0 defined as in (4.21), then

$$(4.44) \quad \gamma_n(x) = \sum_{k=0}^{\infty} r_k x_0^{m_0+m_1+\dots+m_k},$$

which converges for x with $(x, y) \in U$; otherwise

$$(4.45) \quad \gamma_n(x) = \sum_{k=0}^{n-1} r_k x_0^{m_0+m_1+\dots+m_k}.$$

We assume also that the order of r_k is s_k . For $0 \leq k \leq n$, let $(p_{k,l}, q_{k,l})$ be the leftmost vertex of $\mathcal{N}(P_k)$. For $0 \leq k \leq n - 1$, let (p_k, q_k) be the left vertex of the edge that defines $U_{k,b} \subset U_k$, and (p_n, q_n) be defined as follows: if $U_{n,g}$ is defined by a vertex V , then $(p_n, q_n) = V$; else (p_n, q_n) is the left vertex of the edge that defines $U_{n,g}$. One has

$$(4.46) \quad q_n \leq q_{n,l} \leq s_{n-1} \leq q_{n-1} \leq q_{n-1,l} \leq \dots \leq q_1 \leq s_0 \leq q_0.$$

For every supporting line $L_{m_n} \in \mathcal{SL}(P_n)$ through (p_n, q_n) , one also has

$$(4.47) \quad \begin{aligned} p_n + m_n q_n &\leq p_0 + m_0 q_0 + m_1 q_{1,l} + \dots + m_n q_{n,l} \\ &\leq p_0 + q_0 m_0 + s_0 (m_1 + \dots + m_n). \end{aligned}$$

In addition, for any given $L \in \mathbb{N}$, for all $0 \leq \alpha, \beta \leq L$ and $(x, y) = \rho_n^{-1}(x_n, y_n) \in \rho_n^{-1}(U_{n,g}) \cap U$ one has

$$(4.48) \quad |P(x, y)| = |P_n(x_n, y_n)| \sim |x_n^{p_n} y_n^{q_n}|$$

$$(4.49) \quad |\partial_{x_n}^\alpha \partial_{y_n}^\beta P_n(x_n, y_n)| \lesssim \min\{1, |x_n^{p_n-\alpha} y_n^{q_n-\beta}|\}$$

$$(4.50) \quad |\partial_y^\beta P(x, y)| \lesssim \min\{1, |x^{p_n-\beta(m_0+\dots+m_{n-1})} y_n^{q_n-\beta}|\}.$$

Remark 4.12. (i) This theorem also works for $U = (-\epsilon, \epsilon) \times (-\epsilon, \epsilon)$ (with a smaller ϵ) and thus any open subset of U . First of all, we can apply this theorem to a standard region in the left half plane, i.e., to $U = (-\epsilon, 0) \times (-\epsilon, \epsilon)$ by setting $x = -x$. Secondly, one can incorporate the y -axis into the ‘good’ regions defined by the leftmost vertex of $\mathcal{N}(P)$; see (4.10) for the case of incorporating the x -axis.

(ii) In applications, the estimate (4.48) is often used in conjunction with (4.46) and (4.47) to relate the lower bound of $|P(x, y)|$ in ‘good’ regions from higher stages of iterations to the original Newton polyhedron of P . The estimates in (4.46) and (4.47) contain more details than are needed for the proof of Theorem 1.5. For example the middle term of (4.47) is not used explicitly in this proof. We include such details for the purpose of future reference, since we may need strong information from higher stages of iterations.

Proof of Theorem 4.11. The partition in the theorem is a consequence of the algorithm. The region $\rho_{n,g}^{-1}(U_{n,g})$ does not necessarily all lie in U and hence we only

need to take the portion in U , i.e., $\rho_{n,g}^{-1}(U_{n,g}) \cap U$. The change of variables (4.42) is obtained by iterating

$$(4.51) \quad \begin{cases} x_k = x_{k+1}, \\ y_k = (r_k + y_{k+1}) x_{k+1}^{m_k}. \end{cases}$$

The estimate (4.46) follows directly from (4.19) and (4.47) is a consequence of iterating the following estimates:

$$p_n + m_n q_n \leq p_{n,l} + m_n q_{n,l} = (p_{n-1} + m_{n-1} q_{n-1}) + m_n q_{n,l},$$

where first estimate comes from the fact that $(p_{n,l}, q_{n,l})$ lies on or above the supporting line L_{m_n} and the identity is due to (4.19).

In addition, (4.48) and (4.49) come from Lemma 4.2 and Lemma 4.3. Finally, (4.50) is an outcome of (4.42) and the chain rule, since

$$\partial y / \partial y_n = x^{m_0 + \dots + m_{n-1}}. \quad \square$$

4.3. A smooth partition

In the above theorem, U is decomposed into disjoint ‘good’ regions $\rho_n^{-1}(U_{n,g})$ ’s. However, when it comes to application (in analysis), an overlap version is often more suitable since it provides extra room to fit in a smooth partition. Furthermore, there is an extra benefit in our problem: this extra room can help us to overcome the convexity assumption in Theorem 2.1. To be precise, the ‘good’ regions are in general not convex, even if they are localized to pieces with the x -supports being dyadic intervals. If we take the convex hull of one such piece, as needed in Theorem 2.1, the convex hull may lie in more than one ‘good’ regions. If so, in such convex hull we may fail to find a uniform monomial to bound $|P(x, y)|$ from below. Fortunately, the algorithm here provides sufficient flexibility to adjust the boundary of ‘good’ regions, which allows us to cut each of them into a number of pieces so that each piece

1. is small enough with the convex hull lying essentially in the same ‘good’ region; and
2. is large enough with the total number of pieces bounded above uniformly by a harmless constant; See Lemma 5.5 and Lemma 5.6.

For these reasons, we slightly enlarge $U_{n,g}$ to $U_{n,g} \subset U_{n,g}^* \subset U_{n,g}^{**}$ and $U_{n,b}$ to $U_{n,b} \subset U_{n,b}^* \subset U_{n,b}^{**}$. The $U_{n,g}^*$ ’s are used to fit in a smooth partition and $U_{n,g}^{**}$ ’s are used to address the convexity assumption. The first step is to enlarge $I(E)$, $I_g(E)$ and $I_b(E)$ as follows:

$$\begin{aligned} I^*(E) &= [\tfrac{1}{2}c_E, 2C_E], & I^{**}(E) &= [\tfrac{1}{4}c_E, 4C_E], \\ I_g^*(E) &= I^*(E) \setminus (\cup_{1 \leq j \leq J_E} I_j^{\rho_0/2}(E)), & I_g^{**}(E) &= I^{**}(E) \setminus (\cup_{1 \leq j \leq J_E} I_j^{\rho_0/4}(E)) \end{aligned}$$

and

$$I_b^*(E) = \cup_{1 \leq j \leq J_E} I_j^{2\rho_0}(E), \quad I_b^{**}(E) = \cup_{1 \leq j \leq J_E} I_j^{4\rho_0}(E).$$

Notice that our choices of ρ_0 in (4.2) ensures that for all E , $\{I_j^{4\rho_0}(E)\}$ do not overlap. Then we can define the enlarged ‘good’ regions as

$$\begin{aligned} U_{0,g}^*(E) &= \{(x, y) \in U_0 : y = rx^{m_E}, r \in I_g^*(E)\}, \\ U_{0,g}^{**}(E) &= \{(x, y) \in U_0 : y = rx^{m_E}, r \in I_g^{**}(E)\}. \end{aligned}$$

Both $U_{0,g}^*(E)$ and $U_{0,g}^{**}(E)$ consist of (J_E+2) curved triangular regions $\{U_{0,g}^*(E, j)\}$ and $\{U_{0,g}^{**}(E, j)\}$, respectively. In addition, one has

$$U_{0,g}(E, j) \subset U_{0,g}^*(E, j) \subset U_{0,g}^{**}(E, j).$$

The ‘good’ regions defined by a vertex are enlarged to:

$$\begin{aligned} U_{0,g}^*(V) &= \{(x, y) \in U_0 : \frac{C_{E_r}}{2} x^{m_{E_r}} < y < 2c_{E_l} x^{m_{E_l}}\}, \\ U_{0,g}^{**}(V) &= \{(x, y) \in U_0 : \frac{C_{E_r}}{4} x^{m_{E_r}} < y < 4c_{E_l} x^{m_{E_l}}\}, \end{aligned}$$

and finally the enlarged ‘bad’ regions are

$$\begin{aligned} U_{0,b}^*(E, j) &= \{(x, y) \in U_0 : (r_j - 2\rho_0) x^{m_E} < y < (r_j + 2\rho_0) x^{m_E}\}, \\ U_{0,b}^{**}(E, j) &= \{(x, y) \in U_0 : (r_j - 4\rho_0) x^{m_E} < y < (r_j + 4\rho_0) x^{m_E}\}. \end{aligned}$$

For $n \geq 1$, $U_{n,g}^*$ ’s, $U_{n,b}^*$ ’s and $U_{n,g}^{**}$ ’s, $U_{n,b}^{**}$ ’s are defined similarly. Since ρ_k can be chosen sufficiently small, the above definitions do not cause any conflict. We have:

Corollary 4.13. *Let U and P as in Theorem 4.11. Then all the estimates in Theorem 4.11 hold with $U_{n,g}$ replaced by $U_{n,g}^{**}$. In addition, U is contained in the union of all $U_{n,g}^{**}$.*

For a smooth partition, we issue some technical problems first. Let c be a positive constant far away from all the roots of $P_E(x, y)$ for all E , say $c > 2^{10}C_E$ for C_E defined in (4.1) and for all $E \in \mathcal{E}(P)$. Choose a constant $\epsilon = 2^{-5}c$. Then $y = \pm cx$ divides the plane into four regions: R_1, R_2, R_3 and R_4 , which represents the East, North, West and South regions respectively. Let $\{\Psi_j\}_{1 \leq j \leq 4}$ be smooth functions such that

$$1 = \sum_{j=1}^4 \Psi_j(x, y), \quad (x, y) \neq (0, 0).$$

Here $\Psi_1(x, y)$ is supported in

$$R_1^\epsilon = \{(x, y) : x > 0, -(c + \epsilon)x < y < (c + \epsilon)x\},$$

and $\Psi_1(x, y) = 1$ if

$$-(c - \epsilon)x < y < (c - \epsilon)x.$$

In addition, Ψ_1 satisfies

$$|\partial_y^\beta \Psi_1(x, y)| \leq C_\beta |y|^{-\beta}, \quad \text{for any } \beta \in \mathbb{N}.$$

The other functions Ψ_2, Ψ_3 and Ψ_4 are defined similarly in the other regions.

For a given analytic function $P(x, y)$, let W be a small neighborhood of 0 so that one can apply Theorem 4.11 and Corollary 4.13 to $[W, P]$. Let $\Phi(x, y)$ be a smooth function adapted to W , in the sense $\text{supp } \Phi \subset W$ and $\Phi(x, y) = 1$ if $(2x, 2y) \in W$. Then

$$\Phi(x, y) = \Phi(x, y) \sum_{j=1}^4 \Psi_j(x, y), \quad \text{for } (x, y) \neq (0, 0).$$

We focus on $\Phi\Psi_1$, since discussions for the others are similar. Let

$$U = W \cap R_1^\epsilon;$$

then $\Phi\Psi_1$ is supported in U .

We now apply the resolution algorithm to $P(x, y)$ in the region U , which gives rise to a collection of ‘bad’ regions $\{U_{n,b,\alpha,j}^*\}_{(n,\alpha,j)}$. For a fixed $U_{n,b,\alpha,j}^*$, $\rho_{n,\alpha}^{-1}(U_{n,b,\alpha,j}^*)$ is equal to

$$U \cap \{\gamma_n(x) + (r_n - 2\rho_n) x^{m_0+\dots+m_n} < y < \gamma_n(x) + (r_n + 2\rho_n) x^{m_0+\dots+m_n}\},$$

where $\gamma_n(x)$ is given by (4.44) or (4.45). The assumption that U is contained in the east region R_1^ϵ ensure that all the m_0 is at least 1, i.e., the power of the leading term of $\gamma_n(x)$ is at least 1.

We can then define a smooth function $\Phi_{n,b,\alpha,j}$ supported in $\rho_{n,\alpha}^{-1}(U_{n,b,\alpha,j}^*)$ and $\Phi_{n,b,\alpha,j}(x, y) = 1$ if $(x, y) \in \rho_{n,\alpha}^{-1}(U_{n,b,\alpha,j}^*)$. In addition, the following is true

$$(4.52) \quad |\partial_y^\beta \Phi_{n,b,\alpha,j}(x, y)| \leq C_\beta |x|^{-\beta(m_0+\dots+m_n)} \quad \forall \beta \in \mathbb{N}.$$

Then

$$\Phi(x, y)\Psi_1(x, y) \left(1 - \sum_{\alpha} \sum_j \Phi_{0,b,\alpha,j}(x, y)\right)$$

can be written as

$$\sum_{\alpha} \sum_j \Phi_{0,g,\alpha,j}(x, y),$$

where each $\Phi_{0,g,\alpha,j}(x, y)$ is supported in the ‘good’ region $U_{0,g,\alpha,j}^*$. Similarly,

$$\Phi(x, y)\Psi_1(x, y) \left(\sum_{\alpha} \sum_j \Phi_{0,b,\alpha,j}(x, y)\right) \left(1 - \sum_{\alpha} \sum_j \Phi_{1,b,\alpha,j}(x, y)\right)$$

can be written as

$$\sum_{\alpha} \sum_j \Phi_{1,g,\alpha,j}(x, y),$$

where $\Phi_{1,g,\alpha,j}$ is supported in $\rho_{1,\alpha}^{-1}(U_{1,g,\alpha,j}^*)$. Then we iterate the above procedures until the end of the algorithm. Combining (4.52), we obtain a smooth partition version of Theorem 4.11.

Theorem 4.14. *Let Φ , Ψ_1 and P be as above. Then*

$$(4.53) \quad \Phi(x, y)\Psi_1(x, y) = \sum_n \sum_{\alpha} \sum_j \Phi_{n,g,\alpha,j}(x, y),$$

where $\Phi_{n,g,\alpha,j}(x,y)$ is a smooth function supported in $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*)$ and where $\{U_{n,g,\alpha,j}\}$ is the collection of ‘good’ regions as in Theorem 4.11. The behaviors of $P(x,y)$ in ‘good’ regions $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*) \cap U$ and $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^{**}) \cap U$ are the same as Corollary 4.13. Moreover, $\Phi_{n,g,\alpha,j}(x,y)$ satisfies the following derivative conditions:

(1) If $U_{n,g,\alpha,j}$ is defined by an edge, then $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*)$ is given by in a curved triangular region of the form

$$b_{n,g,\alpha,j} x^{m_0+\dots+m_n} \leq y - \gamma_n(x) \leq B_{n,g,\alpha,j} x^{m_0+\dots+m_n}$$

for some nonzero constants $b_{n,g,\alpha,j}$ and $B_{n,g,\alpha,j}$ with the same signs. Then

$$(4.54) \quad |\partial_y^\beta \Phi_{n,g,\alpha,j}(x,y)| \leq C_\beta |x|^{-\beta(m_0+\dots+m_n)}, \quad \forall \beta \in \mathbb{N}.$$

(2) Otherwise, $U_{n,g,\alpha,j}$ is defined by a vertex. The region $U_{n,g,\alpha,j}$ may lie only in the first quadrant, only in the fourth quadrant or in both of them. In the first case, $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*)$ is given by

$$b_{n,g,\alpha,j} x^{m_0+\dots+m_{n-1}+m_{n,r}} < y - \gamma_n(x) < B_{n,g,\alpha,j} x^{m_0+\dots+m_{n-1}+m_{n,l}},$$

where $0 \leq m_{n,l} < m_{n,r} < \infty$. In the upper portion of $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*) \setminus \rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j})$, one has

$$|\partial_y^\beta \Phi_{n,g,\alpha,j}(x,y)| \leq C_\beta |x|^{-\beta(m_0+\dots+m_{n-1}+m_{n,l})}, \quad \forall \beta \in \mathbb{N}$$

and in the lower portion

$$|\partial_y^\beta \Phi_{n,g,\alpha,j}(x,y)| \leq C_\beta |x|^{-\beta(m_0+\dots+m_{n-1}+m_{n,r})}, \quad \forall \beta \in \mathbb{N}.$$

The second case is similar. In the last case, $U_{n,g,\alpha,j}^*$ is defined by the rightmost vertex of $\mathcal{N}(P_n, \alpha)$, and $\rho_{n,\alpha}^{-1}(U_{n,g,\alpha,j}^*)$ is given by $|y - \gamma_n(x)| \lesssim x^{m_0+\dots+m_{n-1}+m_{n,l}}$, one has

$$|\partial_y^\beta \Phi_{n,g,\alpha,j}(x,y)| \leq C_\beta |x|^{-\beta(m_0+\dots+m_{n-1}+m_{n,l})}, \quad \forall \beta \in \mathbb{N}.$$

5. Proof of Theorem 1.5

Set $P = \partial_x \partial_y (\partial_x - \partial_y) S$ and assume $a(x,y)$ is supported in $W/100$, where W is a neighborhood of 0 such that we can apply Theorem 4.11 and Corollary 4.13 to $[W, P]$. As what was done in Theorem 4.14, we divide W into 4 regions by lines $y = cx$ and $y = -cx$ and restrict our discussion in the east region

$$(5.1) \quad U = W \cap \{(x,y) : x > 0, -(c+\epsilon)x < y < (c+\epsilon)x\},$$

since the others can be reduced to U by either changing x to $-x$ or permuting x and y , or both. Let $\Psi_j(x,y)$ and $\Phi(x,y)$ be smooth functions as in Theorem 4.14. Then

$$a(x,y) = a(x,y) \Phi(x,y) \sum_{j=1}^4 \Psi_j(x,y), \quad \text{for } (x,y) \neq (0,0),$$

and

$$\Lambda_S(f_1, f_2, f_3) = \sum_{j=1}^4 \Lambda_S^j(f_1, f_2, f_3),$$

where

$$\Lambda_S^j(f_1, f_2, f_3) = \iint e^{i\lambda S(x,y)} a(x, y) \Phi(x, y) \Psi_j(x, y) f_1(x) f_2(y) f_3(x+y) dx dy.$$

We only focus on $j = 1$. Theorem 4.14 yields

$$a(x, y) \Phi(x, y) \Psi_1(x, y) = \sum_{0 \leq n \leq N_P} \sum_{\alpha} \sum_j a_{n,g,\alpha,j}(x, y),$$

where

$$(5.2) \quad a_{n,g,\alpha,j}(x, y) = \Phi_{n,g,\alpha,j}(x, y) a(x, y).$$

Set

$$(5.3) \quad \Lambda_{n,g,\alpha,j}(f_1, f_2, f_3) = \iint e^{i\lambda S(x,y)} f_1(x) f_2(y) f_3(x+y) a_{n,g,\alpha,j}(x, y) dx dy;$$

then

$$(5.4) \quad \Lambda_S^1(f_1, f_2, f_3) = \sum_{0 \leq n \leq N_P} \sum_{\alpha} \sum_j \Lambda_{n,g,\alpha,j}(f_1, f_2, f_3).$$

Since the summands in (5.4) have only finitely many terms, it suffices to bound each of them separately. For simplicity, we use $\Lambda_{n,g}$ to represent $\Lambda_{n,g} = \Lambda_{n,g,\alpha,j}$ for some α and j . Then Theorem 1.5 follows from:

Theorem 5.1. *If $U_{n,g}$ is defined by $F \in \mathcal{F}(P)$, then*

$$\|\Lambda_{0,g}\| \lesssim |\lambda|^{-\frac{1}{2(3+d_F)}}.$$

We split the proof into three cases (i) $n = 0$ and $U_{0,g}$ is defined by an edge E , (ii) $n = 0$ and $U_{0,g}$ is defined by a vertex V and (iii) $n \geq 1$.

Proposition 5.2. *If $U_{0,g}$ is defined by an edge E , then*

$$\|\Lambda_{0,g}\| \lesssim |\lambda|^{-\frac{1}{2(3+d_E)}}.$$

Proposition 5.3. *If $U_{0,g}$ is defined by a vertex V , then*

$$\|\Lambda_{0,g}\| \lesssim |\lambda|^{-\frac{1}{2(3+d_V)}}.$$

Proposition 5.4. *If $n \geq 1$ and $U_{n,g}$ is defined by an edge E , then*

$$\|\Lambda_{n,g}\| \lesssim |\lambda|^{-\frac{1}{2(3+d_E)}}.$$

Heuristically, one can consider Proposition 5.2 as an estimate of Λ_S in one supporting line, i.e., in $E \in \mathcal{SL}(P)$, and Proposition 5.3 as the sum/integral of such estimates over all supporting lines through V . Although the function $P_n(x_n, y_n)$ becomes more singular as n increases, the region $U_{n,g}$ becomes nicer, which allows one to exploit certain orthogonality. The rest of this section is devoted to the proofs of the above propositions.

5.1. Proof of Proposition 5.2

Let $0 < \sigma < 1$ be a dyadic number and let $\phi_\sigma(x)$ be a smooth function supported in $\sigma/2 < x < 2\sigma$ such that

$$\sum_{0 < \sigma < 1} \phi_\sigma(x) = 1 \quad \text{for } 0 < x < 1/10.$$

Let $\Lambda_{0,g,\sigma_1,\sigma_2}(f_1, f_2, f_3)$ denote

$$\iint e^{i\lambda S(x,y)} f_1(x) f_2(y) f_3(x+y) a_{0,g}(x,y) \phi_{\sigma_1}(x) \phi_{\sigma_2}(y) dx dy,$$

where σ_1 and σ_2 are positive dyadic numbers and $a_{0,g} = a_{0,g,\alpha,j}$ for some (α, j) . Notice that

$$\text{supp}(a_{0,g}) \subset U_{0,g}^*,$$

which is a ‘good’ region defined by E . Notice that $m = \mathcal{M}(E) \geq 1$ due to (5.1). Thus

$$(5.5) \quad |y| \sim x^m \quad \text{and} \quad \sigma_2 \sim \sigma_1^m \leq \sigma_1.$$

This yields, for fixed σ_1 , there are only finitely many choices of σ_2 . We lose no generality in assuming, for a given σ_1 , that σ_2 is fixed. To apply Theorem 2.1, we need to verify its assumptions. For $K \in \mathbb{N}$, equally divide the interval $(\sigma_1/2, 2\sigma_1)$ into K subintervals $\{I_k\}_{1 \leq k \leq K}$, and set

$$U_{0,g,k}^* = \{(x, y) \in U_{0,g}^* : x \in I_k\}.$$

Lemma 5.5. *There is a constant $K \in \mathbb{N}$, dependent of σ_1 and σ_2 , such that*

$$\text{Conv}(U_{0,g,k}^*) \subset U_{0,g}^{**},$$

for all $1 \leq k \leq K$.

The proof of this lemma is postponed to the end of this section. Now let (p_l, q_l) be the left vertex of E . Then for every $(x, y) \in \text{Conv}(U_{0,g,k}^*) \subset U_{0,g}^{**}$, Theorem 4.11 and Corollary 4.13 yield

$$(5.6) \quad |P(x, y)| \gtrsim |x|^{p_l} |y|^{q_l} \sim \sigma_1^{p_l} \sigma_2^{q_l},$$

and for $\beta = 0, 1, 2$,

$$(5.7) \quad |\partial_y^\beta P(x, y)| \lesssim |x|^{p_l} |y|^{q_l - \beta} \sim \sigma_1^{p_l} \sigma_2^{q_l - \beta}.$$

Theorem 4.14 together with (5.2) yields

$$(5.8) \quad |\partial_y^\beta a_{0,g}(x, y)| \lesssim \sigma_2^{-\beta}.$$

By invoking Theorem 2.1, one has

$$\|\Lambda_{0,g,\sigma_1,\sigma_2,k}\| \lesssim |\lambda \sigma_1^{p_l} \sigma_2^{q_l}|^{-1/6},$$

where $\Lambda_{0,g,\sigma_1,\sigma_2,k}(f_1, f_2, f_3)$ is given by

$$\iint e^{i\lambda S(x,y)} f_1(x) \mathbf{1}_{I_k}(x) f_2(y) f_3(x+y) a_{0,g}(x,y) \phi_{\sigma_1}(x) \phi_{\sigma_2}(y) dx dy.$$

Summing over $1 \leq k \leq K$ (K is a constant) yields

$$(5.9) \quad \|\Lambda_{0,g,\sigma_1,\sigma_2}\| \lesssim |\lambda \sigma_1^{p_l} \sigma_2^{q_l}|^{-1/6}.$$

Employing Lemma 1.7, (5.5) and combining (5.9), one obtains

$$(5.10) \quad \|\Lambda_{0,g,\sigma_1,\sigma_2}\| \lesssim \begin{cases} |\lambda \sigma_1^{p_l} \sigma_2^{q_l}|^{-1/6} \sim |\lambda \sigma_2^{q_l+p_l/m}|^{-1/6}, \\ \min\{\sigma_1, \sigma_2\}^{1/2} = \sigma_2^{1/2}. \end{cases}$$

Since for fixed σ_2 , σ_1 is fixed, summing over σ_2 yields

$$\left\| \sum_{\sigma_1, \sigma_2} \Lambda_{0,g,\sigma_1,\sigma_2} \right\| \lesssim |\lambda|^{-\frac{1}{2(3+q_l+p_l/m)}} = |\lambda|^{-\frac{1}{2(3+d_E)}},$$

as desired.

5.2. Proof of Proposition 5.3

Like in the proof of Proposition 5.2, insert the smooth support $\phi_{\sigma_1}(x)\phi_{\sigma_2}(y)$ into $\Lambda_{0,g}$. Set $V = (p, q)$ and assume $-1/m_l$ and $-1/m_r$ be the slopes of the edges left and right to V . Due to the assumption on U (5.1), we may replace m_l with 1 if $m_l < 1$. Thus,

$$(5.11) \quad \infty \geq m_r > m_l \geq 1.$$

Notice that

$$\sigma_2^{1/m_r} \gtrsim \sigma_1 \gtrsim \sigma_2^{1/m_l} \gtrsim \sigma_2.$$

Consider all (σ_1, σ_2) with $\sigma_2 \lesssim \lambda_2 := |\lambda|^{-\frac{1}{3+q+p/m_l}} = |\lambda|^{-\frac{1}{3+d_V}}$. By Lemma 1.7 and the triangle inequality, we have

$$(5.12) \quad \left\| \sum_{\sigma_2 \lesssim \lambda_2} \left(\sum_{\sigma_1} \Lambda_{0,g,\sigma_1,\sigma_2} \right) \right\| \lesssim \sum_{\sigma_2 \lesssim \lambda_2} |\sigma_2|^{1/2} \lesssim |\lambda_2|^{1/2} = |\lambda|^{-\frac{1}{2(3+d_V)}}.$$

Now assume $\sigma_2 \gtrsim \lambda_2$. As it was done in the proof of Proposition 5.2, we invoke the same trick to meet the convexity assumption of Theorem 2.1: splitting $\Lambda_{0,g,\sigma_1,\sigma_2}$ into the sum of $\Lambda_{0,g,\sigma_1,\sigma_2,k}$, applying Theorem 2.1 to each $\Lambda_{0,g,\sigma_1,\sigma_2,k}$ and summing them together. This gives

$$\|\Lambda_{0,g,\sigma_1,\sigma_2}\| \lesssim |\lambda \sigma_1^p \sigma_2^q|^{-1/6}.$$

Since $\sigma_1 \gtrsim \sigma_2^{1/m_l}$, thus

$$\sum_{\sigma_1} \|\Lambda_{0,g,\sigma_1,\sigma_2}\| \lesssim |\lambda \sigma_2^{p/m_l+q}|^{-1/6}.$$

Summing all $\sigma_2 \gtrsim \lambda_2$ we obtain the same bound as (5.12).

5.3. Proof of Proposition 5.4

Assume the change of variables $(x, y) = \rho_k^{-1}(x_k, y_k)$ is given as in Theorem 4.11, as well as the parameters r_k, s_k and (p_k, q_k) for $0 \leq k \leq n$. Theorem 4.11 and Corollary 4.13 yield

$$|P(x, y)| = |P_n(x_n, y_n)| \sim |x_n^{p_n} y_n^{q_n}|$$

for all $(x_n, y_n) \in U_{n,g}^{**}$. Since $U_{n,g}^{**} \subset U_n^{**}$ is a ‘good’ region, we can find m'_n and m_n such that

$$|x_n|^{m'_n} \lesssim |y_n| \lesssim |x_n|^{m_n},$$

where $0 \leq m_n \leq m'_n \leq \infty$. In addition, if $m_n = m'_n$, then $U_{n,g}^{**}$ is defined by an edge; otherwise by a vertex. Dyadically decompose $x_n \sim \sigma_1$ and let Λ_{n,g,σ_1} denote the operator $\Lambda_{n,g}$ when x_n is localized to this region.

Due to the ‘almost orthogonality’ below, we only need to handle one single σ_1 .

Claim 1. *There is a constant L such that if*

$$(5.13) \quad \|\Lambda_{n,g,\sigma_1}\| \leq A$$

for some $A > 0$ then

$$(5.14) \quad \sum_{\sigma_1} \|\Lambda_{n,g,\sigma_1}\| \leq LA.$$

Proof of Claim 1. Recall that

$$(5.15) \quad y = y(x) = \gamma_n(x) + y_n x^{m_0 + \dots + m_{n-1}}$$

where

$$\gamma_n(x) = \sum_{k=0}^* r_k x_0^{m_0 + m_1 + \dots + m_k}.$$

with $* = n - 1$ or ∞ ; see (4.44) and (4.45). Set

$$Y(\sigma) = \{y(x) : x \sim \sigma\}.$$

For σ small enough, one has $|y| \sim |r_0 \sigma^{m_0}|$ for $y \in Y(\sigma)$. Thus, one can find $L \in \mathbb{N}$ such that

$$(5.16) \quad Y(\sigma_1) \cap Y(\sigma_2) = \emptyset$$

given $\sigma_1 \geq 2^L \sigma_2$ and σ_1 is small.

Consider the congruence classes modulo L for all small σ_1 . Let

$$H_\ell = \{\sigma_1 = 2^{-h} : h \equiv \ell \pmod{L}\}$$

for $0 \leq \ell < L$. Notice that

$$\Lambda_{n,g,\sigma_1}(f_1, f_2, f_3) = \Lambda_{n,g,\sigma_1}(f_1 \mathbf{1}_{I_{\sigma_1}}, f_2 \mathbf{1}_{Y(\sigma_1)}, f_3),$$

where I_{σ_1} denotes such x with $x \sim \sigma_1$. The triangle inequality, the Cauchy–Schwartz inequality and (5.20) yield

$$\begin{aligned} \left| \sum_{\sigma_1 \in H_\ell} \Lambda_{n,g,\sigma_1}(f_1, f_2, f_3) \right| &\leq \sum_{\sigma_1 \in H_\ell} \left| \Lambda_{n,g,\sigma_1}(f_1 \mathbf{1}_{I_{\sigma_1}}, f_2 \mathbf{1}_{Y(\sigma_1)}, f_3) \right| \\ &\leq A \sum_{\sigma_1 \in H_\ell} \|f_1 \mathbf{1}_{I_{\sigma_1}}\|_2 \|f_2 \mathbf{1}_{Y(\sigma_1)}\|_2 \|f_3\|_2 \\ &\leq A \left\| \sum_{\sigma_1 \in H_\ell} f_1 \mathbf{1}_{I_{\sigma_1}} \right\|_2 \left\| \sum_{\sigma_1 \in H_\ell} f_2 \mathbf{1}_{Y(\sigma_1)} \right\|_2 \|f_3\|_2, \end{aligned}$$

which is controlled by $A\|f_1\|_2 \|f_2\|_2 \|f_3\|_2$ due to (5.16). The desired estimates then follow by the triangle inequality:

$$\left\| \sum_{\sigma_1} \Lambda_{n,g,\sigma_1} \right\| \leq L \sup_{0 \leq \ell < L} \left\| \sum_{\sigma_1 \in H_\ell} \Lambda_{n,g,\sigma_1} \right\| \leq LA. \quad \square$$

Dyadically decompose (x_n, y_n) as

$$\begin{cases} |x_n| \sim \sigma_1, \\ |y_n| \sim \sigma_2 x_n^{m_n} \sim \sigma_2 \sigma_1^{m_n}. \end{cases}$$

Use $U_{n,g}(\sigma_1, \sigma_2)$ and $\Lambda_{n,g,\sigma_1,\sigma_2}$ to denote the corresponding localization of the ‘good’ region $U_{n,g}$ and the operator $\Lambda_{n,g}$ respectively. In what follows, we shall focus on one single σ_1 and prove a bound for $\|\Lambda_{n,g,\sigma_1}\|$ that is independent of σ_1 .

Even in the region $U_{n,g}(\sigma_1, \sigma_2)$, there is some orthogonality that one can employ. In fact, the change of variables (5.15) yields that the length of every x -section of $U_{n,g}(\sigma_1, \sigma_2)$ denoted by Δy is bounded by

$$(5.17) \quad \Delta y \lesssim \sigma_2 \sigma_1^{m_0 + \dots + m_n}.$$

In addition, (5.15) gives

$$\Delta y \sim \Delta x \frac{dy}{dx} \sim \Delta x \cdot \sigma_1^{m_0 - 1},$$

and thus

$$(5.18) \quad \Delta x \sim \sigma_2 \sigma_1^{m_0 + \dots + m_n - (m_0 - 1)}.$$

Notice also $m_0 \geq 1$ due to the assumption $|y| \lesssim |x|$ on U . We then divide the interval $(\sigma_1/2, 2\sigma_1)$ equally into H subintervals $\{I_h\}_{1 \leq h \leq H}$, where

$$H = K \cdot \sigma_1 (\sigma_2 \sigma_1^{m_0 + \dots + m_n - (m_0 - 1)})^{-1}.$$

Here K is a large constant, independent of (σ_1, σ_2) and designed to treat the convexity assumption in Theorem 2.1. Set

$$U_{n,g,h}^* = \{(x, y) \in U_{n,g}^* : x \in I_h\}$$

and

$$Y(I_h) = \{y(x) : x \in I_h \text{ and } y_n \sim \sigma_2 \sigma_1^{m_n}\},$$

where $y(x)$ is defined in (5.15). Then $\Lambda_{n,g,\sigma_1,\sigma_2}$ can be further decomposed into $\{\Lambda_{n,g,\sigma_1,\sigma_2,h}\}_{1 \leq h \leq H}$ by restricting $x \in I_h$. By (5.15), $y = y(x)$ is monotone given $|x|$ sufficiently small. Hence, given $L = L(P, \epsilon, K)$ large enough, by (5.17) and (5.18) one has

$$(5.19) \quad Y(I_h) \cap Y(I_{h'}) = \emptyset \quad \text{if } |h - h'| \geq L,$$

which implies the following ‘almost orthogonality’ principle,

Claim 2. *If there is a constant A such that*

$$(5.20) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2,h}\| \leq A, \quad \text{for all } 1 \leq h \leq H,$$

then

$$(5.21) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2}\| \leq LA.$$

The proof of Claim 2 is almost identical to that of Claim 1, and it is omitted. To prove (5.20) for a desired bound A , we also need the following lemma, which is similar to Lemma 5.5 and whose proof can be found at the end of this section.

Lemma 5.6. *There is a $K = K(P, \epsilon, n)$, such that for $1 \leq h \leq H$, one has*

$$\text{Conv}(\rho_n^{-1}(U_{n,g,h}^*)) \subset \rho_n^{-1}(U_{n,g}^{**}).$$

With all the preparation above, we now complete the proof of Proposition 5.4. Invoking Lemma 1.7, Theorem 4.11, Corollary 4.13 and Theorem 2.1 yields

$$(5.22) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2,h}\| \lesssim \begin{cases} |\lambda \sigma_1^{p_n+q_n m_n} \sigma_2^{q_n}|^{-1/6}, \\ (\sigma_2 \sigma_1^{m_0+\dots+m_n})^{1/2}, \end{cases}$$

for every $1 \leq h \leq H$. By Claim 2, summing over h yields

$$(5.23) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2}\| \lesssim \begin{cases} |\lambda \sigma_1^{p_n+q_n m_n} \sigma_2^{q_n}|^{-1/6}, \\ (\sigma_2 \sigma_1^{m_0+\dots+m_n})^{1/2}. \end{cases}$$

Balancing these two estimates via σ_1 yields

$$(5.24) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2}\| \lesssim |\lambda|^{-\frac{1}{2(3+(p_n+q_n m_n)/(m_0+\dots+m_n))}} \cdot \sigma_2^{\frac{p_n+m_n q_n - q_n(m_0+\dots+m_n)}{3(m_0+\dots+m_n)+p_n+m_n q_n}}.$$

By (4.46) and (4.47), we conclude that

$$(5.25) \quad p_n + m_n q_n - q_n(m_0 + \dots + m_n) \geq 0.$$

In the case (5.25) is strictly positive, we can sum over σ_2 in (5.24), which yields

$$\sum_{\sigma_2} \|\Lambda_{n,g,\sigma_1,\sigma_2,h}\| \lesssim |\lambda|^{-\frac{1}{2(3+(p_n+q_n m_n)/(m_0+\dots+m_n))}} \leq |\lambda|^{-\frac{1}{2(3+d_E)}},$$

where the latter inequality will be proved in a moment.

Otherwise,

$$(5.26) \quad p_n + m_n q_n - q_n(m_0 + \cdots + m_n) = 0,$$

which implies

$$(5.27) \quad \begin{cases} p_0 = 0 \\ q_0 = s_0 = q_1 = q_2 = \cdots = q_{n-1} = q_n. \end{cases}$$

Then (5.23) becomes

$$(5.28) \quad \|\Lambda_{n,g,\sigma_1,\sigma_2}\| \lesssim \begin{cases} |\lambda \sigma_1^{q_0(m_0+\cdots+m_n)} \sigma_2^{q_0}|^{-1/6}, \\ (\sigma_2 \sigma_1^{m_0+\cdots+m_n})^{1/2}. \end{cases}$$

Summing over σ_2 yields

$$\|\Lambda_{n,g,\sigma_1}\| = \left\| \sum_{\sigma_2} \Lambda_{n,g,\sigma_1,\sigma_2} \right\| \leq \sum_{\sigma_2} \|\Lambda_{n,g,\sigma_1,\sigma_2}\| \lesssim |\lambda|^{-\frac{1}{2(3+q_0)}} = |\lambda|^{-\frac{1}{2(3+d_E)}}.$$

The last equality comes from $p_0 = 0$, $m_0 \geq 1$ and thus $\text{mult}(P) = d_E = q_0$.

It remains to verify $(p_n + q_n m_n)/(m_0 + \cdots + m_n) \leq d_E$. Indeed, (4.47) yields

$$\frac{p_n + q_n m_n}{m_0 + \cdots + m_n} \leq \frac{(p_0 + m_0 q_0) + s_0(m_1 + \cdots + m_n)}{m_0 + (m_1 + \cdots + m_n)} \leq \frac{p_0 + m_0 q_0}{m_0} = d_E,$$

since

$$s_0 \leq q_0 \leq \frac{p_0 + m_0 q_0}{m_0} \leq d_E.$$

5.4. Proofs of Lemmas 5.5 and 5.6

We only provide the proof of Lemma 5.6 for the other is similar and even simpler. First, notice that the upper and the lower boundaries of $U_{n,g,h}^*$ can be represented by two curves:

$$\begin{aligned} \bar{\gamma}_1(x) &= r_0 x^{m_0} + r_1 x^{m_0+m_1} + \cdots + r_{n-1} x^{m_0+\cdots+m_{n-1}} + r_{n,1} x^{m_0+\cdots+m_{n-1}+m_n}, \\ \bar{\gamma}_2(x) &= r_0 x^{m_0} + r_1 x^{m_0+m_1} + \cdots + r_{n-1} x^{m_0+\cdots+m_{n-1}} + r_{n,2} x^{m_0+\cdots+m_{n-1}+m'_n}, \end{aligned}$$

and the upper and the lower boundaries of $U_{n,g,h}^{**}$ by

$$\begin{aligned} \bar{\gamma}_1^*(x) &= r_0 x^{m_0} + r_1 x^{m_0+m_1} + \cdots + r_{n-1} x^{m_0+\cdots+m_{n-1}} + r_{n,1}^* x^{m_0+\cdots+m_{n-1}+m_n}, \\ \bar{\gamma}_2^*(x) &= r_0 x^{m_0} + r_1 x^{m_0+m_1} + \cdots + r_{n-1} x^{m_0+\cdots+m_{n-1}} + r_{n,2}^* x^{m_0+\cdots+m_{n-1}+m'_n}, \end{aligned}$$

where $r_{n,1} < r_{n,1}^*$, $r_{n,2} > r_{n,2}^*$ and $0 \leq m_n \leq m'_n$. If $m_n = m'_n$, we have in addition $r_{n,1} > r_{n,2}$. Notice that all the curves above have only finitely many terms of fractional powers of x , even in the case when $n = n_0 - 1$ for some n_0 defined as in (4.21). Assume without loss of generality $r_0 > 0$ and in particular all the curves

above are increasing functions of x . The assumption $m_0 \geq 1$ (see (5.1)) implies that the curves $\bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_1^*$ and $\bar{\gamma}_2^*$ are concave up. Thus we only need to take care of the upper boundary of $U_{n,g,h}^*$. Use $\sigma_{1,h}$ to denote the left end point of the interval I_h , for $1 \leq h \leq H$. By the definition of convexity, one needs to verify that if K is sufficiently large, then

$$(5.29) \quad t\bar{\gamma}_1(\sigma_{1,h}) + (1-t)\bar{\gamma}_1(\sigma_{1,h+1}) < \bar{\gamma}_1^*(t\sigma_{1,h} + (1-t)\sigma_{1,h+1})$$

for all $0 \leq t \leq 1$ and $\sigma_1 > 0$ sufficiently small.

Since both $\bar{\gamma}_1$ and $\bar{\gamma}_1^*$ are increasing functions, it suffices to show

$$(5.30) \quad \bar{\gamma}_1(\sigma_{1,h+1}) < \bar{\gamma}_1^*(\sigma_{1,h}).$$

By the mean value theorem, there is a $\sigma \in I_h$ such that

$$\bar{\gamma}_1(\sigma_{1,h+1}) - \bar{\gamma}_1(\sigma_{1,h}) = \bar{\gamma}'_1(\sigma)(\sigma_{1,h+1} - \sigma_{1,h}) = \frac{3}{2K}\sigma_2\sigma_1^{m_0+\dots+m_n-(m_0-1)}\bar{\gamma}'_1(\sigma).$$

Since $\sigma_1/2 \leq \sigma \leq 2\sigma_1$, then $|\bar{\gamma}'_1(\sigma)| \leq C\sigma_1^{m_0-1}$ for constant C . Thus

$$(5.31) \quad \bar{\gamma}_1(\sigma_{1,h+1}) - \bar{\gamma}_1(\sigma_{1,h}) \leq \frac{C}{K}\sigma_2\sigma_1^{m_0+\dots+m_n} < \frac{C}{K}\sigma_1^{m_0+\dots+m_n},$$

since $\sigma_2 < 1$. On the other hand,

$$(5.32) \quad \begin{aligned} \bar{\gamma}_1^*(\sigma_{1,h}) - \bar{\gamma}_1(\sigma_{1,h}) &= (r_{1,n}^* - r_{1,n})\sigma_{1,h}^{m_0+\dots+m_n} \\ &\geq (r_{1,n}^* - r_{1,n})(\sigma_1/2)^{m_0+\dots+m_n}. \end{aligned}$$

Thus by choosing

$$K > \frac{C}{r_{1,n}^* - r_{1,n}} \cdot 2^{m_0+\dots+m_n},$$

(5.31) and (5.32) imply (5.30) and thus (5.29).

References

- [1] CHRIST, M.: Bounds for multilinear sublevel sets via Szemerédi’s theorem. Preprint available at arXiv: 1107.2350, 2011.
- [2] CHRIST, M.: Multilinear oscillatory integrals via reduction of dimension. Preprint available at arXiv: 1107.2352, 2011.
- [3] CHRIST, M., LI, X., TAO, T. AND THIELE, C.: On multilinear oscillatory integrals, nonsingular and singular. *Duke Math. J.* **130** (2005), no. 2, 321–351,
- [4] CHRIST, M. AND OLIVEIRA E SILVA, D.: On trilinear oscillatory integrals. *Rev. Mat. Iberoam.* **30** (2014), no. 2, 667–684.
- [5] CUTKOSKY, S. D.: *Resolution of singularities*. Graduate Studies in Mathematics 63, American Mathematical Society, Providence, RI, 2004.
- [6] GREENBLATT, M.: A direct resolution of singularities for functions of two variables with applications to analysis. *J. Anal. Math.* **92** (2004), 233–257.

- [7] GREENBLATT, M.: Sharp L^2 estimates for one-dimensional oscillatory integral operators with C^∞ phase. *Amer. J. Math.* **127** (2005), no. 3, 659–695.
- [8] GREENBLATT, M.: Simply nondegenerate multilinear oscillatory integral operators with smooth phase. *Math. Res. Lett.* **15** (2008), no. 4, 653–660.
- [9] GREENLEAF, A., PRAMANIK, M. AND TANG, W.: Oscillatory integral operators with homogeneous polynomial phases in several variables. *J. Funct. Anal.* **244** (2007), no. 2, 444–487.
- [10] GRESSMAN, P. T. AND XIAO, L.: Maximal decay inequalities for trilinear oscillatory integrals of convolution type. *J. Funct. Anal.* **271** (2016), no. 12, 3695–3726.
- [11] HIRONAKA, H.: Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II. *Ann. of Math. (2)* **79** (1964), 109–203; 205–326.
- [12] HÖRMANDER, L.: Oscillatory integrals and multipliers on FL^p . *Ark. Mat.* **11** (1973), 1–11.
- [13] LI, X.: Bilinear Hilbert transforms along curves I: The monomial case. *Anal. PDE* **6** (2013), no. 1, 197–220.
- [14] PHONG, D. H. AND STEIN, E. M.: Models of degenerate Fourier integral operators and Radon transforms. *Ann. of Math. (2)* **140** (1994), no. 3, 703–722.
- [15] PHONG, D. H. AND STEIN, E. M.: Operator versions of the van der Corput lemma and Fourier integral operators. *Math. Res. Lett.* **1** (1994), no. 1, 27–33.
- [16] PHONG, D. H. AND STEIN, E. M.: The Newton polyhedron and oscillatory integral operators. *Acta Math.* **179** (1997), no. 1, 105–152.
- [17] PHONG, D. H., STEIN, E. M. AND STURM, J.: Multilinear level set operators, oscillatory integral operators, and Newton polyhedra. *Math. Ann.* **319** (2001), no. 3, 573–596.
- [18] RYCHKOV, V. S.: Sharp L^2 bounds for oscillatory integral operators with C^∞ phases. *Math. Z.* **236** (2001), no. 3, 461–489.
- [19] SEEGER, A.: Radon transforms and finite type conditions. *J. Amer. Math. Soc.* **11** (1998), no. 4, 869–897.
- [20] VARČENKO, A. N.: Newton polyhedra and estimates of oscillatory integrals. *Funkcional. Anal. i Priložen.* **10** (1976), no. 3, 13–38.
- [21] XIAO, L.: Endpoint estimates for one-dimensional oscillatory integral operators. Preprint, available at arXiv:1602.05663, 2016.

Received February 20, 2015; revised August 10, 2015.

LECHAO XIAO: Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA.

E-mail: xle@math.upenn.edu