Acta Math., 204 (2010), 151–271
DOI: 10.1007/s11511-010-0047-6


# Estimates for maximal functions associated with hypersurfaces in $\mathbb{R}^3$ and related problems of harmonic analysis

by

Isroil A. Ikromov
*Samarkand State University*
*Samarkand, Uzbekistan*

Michael Kempe
*C.A.-Universität Kiel*
*Kiel, Germany*

Detlef Müller
*C.A.-Universität Kiel*
*Kiel, Germany*

**Contents**



---

We acknowledge the support for this work by the Deutsche Forschungsgemeinschaft.

## 1. Introduction

Let $S$ be a smooth hypersurface in $\mathbb{R}^n$ and let $\varrho\in C_0^\infty(S)$ be a smooth non-negative function with compact support. Consider the associated averaging operators $A_t$, $t>0$, given by

$$A_tf(x):=\int_S f(x-ty)\varrho(y)\,d\sigma(y),$$

where $d\sigma$ denotes the surface measure on $S$. The associated maximal operator is given by

$$\mathcal{M}f(x):=\sup_{t>0}|A_tf(x)|,\quad x\in\mathbb{R}^n. \tag{1.1}$$

We remark that by testing $\mathcal{M}$ on the characteristic function of the unit ball in $\mathbb{R}^n$, it is easy to see that a necessary condition for $\mathcal{M}$ to be bounded on $L^p(\mathbb{R}^n)$ is that $p>n/(n-1)$, provided the transversality assumption 1.1 below is satisfied.

In 1976, E. M. Stein [38] proved that conversely, if $S$ is the Euclidean unit sphere in $\mathbb{R}^n$, $n\geqslant 3$, then the corresponding spherical maximal operator is bounded on $L^p(\mathbb{R}^n)$ for every $p>n/(n-1)$. The analogous result in dimension $n=2$ was later proved by J. Bourgain [4]. These results became the starting point for intensive studies of various classes of maximal operators associated with subvarieties. Stein's monograph [39] is an excellent reference to many of these developments. From these early works, the influence of geometric properties of $S$ on the validity of $L^p$-estimates of the maximal operator $\mathcal{M}$ became evident. For instance, A. Greenleaf [16] proved that $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^n)$ if $n\geqslant 3$ and $p>n/(n-1)$, provided $S$ has everywhere non-vanishing Gaussian curvature and in addition $S$ is starshaped with respect to the origin.

In contrast, the case where the Gaussian curvature vanishes at some points is still wide open, with the exception of the 2-dimensional case $n=2$, i.e., the case of finite-type curves in $\mathbb{R}^2$ studied by A. Iosevich in [21]. As a partial result in higher dimensions, C. D. Sogge and E. M. Stein showed in [37] that if the Gaussian curvature of $S$ does not vanish to infinite order at any point of $S$, then $\mathcal{M}$ is bounded on $L^p$ in a certain range $p>p(S)$. However, the exponent $p(S)$ given in that article is in general far from being optimal, and in dimensions $n\geqslant 3$, sharp results are known only for particular classes of hypersurfaces.

The perhaps best understood class in higher dimensions is the class of convex hypersurfaces of finite line type (see in particular the early work in this setting by M. Cowling and G. Mauceri in [8], [9], the work by A. Nagel, A. Seeger and S. Wainger in [30], and the articles [22], [23] by A. Iosevich and E. Sawyer and [24] by Iosevich, Sawyer and Seeger). In [30], sharp results were for instance obtained for convex hypersurfaces which are given as the graph of a mixed homogeneous convex function $\phi$. Further results were based on a result due to H. Schulz [34] (see also [42]), which states that, possibly after a rotation of coordinates, any smooth convex function $\phi$ of finite line type can be written in the form $\phi=Q+\phi_r$, where $Q$ is a convex mixed homogeneous polynomial that vanishes only at the origin, and $\phi_r$ is a remainder term consisting of terms of higher homogeneous degree than the polynomial $Q$. By means of this result, Iosevich and Sawyer proved in [23] sharp $L^p$-estimates for the maximal operator $\mathcal{M}$ for $p>2$. For further results in the case $p\leqslant 2$, see also [39].

As is well-known since the early work of E. M. Stein on the spherical maximal operator, the estimates of the maximal operator $\mathcal{M}$ on Lebesgue spaces are intimately connected with the decay rate of the Fourier transform

$$\widehat{\varrho\, d\sigma}(\xi)=\int_S e^{-i\xi\cdot x}\varrho(x)\, d\sigma(x), \quad \xi\in\mathbb{R}^n, \tag{1.2}$$

of the surface carried measure $\varrho\, d\sigma$, i.e., to estimates of oscillatory integrals. These in turn are closely related to geometric properties of the surface $S$, and have been considered by numerous authors ever since the early work by B. Riemann on this subject (see [39] for further information). Also the aforementioned results for convex hypersurfaces of finite line type are based on such estimates. Indeed, sharp estimates for the Fourier transform of surface carried measures on $S$ have been obtained by J. Bruna, A. Nagel and S. Wainger in [5], improving on previous results by B. Randol [33] and I. Svensson [40]. They introduced a family of non-isotropic balls on $S$, called "caps", by setting

$$B(x,\delta) := \{y \in S : \operatorname{dist}(y, x+T_xS) < \delta\}, \quad \delta > 0.$$

Here $T_xS$ denotes the tangent space to $S$ at $x \in S$. Suppose that $\xi$ is normal to $S$ at the point $x^0$. Then it was shown that

$$|\widehat{\varrho\, d\sigma}(\xi)| \leqslant C|B(x^0, |\xi|^{-1})|,$$

where $|B(x^0,\delta)|$ denotes the surface area of $B(x^0,\delta)$. These estimate became fundamental also in the subsequent work on associated maximal operators.

However, such estimates fail to be true for non-convex hypersurfaces, which we shall be dealing with in this article, too. More precisely, we shall consider general smooth hypersurfaces in $\mathbb{R}^3$.

Assume that $S \subset \mathbb{R}^3$ is such a hypersurface, and let $x^0 \in S$ be a fixed point in $S$. We can then find a Euclidean motion of $\mathbb{R}^3$, so that in the new coordinates given by this motion, we can assume that $x^0 = (0,0,1)$ and $T_{x^0} = \{(x_1,x_2,x_3) : x_3 = 0\}$. Then, in a neighborhood $U$ of the origin, the hypersurface $S$ is given by the graph

$$U \cap S = \{(x_1, x_2, 1+\phi(x_1,x_2)) : (x_1,x_2) \in \Omega\}$$

of a smooth function $1+\phi$ defined on an open neighborhood $\Omega$ of $0 \in \mathbb{R}^2$ and satisfying the conditions

$$\phi(0,0) = 0 \quad \text{and} \quad \nabla\phi(0,0) = 0. \tag{1.3}$$

With $\phi$ we can then associate the so-called height $h(\phi)$ in the sense of Varchenko [41] defined in terms of the Newton polyhedra of $\phi$ when represented in smooth coordinate systems near the origin (see §2 for details). An important property of this height is that it is invariant under local smooth changes of coordinates fixing the origin. We then define the *height* of $S$ at the point $x^0$ by $h(x^0, S) := h(\phi)$. This notion can easily be seen to be invariant under affine linear changes of coordinates in the ambient space $\mathbb{R}^3$ (cf. §12) because of the invariance property of $h(\phi)$ under local coordinate changes.

Now observe that unlike linear transformations, translations do not commute with dilations, which is why Euclidean motions are not admissible coordinate changes for the study of the maximal operators $\mathcal{M}$. We shall therefore study $\mathcal{M}$ under the following transversality assumption on $S$.

*Assumption* 1.1. For every $x\in S$, the affine tangent plane $x+T_xS$ to $S$ through $x$ does not pass through the origin in $\mathbb{R}^3$. Equivalently, $x\notin T_xS$ for every $x\in S$, so that $0\notin S$, and $x$ is transversal to $S$ for every point $x\in S$.

Notice that this assumption allows us to find a linear change of coordinates in $\mathbb{R}^3$ so that in the new coordinates $S$ can locally be represented as the graph of a function $\phi$ as before, and that the norm of $\mathcal{M}$ when acting on $L^p(\mathbb{R}^3)$ is invariant under such a linear change of coordinates.

If $\phi$ is flat, i.e., if all derivatives of $\phi$ vanish at the origin, and if $\varrho(x^0)>0$, then it is well known and easy to see that the maximal operator $\mathcal{M}$ is $L^p$-bounded if and only if $p=\infty$, so that this case is of no interest. Let us therefore assume in the sequel that $\phi$ is non-flat, i.e., of finite type. Correspondingly, we shall usually assume, often without further mentioning, that the hypersurface $S$ is of finite type in the sense that every tangent plane has finite order of contact.

We can now state the main result of this article.

THEOREM 1.2. *Assume that $S$ is a smooth, finite-type hypersurface in $\mathbb{R}^3$ satisfying Assumption* 1.1, *and let $x^0\in S$ be a fixed point. Then there exists a neighborhood $U\subset S$ of the point $x^0$ such that for every non-negative density $\varrho\in C_0^\infty(U)$ the associated maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ whenever $p>\max\{h(x^0,S),2\}$.*

Notice that even in the case where $S$ is convex this result is stronger than the previously known results, which always assumed that $S$ is of finite line type.

The following result shows the sharpness of Theorem 1.2.

THEOREM 1.3. *Assume that the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for some $p>1$, where $S$ satisfies Assumption* 1.1. *Then, for any point $x^0\in S$ with $\varrho(x^0)>0$, we have $h(x^0,S)\leqslant p$. Moreover, if $S$ is analytic at such a point $x^0$, then $h(x^0,S)<p$.*

From these results, global results can be deduced easily. For instance, if $S$ is a compact hypersurface, then we define the *height* $h(S)$ of $S$ by $h(S):=\sup_{x\in S} h(x,S)$. Then in fact

$$h(S):=\max_{x\in S} h(x,S)<\infty$$

(cf. Corollary 1.15), and from Theorems 1.2 and 1.3, we obtain the following result.

COROLLARY 1.4. *Assume that $S$ is a smooth, compact hypersurface of finite type in $\mathbb{R}^3$ satisfying Assumption* 1.1, *that $\varrho>0$ on $S$ and that $p>2$.*

*If $S$ is analytic, then the associated maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ if and only if $p>h(S)$. If $S$ is only assumed to be smooth, then for $p\neq h(S)$ we still have that the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ if and only if $p>h(S)$.*

Let $H$ be an affine hyperplane in $\mathbb{R}^3$. Following A. Iosevich and E. Sawyer [22], we consider the distance $d_H(x):=\operatorname{dist}(H,x)$ from $x\in S$ to $H$. In particular, if $x^0\in S$, then $d_{T,x^0}(x):=\operatorname{dist}(x^0+T_{x^0}S,x)$ will denote the distance from $x\in S$ to the affine tangent plane to $S$ at the point $x^0$. The following result has been proved in [22] in arbitrary dimensions $n\geqslant 2$ and without requiring Assumption 1.1.

THEOREM 1.5. (Iosevich–Sawyer) *If the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^n)$, where $p>1$, then*

$$\int_S d_H(x)^{-1/p}\varrho(x)\,d\sigma(x)<\infty \tag{1.4}$$

*for every affine hyperplane $H$ in $\mathbb{R}^n$ which does not pass through the origin.*

Moreover, it was conjectured in [22] that for $p>2$ the condition (1.4) is indeed necessary and sufficient for the boundedness of the maximal operator $\mathcal{M}$ on $L^p$, at least if for instance $S$ is compact and $\varrho>0$.

*Remark* 1.6. Notice that condition (1.4) is easily seen to be true for every affine hyperplane $H$ which is nowhere tangential to $S$, so that it is in fact a condition on affine tangent hyperplanes to $S$ only. Moreover, if Assumption 1.1 is satisfied, then there are no affine tangent hyperplanes which pass through the origin, so that in this case it is a condition on all affine tangent hyperplanes.

In §12, we shall prove the following result.

PROPOSITION 1.7. *Suppose that $S$ is a smooth hypersurface of finite type in $\mathbb{R}^3$, and let $x^0\in S$ be a fixed point. Then, for every $p<h(x^0,S)$, we have*

$$\int_{S\cap U} d_{T,x^0}(x)^{-1/p}\,d\sigma(x)=\infty \tag{1.5}$$

*for every neighborhood $U$ of $x^0$. Moreover, if $S$ is analytic near $x^0$, then* (1.5) *holds true also for $p=h(x^0,S)$.*

Notice that this result does not require Assumption 1.1. As an immediate consequence of Theorems 1.2 and 1.5, and Proposition 1.7 we obtain the following result.

COROLLARY 1.8. *Assume that $S\subset\mathbb{R}^3$ is of finite type and satisfies Assumption* 1.1, *and let $x^0\in S$ be a fixed point. Moreover, let $p>2$.*

*Then, if $S$ is analytic near $x^0$, there exists a neighborhood $U\subset S$ of the point $x^0$ such that for any $\varrho\in C_0^\infty(U)$ with $\varrho(x^0)>0$ the associated maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ if and only if condition* (1.4) *holds for every affine hyperplane $H$ in $\mathbb{R}^3$ which does not pass through the origin.*

*If $S$ is only assumed to be smooth near $x^0$, then the same conclusion holds true, with the possible exception of the exponent $p=h(x^0,S)$.*

This confirms the conjecture by Iosevich and Sawyer in our setting for analytic $S$, and for smooth, finite-type $S$ with the possible exception of the exponent $p{=}h(x^0,S)$. For the critical exponent $p{=}h(x^0,S)$, if $S$ is not analytic near $x^0$, examples show that unlike in the analytic case it may happen that $\mathcal{M}$ is bounded on $L^{h(x^0,S)}$ (see, e.g., [23]), and the conjecture remains open for this value of $p$. For further details, we refer to §12.

As mentioned before, the estimates of the maximal operator $\mathcal{M}$ on Lebesgue spaces are intimately connected with the decay rate of the Fourier transform (1.2) of the surface carried measure $\varrho\,d\sigma$. Estimates of such oscillatory integrals will naturally play a central role also in our proof Theorem 1.2. Indeed our proof of Theorem 1.2 will provide enough information so that it will also be easy to derive the following uniform estimate for the Fourier transform of surface carried measures on $S$ from it. Notice that our next few results do not require that $h(x^0,S)\geqslant 2$.

THEOREM 1.9. *Let $S$ be a smooth hypersurface of finite type in $\mathbb{R}^3$ and let $x^0$ be a fixed point in $S$. Then there exists a neighborhood $U\subset S$ of the point $x^0$ such that for every $\varrho\in C_0^\infty(U)$ the following estimate holds true*:

$$|\widehat{\varrho\,d\sigma}(\xi)|\leqslant C\|\varrho\|_{C^3(S)}(1+|\xi|)^{-1/h(x^0,S)}\log(2+|\xi|)\quad \textit{for every } \xi\in\mathbb{R}^3. \tag{1.6}$$

This estimate generalizes Karpushkin's estimates in [25] from the analytic to the smooth, finite-type setting, but just for linear perturbations. In various situations, estimate (1.6) holds true even without the factor $\log(2+|\xi|)$. A complete classification of all the cases when this is true has been obtained in the subsequent work [20].

For sublevel estimates related to Karpushkin's results, we refer to the work by Phong, Stein and Sturm [32], as well as to the recent work by Greenblatt [12]–[15]. We also like to mention the work by Carbery, Wainger and Wright [6], which contains other applications of the study of the Newton diagram to harmonic analysis.

It follows from the work of Greenleaf [16] (cf. also [39, Chapter VIII, 5.15 (b)]), that the uniform estimate (1.6) for the Fourier transform of the surface carried measure $\varrho\,d\sigma$ immediately implies the following $L^p(\mathbb{R}^3)$-$L^2(S)$ restriction theorem for the Fourier transform.

COROLLARY 1.10. *Let $S$ be a smooth hypersurface of finite type in $\mathbb{R}^3$ and let $x^0$ be a fixed point in $S$. Then there exists a neighborhood $U\subset S$ of the point $x^0$ such that for every non-negative $\varrho\in C_0^\infty(U)$ the estimate*

$$\left(\int_S|\hat f|^2\varrho\,d\sigma\right)^{1/2}\leqslant C_p\|f\|_{L^p(\mathbb{R}^3)},\quad f\in\mathcal{S}(\mathbb{R}^3), \tag{1.7}$$

*holds true for every $p\geqslant 1$ such that*

$$p'>2h(x^0,S)+2. \tag{1.8}$$

*In case* (1.6) *holds true without the factor* $\log(2+|\xi|)$, *the estimate* (1.7) *remains valid also for the endpoint* $p'=2h(x^0,S)+2$.

Restriction theorems for the Fourier transform have a long history by now, starting with the seminal work by E. M. Stein and P. Tomas for the case of the Euclidean sphere (see, e.g., [39]). Our results improve on work by A. Magyar [27] on analytic hypersurfaces in $\mathbb{R}^3$.

*Remarks* 1.11. (a) If $S$ is represented near $x^0$ as the graph of a function $\phi(x_1,x_2)$, where $\varrho(x^0)\neq 0$, then it can be easily seen by means of Knapp-type examples that the restriction estimate (1.7) can hold true only if

$$p' \geqslant 2d(\phi)+2$$

(see §12). Here, $d(\phi)$ denotes the Newton distance between the Newton diagram associated with $\phi$ and the origin.

If the coordinates $(x_1,x_2)$ are adapted to $\phi$ (see §2 for the notion of an adapted coordinate system), then we have $d(\phi)=h(\phi)=h(x^0,S)$, so that the result in Corollary 1.10 is sharp, except for the endpoint.

(b) In a subsequent work [20], we have been able to improve the result above by means of Littlewood–Paley theory and show that the restriction estimate (1.7) holds true also at the endpoint $p'=2h(x^0,S)+2$. Moreover, in the case where the coordinates are not adapted to $\phi$, it has turned out that the restriction theorem can be improved to a wider range of values for $p$.

The next result establishes a direct link between the decay rate of $\widehat{\varrho\, d\sigma}(\xi)$ and Iosevich–Sawyer's condition (1.4). It is an almost immediate consequence of some discussion on p. 539 in the work of Phong, Stein and Sturm [32]. In combination with Proposition 1.7, it shows in particular that the exponent $-1/h(x^0,S)$ in estimate (1.6) is sharp (for the case of analytic hypersurfaces this follows also from Varchenko's asymptotic expansions of oscillatory integrals in [41]).

THEOREM 1.12. *Let $S$ be a smooth hypersurface in $\mathbb{R}^n$, let $\varrho\in C_0^\infty(S)$ be a smooth, non-negative cut-off function, and assume that*

$$|\widehat{\varrho\, d\sigma}(\xi)| \leqslant C_\beta(1+|\xi|)^{-\beta} \quad \textit{for every } \xi\in\mathbb{R}^n, \tag{1.9}$$

*for some $\beta>0$. Then, for every $p>1$ such that $p>1/\beta$,*

$$\int_S d_H(x)^{-1/p}\varrho(x)\,d\sigma(x)<\infty \tag{1.10}$$

*for every affine hyperplane $H$ in $\mathbb{R}^n$.*

Observe that if $\varrho>0$ at some flat point $x^0$ of $S$, then the integral in (1.10) is infinite. Thus, the assumptions in Theorem 1.12 imply that necessarily $S$ is of finite type near such a point.

In combination with Proposition 1.7 this result easily implies (see §12) the following consequence.

COROLLARY 1.13. *Suppose that $S$ is a smooth hypersurface in $\mathbb{R}^3$, let $x^0\in S$ be a fixed point and assume that the estimate* (1.9) *holds true for some $\beta>0$. If $\varrho(x^0)>0$, and if $\varrho$ is supported in a sufficiently small neighborhood of $x^0$, then necessarily*

$$\beta\leqslant\frac{1}{h(x^0,S)}.$$

Indeed, more is true. Let us introduce the following quantities. In analogy with V. I. Arnold's notion of the "singularity index" [2], we define the *uniform oscillation index* $\beta_u(x^0,S)$ of the hypersurface $S\subset\mathbb{R}^n$ at the point $x^0\in S$ as follows:

Let $\mathfrak{B}_u(x^0,S)$ denote the set of all $\beta\geqslant 0$ for which there exists an open neighborhood $U_\beta$ of $x^0$ in $S$ such that estimate (1.9) holds true for every function $\varrho\in C_0^\infty(U_\beta)$. Then

$$\beta_u(x^0,S):=\sup\{\beta:\beta\in\mathfrak{B}_u(x^0,S)\}.$$

If we restrict our attention to the normal direction to $S$ at $x^0$ only, then we can define analogously the notion of *oscillation index* of the hypersurface $S$ at the point $x^0\in S$. More precisely, if $n(x^0)$ is a unit normal to $S$ at $x^0$, then we let $\mathfrak{B}(x^0,S)$ denote the set of all $\beta\geqslant 0$ for which there exists an open neighborhood $U_\beta$ of $x^0$ in $S$ such that the estimate (1.9) holds true along the line $\mathbb{R}n(x^0)$ for every function $\varrho\in C_0^\infty(U_\beta)$, i.e.,

$$|\widehat{\varrho\,d\sigma}(\lambda n(x^0))|\leqslant C_\beta(1+|\lambda|)^{-\beta}\quad\text{for every }\lambda\in\mathbb{R}. \tag{1.11}$$

Then

$$\beta(x^0,S):=\sup\{\beta:\beta\in\mathfrak{B}(x^0,S)\}.$$

If we regard $S$ locally as the graph of a function $\phi$, then we can introduce related notions $\beta_u(\phi)$ and $\beta(\phi)$ for $\phi$, regarded as the phase function of an oscillatory integral (cf. [19], and also §12).

We also define the *uniform contact index* $\gamma_u(x^0,S)$ of the hypersurface $S$ at the point $x^0\in S$ as follows: Let $\mathfrak{C}_u(x^0,S)$ denote the set of all $\gamma\geqslant 0$ for which there exists an open neighborhood $U_\gamma$ of $x^0$ in $S$ such that the estimate

$$\int_{U_\gamma} d_H(x)^{-\gamma}\,d\sigma(x)<\infty \tag{1.12}$$

holds true for every affine hyperplane $H$ in $\mathbb{R}^n$. Then we put

$$\gamma_u(x^0,S):=\sup\{\gamma:\gamma\in\mathfrak{C}_u(x^0,S)\}.$$

Similarly, we let $\mathfrak{C}(x^0,S)$ denote the set of all $\gamma\geqslant 0$ for which there exists an open neighborhood $U_\gamma$ of $x^0$ in $S$ such that

$$\int_{U_\gamma} d_{T,x^0}(x)^{-\gamma}\,d\sigma(x)<\infty, \tag{1.13}$$

and call

$$\gamma(x^0,S):=\sup\{\gamma:\gamma\in\mathfrak{C}(x^0,S)\}$$

the *contact index* $\gamma(x^0,S)$ of the hypersurface $S$ at the point $x^0\in S$. Then clearly

$$\beta_u(x^0,S)\leqslant\beta(x^0,S)\quad\text{and}\quad\gamma_u(x^0,S)\leqslant\gamma(x^0,S). \tag{1.14}$$

At least for hypersurfaces in $\mathbb{R}^3$, a lot more is true.

THEOREM 1.14. *Let $S$ be a smooth, finite-type hypersurface in $\mathbb{R}^3$, and let $x^0\in S$ be a fixed point. Then*

$$\beta_u(x^0,S)=\beta(x^0,S)=\gamma_u(x^0,S)=\gamma(x^0,S)=\frac{1}{h(x^0,S)}.$$

Note that for analytic hypersurfaces the estimate $\gamma(x^0,S)\geqslant 1/h(x^0,S)$ is also a consequence of Theorem 4 in the article [32] by Phong, Stein and Sturm. Partial results on the value of $\gamma(x^0,S)$ can be found in Greenblatt's articles [12] and [14]. These articles approach these estimates in a different way by means of sublevel estimates.

As an immediate consequence of Theorem 1.14, we obtain the following result.

COROLLARY 1.15. *Let $S$ be a smooth, finite-type hypersurface in $\mathbb{R}^3$, and let $x^0\in S$ be a fixed point. Then there exists a neighborhood $U\subset S$ of $x^0$ such that $h(x,S)\leqslant h(x^0,S)$ for every $x\in U$.*

This shows in particular that if $\Phi(x,s)=\phi(x_1,x_2)+s_1x_1+s_2x_2$ is a smooth deformation by linear terms of a smooth, finite-type function $\phi$ defined near the origin in $\mathbb{R}^2$ and satisfying (1.3), then the height of $\Phi(\cdot\,,s)$ at any critical point of the function $x\mapsto\Phi(x,s)$ is bounded by the height at $h(\Phi(\cdot\,,0))=h(\phi)$ for sufficiently small perturbation parameters $s_1$ and $s_2$. This proves a conjecture by V. I. Arnold [2] in the smooth setting at least for linear perturbations. For analytic functions $\phi$ of two variables, such a result has been proved for arbitrary analytic deformations by V. N. Karpushkin [25].

Let us recall at this point a result by A. Greenleaf. In [16] he proved that if

$$\widehat{\varrho\, d\sigma}(\xi) = O(|\xi|^{-\beta}), \quad \text{as } |\xi| \to \infty,$$

and if $\beta > \frac{1}{2}$, then the maximal operator is bounded on $L^p$ whenever $p > 1 + 1/2\beta$. The case $\beta \leqslant \frac{1}{2}$ remained open.

E. M. Stein, for $\beta = \frac{1}{2}$, and later A. Iosevich and E. Sawyer [23], for the full range $\beta \leqslant \frac{1}{2}$, conjectured that *if $S$ is a smooth, compact hypersurface in $\mathbb{R}^n$ such that*

$$|\widehat{\varrho\, d\sigma}(\xi)| = O(|\xi|^{-\beta}) \quad \textit{for some } 0 < \beta \leqslant \tfrac{1}{2},$$

*then the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^n)$ for every $p > 1/\beta$, at least if we assume that $\varrho > 0$.*

A partial confirmation of Stein's conjecture has been given by C. D. Sogge [36], who proved that if the surface has at least one non-vanishing principal curvature everywhere, then the maximal operator is $L^p$-bounded for every $p > 2$. Certainly, if the surface has at least one non-vanishing principal curvature then the estimate above holds for $\beta = \frac{1}{2}$.

Now, if $n = 3$ and $0 < \beta \leqslant \frac{1}{2}$, then $\beta_u(x^0, S) \geqslant \beta$ for every point $x^0 \in S$, so that our Theorem 1.14 implies that $1/\beta \geqslant h(x^0, S)$. Then, if $p > 1/\beta$, we have $p > \max\{2, h(x^0, S)\}$. Therefore, by means of a partition of unity argument, we obtain from Theorem 1.2 the following confirmation of the Stein–Iosevich–Sawyer conjecture in this case.

COROLLARY 1.16. *Let $S$ be a smooth compact hypersurface in $\mathbb{R}^3$ satisfying Assumption* 1.1, *and let $\varrho > 0$ be a smooth density on $S$. We assume that there is some $0 < \beta \leqslant \frac{1}{2}$ such that*

$$|\widehat{\varrho\, d\sigma}(\xi)| = O(|\xi|^{-\beta}).$$

*Then the associated maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p > 1/\beta$.*

We finally remark that the case $p \leqslant 2$ behaves quite differently, and examples show that neither condition (1.4) nor the notion of height will be suitable to determine the range of exponents $p$ for which the maximal operator $\mathcal{M}$ is $L^p$-bounded (see, e.g., [24]). The study of this range for $p \leqslant 2$ is work in progress.

### 1.1. Outline of the proof of Theorem 1.2 and organization of the article

The proof of our main result, Theorem 1.2, will strongly make use of the results in [19] on the existence of the so-called "adapted" coordinate systems for a smooth, finite-type function $\phi$ defined near the origin in $\mathbb{R}^2$, which will be briefly reviewed in §2.4 and §2.5. These results generalize the corresponding classical results for analytic $\phi$, due to A. N. Varchenko [41], by means of a simplified approach inspired by the seminal work of D. H. Phong and E. M. Stein [31]. According to these results, possibly after switching the coordinates $x_1$ and $x_2$, one can always find a change of coordinates of the form

$$y_1 := x_1, \quad y_2 := x_2 - \psi(x_1)$$

which leads to adapted coordinates $y$. The function $\psi$ can be constructed from the Pusieux series expansion of roots of $\phi$ (at least if $\phi$ is analytic) as the so-called principal root jet (cf. [19], and also the earlier work by D. H. Phong, E. M. Stein and J. A. Sturm [32], which contains a construction of adapted coordinates in the analytic setting by means of Puiseux series expansions of roots, too). The relations between the Newton diagram and the Pusieux series expansions of roots will be reviewed in §3. Somewhat simplifying, $\psi$ agrees with a real-valued leading part of the (complex) root of $\phi$ near which the function $\phi$ is "small to highest order" in an averaged sense. One would preferably like to work in these adapted coordinates $y$, since the height of $\phi$ when expressed in the adapted coordinates can be read off directly from the Newton polyhedron of $\phi$ as the so-called "Newton distance". However, this change of coordinates leads to substantial problems, since it is in general non-linear.

Now, in domains away from the curve $x_2 = \psi(x_1)$, it turns out that one can find some $k$ with $2 \leqslant k \leqslant h(\phi)$ such that $\partial_2^k \phi \neq 0$. This suggests that one may apply the results on maximal functions on curves in [21]. Indeed this is possible, but we need estimates for such maximal operators along curves which are stable under small perturbations of the given curve. Such results, which will be based on the local smoothing estimates by G. Mockenhaupt, A. Seeger and C. Sogge in [29], and related estimates for maximal operators along surfaces, are derived in §4. The necessary control on partial derivatives $\partial_2^k \phi$ will be obtained from the study of mixed homogeneous polynomials in §2.3. Indeed, in a similar way as the Schulz polynomial [34] is used in the convex case to approximate the given function $\phi$, we shall approximate the function $\phi$ in domains close to a given root of $\phi$ by a suitable mixed homogeneous polynomial, following here some ideas of Phong and Stein [31].

The case where our original coordinates $x$ are adapted or where the height $h(\phi)$ is strictly less than 2 is the simplest one, since we can here avoid non-linear changes of coordinates. This case is dealt with in §7.

We then concentrate in §§8–10 on the situation where $h(\phi) \geqslant 2$ and where the coordinates are not adapted.

The contributions to the maximal operator $\mathcal{M}$ by the complement of a narrow domain containing the curve $x_2 = \psi(x_1)$ are estimated in §8.1 by essentially the same tools as for the case of adapted coordinates.

The narrow domain containing the principal root jet $x_2=\psi(x_1)$ remains to be considered. For this domain, it is in general no longer possible to reduce its contribution to the maximal operator $\mathcal{M}$ to maximal operators along curves, and we have to apply 2-dimensional oscillatory integral techniques, too. Indeed, we shall need estimates for certain classes of oscillatory integrals with small parameters, which will be provided in §5 and §6. These results will be applied in §9 and §10 in order to complete the proof of Theorem 1.2.

Our approach requires some rather delicate domain decompositions, which will be described in §8.2 and §9.1. The decompositions from §8.2 have been considered before in [31], and we recommend that the reader consults that article for additional motivation. They are based on classical relations between the edges of the Newton diagram and the roots of $\phi$, which are reviewed in §3.

The decompositions from §9.1 share some features of the ones from §8.2, but are even more complex. They are based on a stopping time argument motivated by the behavior of the oscillatory integrals that arise, after rescaling, from dyadic pieces of the oscillatory integral associated with $\phi$. The crucial control quantity for the stopping time will be the size of $\partial_2\phi$.

We remark that our proof does not make use of any damping techniques, which had been crucial to many earlier approaches.

The proof of Theorem 1.9, which will be given in §11, can easily be obtained from the results established in the course of the proof of Theorem 1.2, except for the case $h(x^0,S)<2$, which, however, has been studied by J. J. Duistermaat [11] in a complete way. The main difference is that we have to replace the estimates for maximal operators in §4 by van der Corput type estimates due to J. E. Björk and G. I. Arkhipov.

In the last section, §12, we shall give proofs of all the other results stated above as well as refinements of some of them.

## 2. Newton diagrams and adapted coordinates

### 2.1. Basic definitions

We first recall some basic notions from [19], which essentially go back to the paper [41] by A. N. Varchenko. Let $\phi$ be a smooth real-valued function defined on an open neighborhood $\Omega$ of the origin in $\mathbb{R}^2$ with $\phi(0,0)=0$ and $\nabla\phi(0,0)=0$, and consider the associated Taylor series

$$\phi(x_1,x_2)\sim\sum_{j,k=0}^{\infty} c_{jk}x_1^j x_2^k$$

of $\phi$ centered at the origin. The set

$$\mathcal{T}(\phi) := \left\{ (j,k) \in \mathbb{N}^2 : c_{jk} = \frac{1}{j!k!} \partial_{x_1}^j \partial_{x_2}^k \phi(0,0) \neq 0 \right\}$$

will be called the *Taylor support* of $\phi$ at $(0,0)$. We shall always assume that

$$\mathcal{T}(\phi) \neq \varnothing,$$

i.e., that the function $\phi$ is of finite type at the origin. If $\phi$ is real-analytic, so that the Taylor series converges to $\phi$ near the origin, this just means that $\phi \neq 0$. The *Newton polyhedron* $\mathcal{N}(\phi)$ of $\phi$ at the origin is defined to be the convex hull of the union of all the quadrants $(j,k)+\mathbb{R}_+^2$ in $\mathbb{R}^2$, with $(j,k) \in \mathcal{T}(\phi)$. The associated *Newton diagram* $\mathcal{N}_d(\phi)$ is the union of all compact faces of the Newton polyhedron; here, by a *face,* we shall mean an edge or a vertex.

We shall use coordinates $(t_1, t_2)$ for points in the plane containing the Newton polyhedron, in order to distinguish this plane from the $(x_1, x_2)$-plane.

The *Newton distance* (or shorter *distance*) $d=d(\phi)$ between the Newton polyhedron and the origin is given by the coordinate $d$ of the point $(d,d)$ at which the bisectrix $t_1=t_2$ intersects the boundary of the Newton polyhedron.

The *principal face* $\pi(\phi)$ of the Newton polyhedron of $\phi$ is the face of minimal dimension containing the point $(d,d)$. Deviating from the notation in [41], we shall call the series

$$\phi_{\mathrm{pr}}(x_1, x_2) := \sum_{(j,k) \in \pi(\phi)} c_{jk} x_1^j x_2^k \tag{2.1}$$

the *principal part* of $\phi$. In the case where $\pi(\phi)$ is compact, $\phi_{\mathrm{pr}}$ is a polynomial; otherwise, we shall consider $\phi_{\mathrm{pr}}$ as a formal power series.

Note that the distance between the Newton polyhedron and the origin depends on the chosen local coordinate system in which $\phi$ is expressed. By a *local analytic* (respectively, *smooth*) *coordinate system at the origin* we shall mean an analytic (respectively, smooth) coordinate system defined near the origin which preserves 0. If we work in the category of smooth functions $\phi$, we shall always consider smooth coordinate systems, and if $\phi$ is analytic, then one usually restricts oneself to analytic coordinate systems (even though this will not really be necessary for the questions we are going to study, as we shall see). The *height* of the analytic (respectively, smooth) function $\phi$ is defined by

$$h(\phi) := \sup d_x,$$

where the supremum is taken over all local analytic (respectively, smooth) coordinate systems $x$ at the origin, and where $d_x$ is the distance between the Newton polyhedron and the origin in the coordinates $x$.

A given coordinate system $x$ is said to be *adapted* to $\phi$ if $h(\phi)=d_x$.

### 2.2. The $\varkappa$-principal part $\phi_\varkappa$ of $\phi$ associated with a supporting line of the Newton polyhedron as a mixed homogeneous polynomial

Let $\varkappa=(\varkappa_1,\varkappa_2)$ with $\varkappa_1,\varkappa_2\geqslant 0$ and $|\varkappa|:=\varkappa_1+\varkappa_2>0$ be a given weight, with associated one-parameter family of dilations $\delta_r(x_1,x_2):=(r^{\varkappa_1}x_1,r^{\varkappa_2}x_2)$, $r>0$. A function $\phi$ on $\mathbb{R}^2$ is said to be *$\varkappa$-homogeneous of degree* $a$, if $\phi(\delta_r x)=r^a\phi(x)$ for every $r>0$, $x\in\mathbb{R}^2$. Such functions will also be called *mixed homogeneous.* The exponent $a$ will be called the *$\varkappa$-degree* of $\phi$. For instance, the monomial $x_1^j x_2^k$ has $\varkappa$-degree $\varkappa_1 j+\varkappa_2 k$.

If $\phi$ is an arbitrary smooth function near the origin, consider its Taylor series $\sum_{j,k=0}^\infty c_{jk}x_1^j x_2^k$ around the origin. We choose $a$ so that the line

$$L_\varkappa:=\{(t_1,t_2)\in\mathbb{R}^2:\varkappa_1 t_1+\varkappa_2 t_2=a\}$$

is the supporting line to the Newton polyhedron $\mathcal{N}(\phi)$ of $\phi$. If we assume that $\varkappa_1>0$ and $\varkappa_2>0$, then

$$\phi_\varkappa(x_1,x_2):=\sum_{(j,k)\in L_\varkappa} c_{jk}x_1^j x_2^k$$

is a non-trivial polynomial which is $\varkappa$-homogeneous of degree $a$; it will be called the *$\varkappa$-principal part* of $\phi$. By definition, we then have

$$\phi(x_1,x_2)=\phi_\varkappa(x_1,x_2)+\text{terms of higher }\varkappa\text{-degree.} \tag{2.2}$$

More precisely, we mean by this that every point $(j,k)$ in the Taylor support of the remainder term $\phi_r:=\phi-\phi_\varkappa$ lies on a line $\varkappa_1 t_1+\varkappa_2 t_2=d$ with $d>a$ parallel to, but above the line $L_\varkappa$, i.e., we have $\varkappa_1 j+\varkappa_2 k>a$. Moreover, clearly

$$\mathcal{N}_d(\phi_\varkappa)\subset\mathcal{N}_d(\phi).$$

In the sequel, we shall often encounter polynomial functions $P$ satisfying

$$\nabla P(0,0)=0,$$

which are $\varkappa$-homogeneous of degree 1. The quantity

$$d_h(P):=\frac{1}{\varkappa_1+\varkappa_2} \tag{2.3}$$

will then be called the *homogeneous distance* of the mixed homogeneous polynomial $P$. We recall that $(d_h(P),d_h(P))$ is just the point of intersection of the bisectrix with the line $\varkappa_1 t_1+\varkappa_2 t_2=1$ on which the Newton diagram $\mathcal{N}_d(P)$ lies, and that

$$d_h(P)\leqslant d(P) \tag{2.4}$$

(cf. [19]). By

$$m(P) := \mathrm{ord}_{S^1} P \tag{2.5}$$

we shall denote the maximal order of vanishing of $P$ along the unit circle $S^1$ centered at the origin. We note that, by [19, Corollary 3.4], we then have

$$h(P) = \max\{m(P), d_h(P)\}. \tag{2.6}$$

We also recall the following result (cf. Proposition 2.2 and Corollary 2.3 in [19]). If $m_1, ..., m_n$ are positive integers, then we denote their greatest common divisor by $(m_1, ..., m_n)$.

PROPOSITION 2.1. *Let $P$ be a $\varkappa$-homogeneous polynomial of degree* 1, *and assume that $P$ is not of the form $P(x_1, x_2) = cx_1^{\nu_1} x_2^{\nu_2}$. Then $\varkappa_1$ and $\varkappa_2$ are uniquely determined by $P$, and $\varkappa_1, \varkappa_2 \in \mathbb{Q}$.*

*Let us assume that $\varkappa_1 \leqslant \varkappa_2$, and write*

$$\varkappa_1 = \frac{q}{m} \quad \textit{and} \quad \varkappa_2 = \frac{p}{m}, \quad (p, q, m) = 1,$$

*so that in particular $p \geqslant q$. Then $(p, q) = 1$ and there exist non-negative integers $\alpha_1$ and $\alpha_2$ and a $(1, 1)$-homogeneous polynomial $Q$ such that the polynomial $P$ can be written as*

$$P(x_1, x_2) = x_1^{\alpha_1} x_2^{\alpha_2} Q(x_1^p, x_2^q). \tag{2.7}$$

*More precisely, $P$ can be written in the form*

$$P(x_1, x_2) = cx_1^{\nu_1} x_2^{\nu_2} \prod_{l=1}^{M} (x_2^q - \lambda_l x_1^p)^{n_l}, \tag{2.8}$$

*with $M \geqslant 1$, distinct $\lambda_l \in \mathbb{C} \setminus \{0\}$ and multiplicities $n_l \in \mathbb{N} \setminus \{0\}$, and with $\nu_1, \nu_2 \in \mathbb{N}$ (possibly different from $\alpha_1$ and $\alpha_2$ in* (2.7)).

*Let $n := \sum_{l=1}^{M} n_l$. The distance $d(P)$ of $P$ can then be read off from* (2.8) *as follows: If the principal face of $\mathcal{N}(P)$ is compact, then it lies on the line $\varkappa_1 t_1 + \varkappa_2 t_2 = 1$, and the distance is given by*

$$d(P) = \frac{1}{\varkappa_1 + \varkappa_2} = \frac{\nu_1 q + \nu_2 p + pqn}{q + p}. \tag{2.9}$$

*Otherwise, we have $d(P) = \max\{\nu_1, \nu_2\}$. In particular, in any case we have*

$$d(P) = \max\{\nu_1, \nu_2, d_h(P)\}.$$

The proposition shows that every zero (or "root") $(x_1, x_2)$ of $P$ which does not lie on a coordinate axis is of the form $x_2 = \lambda_l^{1/q} x_1^{p/q}$. We also note that

$$m(P) = \max\{\nu_1, \nu_2, \max\{n_l : \lambda_l \in \mathbb{R} \text{ and } l = 1, ..., M\}\}. \tag{2.10}$$

In view of the homogeneity of $P$, we shall often restrict our considerations to roots lying on the unit circle. For the next result, cf. [19, Corollaries 2.3 and 3.4].

COROLLARY 2.2. *Let $P$ be a $(\varkappa_1, \varkappa_2)$-homogeneous polynomial of degree* 1 *as in Proposition* 2.1, *and consider the representation* (2.8) *of $P$. We put again*

$$n := \sum_{l=1}^{M} n_l.$$

(a) *If $\varkappa_2/\varkappa_1 \notin \mathbb{N}$, that is, if $q \geqslant 2$, then $n < d_h(P)$. In particular, every real root $x_2 = \lambda_l^{1/q} x_1^{p/q}$ of $P$ has multiplicity $n_l < d_h(P)$.*

(b) *If $\varkappa_2/\varkappa_1 \in \mathbb{N}$, that is, if $q=1$, then there exists at most one real root of $P$ on the unit circle $S^1$ of multiplicity greater than $d_h(P)$. More precisely, if we put $n_0 := \nu_1$ and $n_{M+1} := \nu_2$ and choose $l_0 \in \{0, ..., M+1\}$ so that*

$$n_{l_0} = \max\{n_l : l = 0, ..., M+1\} > d_h(P),$$

*then $n_l < d_h(P)$ for every $l \neq l_0$.*

(c) *If $1 < \varkappa_2/\varkappa_1 \in \mathbb{N}$, then there exists at most one real root $(x_1^0, x_2^0)$ of $P$ on the unit circle $S^1$ with $x_1^0 \neq 0$, which has multiplicity greater than or equal to $d_h(P)$.*

In particular, we see that the multiplicity of every real root of $P$ not lying on a coordinate axis is bounded by the distance $d(P)$, unless $q=1$, in which case there can at most be one real root $x_2 = \lambda_{l_0} x_1^p$ with multiplicity exceeding $d(P)$. If such a root exists, we shall call it the *principal root* of $P$.

### 2.3. On the multiplicity of roots of derivatives of a mixed homogeneous polynomial function

Assume that $P$ is a $\varkappa$-homogeneous polynomial of degree 1 such that $\nabla P(0) = 0$, and that $0 < \varkappa_1 \leqslant \varkappa_2$. We shall show that the multiplicities of roots of $\partial_2 P$ (respectively, $\partial_2^2 P$) can be controlled in many cases in a suitable way by the homogeneous distance of $P$.

These results will for instance allow us later to reduce the estimation of the maximal operator in Theorem 1.2 for large parts of the surface measure of $S$ to the estimation of maximal operators along curves in a plane, which will be provided in Proposition 4.5 (respectively, Corollary 4.6).

PROPOSITION 2.3. *Let $P$ be a $\varkappa$-homogeneous polynomial of degree* 1 *such that* $\nabla P(0)=0$.

(a) *Suppose that $\partial_2^2 P$ does not vanish identically, and that $2<\varkappa_2/\varkappa_1\in\mathbb{N}$, so that $p:=\varkappa_2/\varkappa_1\geqslant 3$ and $q=1$ in* (2.8). *If $x^0\in S^1$, then denote by $m_2(x^0)$ the order of vanishing of $\partial_2^2 P$ along the circle $S^1$ in the point $x^0$. By $\mathcal{R}_2$ we shall denote the set of all roots $x^0=(x_1^0,x_2^0)$ of $\partial_2^2 P$ on the unit circle such that $x_1^0\neq 0$. We assume that $\mathcal{R}_2\neq\varnothing$, and let $x^m\in\mathcal{R}_2$ be a root of maximal multiplicity $m_2(x^m)\geqslant 1$ among all roots in $\mathcal{R}_2$. Then, for any other root $x^0\neq x^m$ in $\mathcal{R}_2$, we have $m_2(x^0)\leqslant d_h(P)-2$.*

*In particular, for every point $x=(x_1,x_2)\in S^1$ such that $x_1\neq 0$ and $x\neq x^m$ there exists some $j$, with $2\leqslant j\leqslant d_h(P)$, such that $\partial_2^j P(x)\neq 0$.*

(b) *Assume that $P$ vanishes along $S^1$ of exact order $\nu_2=d(P)$ in the point* $(1,0)$ *on the $x_1$-axis, that $\varkappa_2/\varkappa_1>2$ and that $d(P)>2$. Then $m_2(x^0)\leqslant d_h(P)-2$ for every root $x^0\in\mathcal{R}_2$ satisfying $x_2^0\neq 0$.*

*In particular, for every point $x\in S^1$ which does not lie on a coordinate axis, there exists some $j$ with $2\leqslant j\leqslant d_h(P)$ such that $\partial_2^j P(x)\neq 0$.*

(c) *Assume that $\varkappa_2/\varkappa_1\notin\mathbb{N}$, and that $\partial_2 P$ does not vanish identically. If $x^0\in S^1$, then denote by $m_1(x^0)$ the order of vanishing of $\partial_2 P$ along $S^1$ in the point $x^0$. Then $m_1(x_0)<d_h(P)-1$ for every root $x^0$ of $\partial_2 P$ with $x_1^0\neq 0\neq x_2^0$.*

*In particular, for every point $x\in S^1$ which does not lie on a coordinate axis, there exists some $j$ with $1\leqslant j<d_h(P)$ such that $\partial_2^j P(x)\neq 0$.*

*Remarks.* (i) In case (a), if $m(P)>d(P)$, so that $P$ has a (unique) principal root $x^{\mathrm{pr}}\in S^1$, then $x^m=x^{\mathrm{pr}}$.

(ii) In analogy with (c), one can prove that if $\varkappa_2/\varkappa_1\notin\mathbb{N}$, then $m_2(x_0)\leqslant d_h(P)-2$ for every root with $x_1^0\neq 0\neq x_2^0$, unless the polynomial $P$ is of the form

$$P(x_1,x_2)=c(x_2^2-\lambda_1 x_1^5)(x_2^2-\lambda_2 x_1^5),$$

with $\lambda_1+\lambda_2\in\mathbb{R}\setminus\{0\}$ and $\lambda_1\lambda_2\in\mathbb{R}$. Our estimate in (c) for $m_1(x^0)$ will allow us to avoid a separate discussion of these exceptional polynomials.

*Proof.* We first prove (a) and remark (i). In order to prepare also the proof of part (b), let us initially only assume that $\varkappa_2/\varkappa_1>2$. By our assumptions, $\partial_2^2 P$ is a $\sigma$-homogeneous polynomial of degree 1 with respect to the weight $\sigma=(\sigma_1,\sigma_2)$, with

$$\sigma_1:=\frac{\varkappa_1}{1-2\varkappa_2}\quad\text{and}\quad\sigma_2:=\frac{\varkappa_2}{1-2\varkappa_2}.$$

According to the Proposition 2.1, we can write the polynomial $\partial_2^2 P$ in the form

$$\partial_2^2 P(x_1,x_2)=x_1^{\nu_1}x_2^{\nu_2}Q_2(x_1^p,x_2^q),$$

where $Q_2$ is a homogeneous polynomial of degree $n_2$, and

$$a := \frac{p}{q} = \frac{\varkappa_2}{\varkappa_1} = \frac{\sigma_2}{\sigma_1} > 2. \tag{2.11}$$

We shall also assume that no power of $x_2^q$ can be factored from $Q_2(x_1^p, x_2^q)$, so that we have

$$\sigma_1 = \frac{q}{\nu_1 q+\nu_2 p+n_2 pq} \quad \text{and} \quad \sigma_2 = \frac{p}{\nu_1 q+\nu_2 p+n_2 pq}. \tag{2.12}$$

Let us put $N:=\nu_2+n_2$.

To prove (a), assume now that $q=1$. If $x^0$ is a root different from $x^m$ in $\mathcal{R}_2$, then $1\leqslant m_2(x^0)\leqslant m_2(x^m)$, and so we have

$$2m_2(x^0) \leqslant m_2(x^0)+m_2(x^m) \leqslant N,$$

and hence in particular $N\geqslant 2$. Assume we had $m_2(x^0)>d_h(P)-2$. Then

$$\frac{1+2\sigma_2}{\sigma_1+\sigma_2} = d_h(P) < \frac{N}{2}+2, \tag{2.13}$$

and in view of (2.12), one computes that $N<2(2-\nu_1)/(p-1)$. Since $N\geqslant 2$, this implies that $p<3-\nu_1$, contradicting our assumption (2.11), which implies here that $p\geqslant 3$.

To prove remark (i), assume that $m(P)>d(P)$, so that $P$ has a (unique) principal root $x^{\mathrm{pr}}\in S^1$ of multiplicity $m(x^{\mathrm{pr}})=m(P)$.

If $m(P)\geqslant 3$, then $x^{\mathrm{pr}}\in\mathcal{R}_2$, with multiplicity

$$m_2(x^{\mathrm{pr}}) = m(P)-2 > d(P)-2 \geqslant d_h(P)-2,$$

so that, by (a), we must have $x^{\mathrm{pr}}=x^m$, and the conclusion in (i) is obvious.

If $m(P)\leqslant 2$, then $d(P)<2$, and hence

$$\frac{1+2\sigma_2}{\sigma_1+\sigma_2} < 2,$$

which implies that $\nu_1+pN\leqslant 1$. Consequently, we have $N=0$ and $\nu_1\leqslant 1$, so that $P$ would be a polynomial of degree at most one, and hence $\partial_2^2 P$ would vanish identically. This shows that this case actually cannot arise.

We next prove (b). So, assume that $\nu_2=d(P)>2$. Then $\partial_2^2 P$ vanishes exactly of order $d(P)-2\geqslant 1$ in the point $e:=(1,0)$, i.e., $m_2(e)=d(P)-2$. Let $x^0$ be any root of $\partial_2^2 P$ with $x_1^0\neq 0\neq x_2^0$. We want to show that $m_2(x^0)\leqslant d_h(P)-2$. Assume to the contrary that $m_2(x^0)>d_h(P)-2$.

If $m_2(x^0)<m_2(e)$, then

$$2m_2(x^0) < m_2(e)+m_2(x^0) \leqslant N,$$

and we obtain $d_h(P)<\frac{1}{2}N+2$. And, if $m_2(x^0)\geqslant m_2(e)$, then

$$2m_2(e)\leqslant m_2(e)+m_2(x^0)\leqslant N,$$

and hence $d(P)\leqslant\frac{1}{2}N+2$. So, in both cases, we have $d_h(P)\leqslant\frac{1}{2}N+2$. As in the proof of (a), this leads to a contradiction: Indeed, (2.12) and (2.13) imply that

$$\frac{\nu_1 q+(\nu_2+2)p+n_2 pq}{p+q}\leqslant\frac{N}{2}+2,$$

and hence

$$\frac{\nu_1+(\nu_2+2)a+n_2 p}{a+1}\leqslant\frac{\nu_2+n_2+4}{2}.$$

Since $a>2$, $\nu_2\geqslant 3$ and $n_2\geqslant 1$, this implies that

$$p<\frac{-\frac{1}{2}\nu_2+\frac{3}{2}n_2+2}{n_2}\leqslant\frac{3}{2}+\frac{1}{2n_2}\leqslant 2,$$

and hence $p<2$, which is not possible.

Let us finally prove (c). So, assume that $x^0$ is a root of $\partial_2 P$ which does not lie on a coordinate axis. Arguing similarly as in (a), we see that $\partial_2 P(x)$ is a $\sigma$-homogeneous polynomial of degree 1 with respect to the weight $\sigma=(\sigma_1,\sigma_2)$, with

$$\sigma_1:=\frac{\varkappa_1}{1-\varkappa_2}\quad\text{and}\quad\sigma_2:=\frac{\varkappa_2}{1-\varkappa_2},$$

and $\partial_2 P(x)$ can be written in the form $\partial_2 P(x_1,x_2)=x_1^{\nu_1}x_2^{\nu_2}Q_1(x_1^p,x_2^q)$, where $p$ and $q$ are coprime, $Q_1$ is a homogeneous polynomial of degree $n_1$ and

$$\frac{p}{q}=\frac{\varkappa_2}{\varkappa_1}=\frac{\sigma_2}{\sigma_1}>1.$$

We shall also assume that no power of $x_2^q$ can be factored from $Q_1(x_1^p,x_2^q)$, so that we have

$$\sigma_1=\frac{q}{\nu_1 q+\nu_2 p+pqn_1}\quad\text{and}\quad\sigma_2=\frac{p}{\nu_1 q+\nu_2 p+pqn_1}.$$

Since $\varkappa_2/\varkappa_1\notin\mathbb{N}$, we have $q\geqslant 2$. Moreover, our assumption $x_1^0\neq 0\neq x_2^0$ implies that

$$qm_1(x^0)\leqslant n_1.$$

Assume that $m_1(x^0)\geqslant d_h(P)-1$. Then

$$\frac{1+\sigma_2}{\sigma_1+\sigma_2}=\frac{1}{\varkappa_1+\varkappa_2}\leqslant 1+m_1(x^0)\leqslant 1+\frac{n_1}{q},$$

and hence

$$n_1 \leqslant \frac{q-(\nu_1 q+\nu_2 p)}{pq-1-p/q}.$$

Let us first consider the case $\nu_1+\nu_2\geqslant 1$. Since $\nu_1 q+\nu_2 p\geqslant(\nu_1+\nu_2)q\geqslant q$, we then obtain $n_1\leqslant 0$, so that $P=cx_1^{\nu_1}x_2^{\nu_2}$, which cannot vanish at $x^0$, so that we arrive at the contradiction $m_1(x^0)=0$.

Assume next that $\nu_1=\nu_2=0$. Then

$$1\leqslant n_1 < \frac{q}{pq-1-p/q}, \quad \text{that is,} \quad p<\frac{1+q}{q-1/q}.$$

But we have $q\geqslant 2$, so that

$$\frac{1+q}{q-1/q}\leqslant 2.$$

This contradicts our assumption that $p/q>1$. $\square$

### 2.4. Adaptedness of coordinates

Let us begin by remarking that if the principal face $\pi(\phi)$ of the Newton polyhedron $\mathcal{N}(\phi)$ of $\phi$ is a compact edge, then there is a unique weight $\varkappa=(\varkappa_1,\varkappa_2)$ so that $\varkappa_1,\varkappa_2>0$ and such that $\pi(\phi)$ lies on the line $\varkappa_1 t_1+\varkappa_2 t_2=1$. Without loss of generality, by flipping coordinates, if necessary, we may and shall assume in the sequel that $\varkappa_2\geqslant\varkappa_1$ (cf. [19]). Thus, $0<\varkappa_1\leqslant\varkappa_2$ and $\phi_{\mathrm{pr}}=\phi_\varkappa$ is a $\varkappa$-homogeneous polynomial of degree 1. We then recall from [19, Corollaries 4.3 and 5.2], the following facts:

The coordinates $x$ are adapted to $\phi$ if and only if the principal face $\pi(\phi)$ of the Newton polyhedron $\mathcal{N}(\phi)$ satisfies one of the following conditions:

(a) $\pi(\phi)$ is a compact edge, and either $\varkappa_2/\varkappa_1\notin\mathbb{N}$, or $\varkappa_2/\varkappa_1\in\mathbb{N}$ and $m(\phi_{\mathrm{pr}})\leqslant d(\phi)$;

(b) $\pi(\phi)$ consists of a vertex;

(c) $\pi(\phi)$ is unbounded.

Moreover, in case (a) we have $h(\phi)=h(\phi_{\mathrm{pr}})=d(\phi_{\mathrm{pr}})$. Notice also that if $\pi(\phi)$ is a compact edge and $\varkappa_2/\varkappa_1\notin\mathbb{N}$, then we even have that $m(\phi_{\mathrm{pr}})<d(\phi)$ (cf. [19, Corollary 2.3]). In the sequel, we shall often refer to these as the cases (a)–(c), without further mentioning.

LEMMA 2.4. *If the coordinates are adapted to $\phi$, then there is a weight $\varkappa=(\varkappa_1,\varkappa_2)$ such that, without loss of generality, $0<\varkappa_1\leqslant\varkappa_2<1$ and $h(\phi)=h(\phi_\varkappa)$.*

*Proof.* In case (a), this is immediate from the previous discussion. Observe also that $\varkappa_j<1$ for $j=1,2$, since $\phi(0)=0$ and $\nabla\phi(0)=0$.

Consider next case (b), where $\pi(\phi)$ consists of a vertex $(N,N)$. Then $h(\phi)=N\geqslant 1$. We can then choose a weight with $0<\varkappa_1\leqslant\varkappa_2<1$ (possibly after flipping coordinates) so that the line $\varkappa_1 t_1+\varkappa_2 t_2=1$ intersects the Newton polyhedron $\mathcal{N}(\phi)$ in the single point $(N,N)$, i.e., so that $\phi_\varkappa=cx_1^N x_2^N$, where $c$ is a non-trivial constant. Then again $h(\phi)=N=d(\phi_\varkappa)=h(\phi_\varkappa)$.

Case (c) can be treated similarly. Here, the principal face $\pi(\phi)$ may be assumed to be a horizontal half-line, with left endpoint $(\nu_1,N)$, where $\nu_1<N=h(\phi)$. Notice that $N\geqslant 2$, since for $N=1$ we had $\nu_1=0$, which is not possible given our assumption $\nabla\phi(0,0)=0$. We can then choose $\varkappa$, with $0<\varkappa_1<\varkappa_2$, so that the line $\varkappa_1 t_1+\varkappa_2 t_2=1$ is a supporting line to the Newton polyhedron of $\phi$ and that the point $(\nu_1,N)$ is the only point of $\mathcal{N}(\phi)$ on this line (we just have to choose $\varkappa_2/\varkappa_1$ sufficiently large!). Then necessarily $\varkappa_2<1$, and the $\varkappa$-principal part $\phi_\varkappa$ of $\phi$ is of the form $\phi_\varkappa(x)=cx_1^{\nu_1}x_2^N$, with $c\neq 0$. Since the coordinates are clearly also adapted to $\phi_\varkappa$, we find that $h(\phi)=N=d(\phi_\varkappa)=h(\phi_\varkappa)$. $\square$

### 2.5. Construction of adapted coordinates

Adapted coordinates can be constructed by means of a slight modification of an algorithm going back to Varchenko, even in the smooth, finite-type setting. We briefly recall the principal steps here, and refer the reader to [19] for further details.

Assume that $\phi$ is as before. Then, by [19, Theorem 5.1], there exists a smooth function $\psi=\psi(x_1)$ defined on a neighborhood of the origin with $\psi(0)=0$ such that an adapted coordinate system $(y_1,y_2)$ for $\phi$ is given locally near the origin by means of the (in general non-linear) shear

$$y_1 := x_1 \quad \text{and} \quad y_2 := x_2-\psi(x_1).$$

In these coordinates, $\phi$ is given by

$$\phi^a(y) := \phi(y_1, y_2+\psi(y_1)). \tag{2.14}$$

The function $\psi$ can be constructed as follows:

If the coordinates are adapted to $\phi$, then we may choose $\psi:=0$ and are finished.

Otherwise, by our previous discussion, the principal face $\pi(\phi)$ is a compact edge such that $m_1:=\varkappa_2/\varkappa_1\in\mathbb{N}$ and $m(\phi_{\mathrm{pr}})>d(\phi_{\mathrm{pr}})$. We then choose a real root $x\mapsto b_1x_1^{m_1}$ of the principal part $\phi_{\mathrm{pr}}$ of $\phi$ of maximal multiplicity $N_0:=m(\phi_{\mathrm{pr}})$, i.e., the principal root. Note that $b_1\neq 0$.

*Step* 1. We apply the real change of variables $x=s(y)$ given by

$$y_1 := x_1 \quad \text{and} \quad y_2 := x_2-b_1x_1^{m_1},$$

and put $\tilde\phi:=\phi\circ s$. Let us endow all quantities associated with $\tilde\phi$ with a tilde. This change of coordinates has the following effects on the Newton polyhedron: all edges of $\mathcal{N}(\phi)$

lying "to the left" of the principal face $\pi(\phi)$ are preserved in $\mathcal{N}(\tilde{\phi})$, but the principal edge $\mathcal{N}(\phi)$ shrinks along the line on which it lies, keeping the left endpoint fixed. It may either become an edge of $\mathcal{N}(\tilde{\phi})$, lying strictly above the bisectrix, or even a single point (cf. the discussion in [19, §4.2]).

Now, if the coordinates $y$ are adapted to $\phi$, we choose $\psi(x_1):=b_1x_1^{m_1}$ and are done.

Otherwise, the principal face $\pi(\tilde{\phi})$ is a compact edge, and we have $\tilde{\varkappa}_2/\tilde{\varkappa}_1=:m_2\in\mathbb{N}$ and $N_1:=m(\tilde{\phi}_{\mathrm{pr}})>d(\tilde{\phi})$. Recall that the principal part $\tilde{\phi}_{\mathrm{pr}}$ of $\tilde{\phi}$ is $\tilde{\varkappa}$-homogeneous of degree 1. One can then show that

$$m_2>m_1, \quad N_1\leqslant N_0 \quad \text{and} \quad d(\tilde{\phi})>d(\phi). \tag{2.15}$$

*Subsequent steps.* Now, either the new coordinates $y$ are adapted, in which case we are done. Or we can apply the same procedure to $\tilde{\phi}$, etc. In this way, we obtain a sequence of functions $\phi_{(k)}=\phi\circ s_{(k)}$, with $\phi_{(0)}:=\phi$ and $\phi_{(k+1)}:=\tilde{\phi}_{(k)}$, which can be obtained from the original coordinates $x$ by means of a change of coordinates $x=s_{(k)}(y)$ of the form

$$y_1:=x_1 \quad \text{and} \quad y_2:=x_2-\sum_{l=1}^{k} b_l x_1^{m_l},$$

with positive integers $m_1<m_2<...<m_k<m_{k+1}<...$ and real coefficients $b_l\neq 0$. Moreover, if $N_k:=m((\phi_{(k)})_{\mathrm{pr}})$ denotes the maximal order of vanishing of the principal part of $\phi_{(k)}$ along the unit circle $S^1$, we then have

$$N_0\geqslant N_1\geqslant ...\geqslant N_k\geqslant N_{k+1}\geqslant ... \,. \tag{2.16}$$

Either this procedure will stop after finitely many steps, or it will continue infinitely. If it stops, say, at the $K$th step, it is clear that we will have arrived at an adapted coordinate system $x=s_{(K)}(y)$, with a polynomial function $\psi(x_1)=\sum_{l=1}^{K} b_l x_1^{m_l}$.

*Final step.* Assume that the procedure does not terminate. Since by (2.16) the maximal multiplicities $N_k$ of the real roots of the principal part of $\phi_{(k)}$ form a decreasing sequence, there exist $K_0, N\in\mathbb{N}$ such that $N_k=N$ for every $k\geqslant K_0$. We assume that $K_0$ is chosen minimal with this property. It has been shown in [19] that the principal part of $\phi_{(k)}$ is then of the form

$$(\phi_{(k)})_{\mathrm{pr}}(x)=c_k x_1^{\nu_1}(x_2-b_{k+1}x_1^{m_{k+1}})^N \quad \text{for every } k\geqslant K_0, \tag{2.17}$$

where $\nu_1<N$.

Now, according to a classical theorem of E. Borel (cf. [17, Theorem 1.2.6]) one can find a smooth function $\psi(x_1)$ near the origin whose Taylor series is the formal series

$\sum_{l=1}^{\infty} b_l x_1^{m_l}$. Then the coordinates $y_1{:=}x_1$ and $y_2{:=}x_2-\psi(x_1)$ turn out to be adapted to $\phi$. More precisely, one finds that in these coordinates, $\phi$ is given by $\phi^a(y)$, where the principal face of $\mathcal{N}(\phi^a)$ is the unbounded horizontal half-line given by $t_1\geqslant\nu_1$ and $t_2=N$. Notice that here $N=h(\phi)$.

In the case of an analytic function $\phi$, the function $\psi$ can be chosen to be analytic, in which case it can be identified with one of the roots of $\phi$, called the principal root, or a partial sum of the Puiseux series expansion of it (see [19], and cf. also §3), which is why we called it the *principal root jet* in [19]. With some slight abuse of notation, it will be given the same name also in the smooth case.

Notice that only when the principal face of the Newton polyhedron of $\phi^a$ is unbounded, $\psi$ may not be a polynomial.

We shall in the sequel distinguish the following three cases (cf. §2.4):

(a) $\pi(\phi^a)$ is a compact edge, and either $\varkappa_2^a/\varkappa_1^a\notin\mathbb{N}$, or $\varkappa_2^a/\varkappa_1^a\in\mathbb{N}$ and $m(\phi^a_{\mathrm{pr}})\leqslant d(\phi)$;

(b) $\pi(\phi^a)$ consists of a vertex;

(c) $\pi(\phi^a)$ is unbounded.

Here, in case (a), the *principal weight* $\varkappa^a=(\varkappa_1^a,\varkappa_2^a)$ is the unique weight such that the principal part $\phi^a_{\mathrm{pr}}$ of $\phi^a$ is $\varkappa^a$-homogeneous of degree 1.

## 3. Newton diagram, Pusieux expansion of roots and weights associated with $\phi^a$

It is well known that with each edge $\gamma_l$ of the Newton diagram of $\phi^a$ one can associate a unique weight $\varkappa^l$ and a corresponding principal part $\phi^a_{\varkappa^l}$ of $\phi^a$, which can indeed be read off from the Pusieux series expansion of roots of $\phi^a$, at least in the analytic case (cf. [31]).

In order to recall these results, let us assume here that $\phi$ is real-analytic (in §8.5 we shall explain how the general case can be reduced to the analytic setting). As in [19], we shall make use of results and notation from Phong and Stein's article [31]. It is well known that we may write

$$\phi^a(y_1,y_2)=U(y_1,y_2)y_1^{\nu_1}y_2^{\nu_2}\prod_r(y_2-r(y_1)),$$

where the product runs over all non-trivial roots $r=r(y_1)$ of $\phi^a$ (which may also be empty) and where $U(0,0)\neq0$. Moreover, according to [31], these roots can be expressed in a small neighborhood of 0 as Puiseux series

$$r(y_1)=c_{l_1}^{\alpha_1}y_1^{a_{l_1}}+c_{l_1l_2}^{\alpha_1\alpha_2}y_1^{a_{l_1l_2}^{\alpha_1}}+\ldots+c_{l_1\ldots l_p}^{\alpha_1\ldots\alpha_p}y_1^{a_{l_1\ldots l_p}^{\alpha_1\ldots\alpha_{p-1}}}+\ldots,$$

where

$$c_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}\beta}\neq c_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}\gamma}\quad\text{for}\quad \beta\neq\gamma,$$

and

$$a_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}}>a_{l_1\dots l_{p-1}}^{\alpha_1\dots\alpha_{p-2}},$$

with strictly positive exponents $a_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}}>0$ and non-zero complex coefficients $c_{l_1\dots l_p}^{\alpha_1\dots\alpha_p}$, and where we have kept enough terms to distinguish between all the non-identical roots of $\phi^a$. These roots can be grouped into clusters:

The *cluster*

$$\begin{bmatrix}\alpha_1 & \dots & \alpha_p\\ l_1 & \dots & l_p\end{bmatrix}$$

designates all the roots $r(y_1)$, counted with their multiplicities, which satisfy

$$r(y_1)-\Big(c_{l_1}^{\alpha_1}y_1^{a_{l_1}}+c_{l_1l_2}^{\alpha_1\alpha_2}y_1^{a_{l_1l_2}^{\alpha_1}}+\dots+c_{l_1\dots l_p}^{\alpha_1\dots\alpha_p}y_1^{a_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}}}\Big)=O(y_1^b)\tag{3.1}$$

for some exponent $b>a_{l_1\dots l_p}^{\alpha_1\dots\alpha_{p-1}}$. We also introduce the clusters

$$\begin{bmatrix}\alpha_1 & \dots & \alpha_{p-1} & \cdot\\ l_1 & \dots & l_{p-1} & l_p\end{bmatrix}:=\bigcup_{\alpha_p}\begin{bmatrix}\alpha_1 & \dots & \alpha_p\\ l_1 & \dots & l_p\end{bmatrix}.$$

Each index $\alpha_p$ or $l_p$ varies in some finite range which we shall not specify here. We finally put

$$N\begin{bmatrix}\alpha_1 & \dots & \alpha_p\\ l_1 & \dots & l_p\end{bmatrix}:=\text{number of roots in}\begin{bmatrix}\alpha_1 & \dots & \alpha_p\\ l_1 & \dots & l_p\end{bmatrix},$$

$$N\begin{bmatrix}\alpha_1 & \dots & \alpha_{p-1} & \cdot\\ l_1 & \dots & l_{p-1} & l_p\end{bmatrix}:=\text{number of roots in}\begin{bmatrix}\alpha_1 & \dots & \alpha_{p-1} & \cdot\\ l_1 & \dots & l_{p-1} & l_p\end{bmatrix}.$$

Let $a_1<\dots<a_l<\dots<a_n$ be the distinct leading exponents of all the roots $r$. Each exponent $a_l$ corresponds to the cluster $\left[\begin{smallmatrix}\cdot\\ l\end{smallmatrix}\right]$, so that the set of all roots $r$ can be divided as $\bigcup_{l=1}^n\left[\begin{smallmatrix}\cdot\\ l\end{smallmatrix}\right]$. Correspondingly, we can decompose

$$\phi^a(y_1,y_2)=U(y_1,y_2)y_1^{\nu_1}y_2^{\nu_2}\prod_{l=1}^n\Phi\begin{bmatrix}\cdot\\ l\end{bmatrix}(y_1,y_2),$$

where

$$\Phi\begin{bmatrix}\cdot\\ l\end{bmatrix}(y_1,y_2):=\prod_{r\in\left[\begin{smallmatrix}\cdot\\ l\end{smallmatrix}\right]}(y_2-r(y_1)).$$

We introduce the following quantities:

$$A_l = A\begin{bmatrix} \cdot \\ l \end{bmatrix} := \nu_1 + \sum_{\mu \leqslant l} a_\mu N\begin{bmatrix} \cdot \\ \mu \end{bmatrix} \quad \text{and} \quad B_l = B\begin{bmatrix} \cdot \\ l \end{bmatrix} := \nu_2 + \sum_{\mu \geqslant l+1} N\begin{bmatrix} \cdot \\ \mu \end{bmatrix}. \tag{3.2}$$

Then the vertices of the Newton diagram $\mathcal{N}_d(\phi^a)$ of $\phi^a$ are the points $(A_l, B_l)$, $l=0, ..., n$, and the Newton polyhedron $\mathcal{N}(\phi^a)$ is the convex hull of the set $\bigcup_l ((A_l, B_l)+\mathbb{R}^2_+)$.

We note that there is at least one vertex, namely $(A_0, B_0)$, and it lies strictly above the bisectrix, i.e., $A_0 < B_0$, because the original coordinates $x$ were assumed to be non-adapted, so that the left endpoint of the principal face $\pi(\phi)$ is a vertex of $\mathcal{N}(\phi)$ lying strictly above the bisectrix, which thus will also be a vertex of $\mathcal{N}(\phi^a)$.

Let $L_l := \{(t_1, t_2) \in \mathbb{N}^2 : \varkappa_1^l t_1 + \varkappa_2^l t_2 = 1\}$ denote the line passing through the points $(A_{l-1}, B_{l-1})$ and $(A_l, B_l)$. Then

$$\frac{\varkappa_2^l}{\varkappa_1^l} = a_l, \tag{3.3}$$

which in return is the reciprocal of the slope of the line $L_l$. The line $L_l$ intersects the bisectrix at the point $(d_l, d_l)$, where

$$d_l := \frac{A_l + a_l B_l}{1+a_l} = \frac{A_{l-1} + a_l B_{l-1}}{1+a_l}. \tag{3.4}$$

Finally, the $\varkappa^l$-principal part $\phi^a_{\varkappa^l}$ of $\phi^a$ corresponding to the supporting line $L_l$ is given by

$$\phi^a_{\varkappa^l}(y) = c_l y_1^{A_{l-1}} y_2^{B_l} \prod_\alpha (y_2 - c_l^\alpha y_1^{a_l})^{N\left[\begin{smallmatrix} \alpha \\ l \end{smallmatrix}\right]}. \tag{3.5}$$

In view of this identity, we shall say that the edge $\gamma_l := [(A_{l-1}, B_{l-1}), (A_l, B_l)]$ is *associated* with the cluster of roots $\left[\begin{smallmatrix} \cdot \\ l \end{smallmatrix}\right]$.

Distinguishing the cases listed below, we next single out a particular edge by fixing the corresponding index $l=\lambda$:

(a) In case (a), where the principal face of $\phi^a$ is a compact edge, we choose $\lambda$ so that the edge $\gamma_\lambda = [(A_{\lambda-1}, B_{\lambda-1}), (A_\lambda, B_\lambda)]$ is the principal face $\pi(\phi^a)$ of the Newton polyhedron of $\phi^a$.

(b) In case (b), where $\pi(\phi^a)$ is the vertex $(h, h)$, with $h=h(\phi)$, we choose $\lambda$ so that $(h, h) = (A_\lambda, B_\lambda)$. Then $\lambda \geqslant 1$, and $(h, h)$ is the right endpoint of the compact edge $\gamma_\lambda$.

(c) Consider finally case (c), in which the principal face $\pi(\phi^a)$ is unbounded, namely the half-line given by $t_1 \geqslant \nu_1$ and $t_2 = h$, where $\nu_1 < h$. Here, we distinguish two subcases:

(c1) If the point $(\nu_1, h)$ is the right endpoint of a compact edge of $\mathcal{N}(\phi^a)$, then we choose again $\lambda \geqslant 1$ so that this edge is given by $\gamma_\lambda$.

(c2) Otherwise, $(\nu_1, h) = (A_0, B_0)$ is the only vertex of $\mathcal{N}(\phi^a)$, that is,

$$\mathcal{N}(\phi^a) = (\nu_1, h) + \mathbb{R}^2_+.$$

In this case, there is no non-trivial root $r$, and hence $n=0$.

In the cases (a), (b) and (c1), let us put

$$a := a_\lambda = \frac{\varkappa_2^\lambda}{\varkappa_1^\lambda}. \tag{3.6}$$

Notice that our dicussion of the algorithm in §2.5 (cf. step 1 and (2.15)) shows that $a > m_1$ in case (a), and $a \geqslant m_1$ in cases (b) and (c1), since the original coordinates had not been adapted.

The following result will become useful later.

LEMMA 3.1. *If* $h(\phi) \geqslant 2$, *then in the cases* (a), (b) *and* (c1) *described before* $\partial_2^2 \phi^a_{\varkappa^l}$ *does not vanish identically*, *and* $\varkappa_2^l < 1$ *for every* $l \leqslant \lambda$.

*Proof.* Consider first the situation where the principal face $\pi(\phi^a)$ is a compact edge. Here the statement will already follow from our general assumption $\nabla\phi(0)=0$. Indeed, write $\phi^a_{\varkappa^l}$ according to (3.5) in the form

$$\phi^a_{\varkappa^l}(y) = c y_1^{\nu_1} y_2^{\nu_2} \prod_{s=1}^{M} (y_2 - \lambda_s y_1^{a_l})^{n_s},$$

with $\lambda_s \neq 0$. Note that $M \geqslant 1$, since $\gamma_\lambda = \pi(\phi^a)$ is a compact edge.

If we assume that $\partial_2^2 \phi^a_{\varkappa^l} = 0$, then clearly $\nu_2 + \sum_s n_s \leqslant 1$, so that there is only one non-trivial, real root $\lambda_1 y_1^{a_l}$ of multiplicity 1. This implies that $\phi^a_{\varkappa^l}(y) = c y_1^{\nu_1}(y_2 - \lambda_1 y_1^{a_l})$. Thus the Newton diagram $\gamma_l = \mathcal{N}_d(\phi^a_{\varkappa^l})$ is the interval $[(\nu_1, 1), (\nu_1 + a_l, 0)]$. Since $l \leqslant \lambda$, its left endpoint must lie above the bisectrix, so that $\nu_1 = 0$. But then $\nabla \phi^a_{\varkappa^l}(0) \neq 0$, and hence $\nabla\phi(0) \neq 0$, a contradiction.

A similar argument applies to show that $\varkappa_2^l < 1$. Indeed, since the polynomial $\tilde{\phi}_{\varkappa^l}$ is $\varkappa^l$-homogeneous of degree 1, as $M \geqslant 1$, $\varkappa_2^l \geqslant 1$ would imply that $\nu_1 = \nu_2 = 0$ and $\sum_s n_s = 1$, so that we could conclude as before that $\nabla\phi(0) \neq 0$.

Finally, if $\pi(\phi^a)$ is a vertex or an unbounded edge of type (c1), then the previous arguments still apply, with minor modifications. □

COROLLARY 3.2. *Denote by* $S_0^1$ *the set of all points* $x = (x_1, x_2)$ *on the unit circle* $S^1$ *such that* $x_1 \neq 0$, *and suppose that* $a = \varkappa_2^\lambda / \varkappa_1^\lambda > 2$. *We also assume that* $h := h(\phi) \geqslant 2$.

(i) *If* $1 \leqslant l < \lambda$, *then for every* $y^0 \in S_0^1$, *with* $y_2^0 \neq 0$, *there is some* $j$,

$$2 \leqslant j \leqslant \max\{2, d_h(\phi^a_{\varkappa^l})\},$$

*such that*

$$\partial_2^j \phi^a_{\varkappa^l}(y^0) \neq 0. \tag{3.7}$$

*Moreover*, *we have* $d_h(\phi^a_{\varkappa^l}) < h$.

(ii) *If* $l=\lambda$, *and if the either* $\pi(\phi^a)$ *is a vertex as in case* (b), *or an unbounded face as in case* (c1), *then the conclusion in* (i) *remains true. Moreover, if* $y^0\in S_0^1$ *and* $y_2^0=0$, *then* (3.7) *holds true for* $j=h$. *We also note that* $d_h(\phi^a_{\varkappa^\lambda})=h$ *in case* (b) *and* $d_h(\phi^a_{\varkappa^\lambda})<h$ *in case* (c1).

(iii) *If* $l=\lambda$, *and if* $\pi(\phi^a)$ *is a compact edge given by* $\gamma_\lambda$ *as in case* (a), *then there is at most one point* $y^m$ *in* $S_0^1$ *such that*

$$\partial_2^j \phi^a_{\varkappa^\lambda}(y^m)=0 \quad \text{for } 1\leqslant j\leqslant h. \tag{3.8}$$

*More precisely, if* $a\in\mathbb{N}$, *then* (3.7) *holds true for every point* $y^0\in S_0^1$ *for some* $j$, $2\leqslant j\leqslant h$, *with the possible exception of one single point* $y^m$.

*Also, if* $a\notin\mathbb{N}$, *then for every* $y^0\in S_0^1$ *with* $y_2^0\neq 0$ *there is some* $j$, $1\leqslant j<d_h(\phi^a_{\varkappa^\lambda})=h$, *so that* (3.7) *holds true.*

*Proof.* Let us write $\phi_l^a:=\phi^a_{\varkappa^l}$. Assume first that $l<\lambda$. From the geometry of the Newton polyhedron of $\phi^a$, it is evident that $d_h(\phi_l^a)<d_h(\phi_\lambda^a)\leqslant h$, and that the principal face of the Newton polyhedron of $\phi_l^a$ is a horizontal half-line contained in the line $t_2=h(\phi_l^a)>h\geqslant 2$, so that $d(\phi_l^a)=h(\phi_l^a)>d_h(\phi_l^a)$. The statements in (i) are thus immediate from Lemma 3.1 and Proposition 2.3 (b), applied to $P=\phi_l^a$.

Assume next that $l=\lambda$. The claim in (ii) follows in exactly the same way from Proposition 2.3 (b), applied to $P=\phi_\lambda^a$, if $y_2^0\neq 0$. So assume that $y^0$ lies on the $y_1$-axis, say $y^0=(1,0)$. But, in the cases (b) and (c1), the $\varkappa^\lambda$- principal part of $\phi^a$ is of the from $\phi_\lambda^a(y)=y_2^h Q(y_1,y_2)$, where $Q$ is a polynomial such that $Q(1,0)\neq 0$. This implies that $\partial_2^h\phi_\lambda^a(1,0)\neq 0$, which completes the proof of (ii).

Finally, the statements in (iii) follow from Lemma 3.1 and Proposition 2.3, statements (a) and (c), applied to $P=\phi_\lambda^a$. □

If the coordinates $x$ are not adapted to $\phi$, then Corollary 3.2 will eventually allow us to control the contributions to the maximal operator $\mathcal{M}$ of major parts of the surface $S$, except possibly for a part corresponding to a small neighborhood of the principal root jet.

In case we can choose $2\leqslant j\leqslant h$ in (3.7) or (3.8), this can be achieved by means of a reduction to maximal functions along curves. The corresponding estimates will be provided in the next section.

If we need to choose $j=1$ in (3.7) or (3.8), such a 1-dimensional reduction is no longer possible, and we shall have to deal with uniform estimates of 2-dimensional oscillatory integrals depending on small parameters. The same applies to the small neighborhood of the principal root jet mentioned before, only that the required estimates in this case are deeper. These estimates will be provided in §5.

## 4. Uniform estimates for maximal operators associated with families of finite-type curves and related surfaces

### 4.1. Finite-type curves

In this subsection, we shall prove an extension of some results by Iosevich [21], which allows for uniform estimates for maximal operators associated with families of curves which arise as small perturbations of a given curve.

We begin with a result whose proof is based on Iosevich's approach in [21].

PROPOSITION 4.1. *Consider averaging operators along curves in the plane of the form*

$$A_t f(x)=A_t^{(\varrho,\eta,\tau)} f(x):=\int_{\mathbb{R}} f(x_1-t(\varrho_1 s+\eta_1), x_2-t(\eta_2+\tau s+\varrho_2 g(s)))\psi(s)\,ds,$$

*where* $\varrho=(\varrho_1,\varrho_2)$, $\eta=(\eta_1,\eta_2)\in\mathbb{R}^2$, $\varrho_1>0$, $\varrho_2>0$, $\tau\in\mathbb{R}$, *and* $\psi\in C_0^\infty(\mathbb{R})$ *is supported in a bounded interval* $I$ *containing the origin, and where*

$$g(s)=s^m(b(s)+R(s)),\quad s\in I,\ m\in\mathbb{N},\ m\geqslant 2, \tag{4.1}$$

*with* $b\in C^\infty(I,\mathbb{R})$ *satisfying* $b(0)\neq 0$. *Also,* $R\in C^\infty(I,\mathbb{R})$ *is a smooth perturbation term.*

*By* $\mathcal{M}^{(\varrho,\eta,\tau)}$ *we denote the associated maximal operator*

$$\mathcal{M}^{(\varrho,\eta,\tau)} f(x):=\sup_{t>0}|A_t^{(\varrho,\eta,\tau)} f(x)|.$$

*Then, there exist a neighborhood* $U$ *of the origin in* $I$, $M\in\mathbb{N}$ *and* $\delta>0$, *such that, for* $p>m$,

$$\|\mathcal{M}^{(\varrho,\eta,\tau)} f\|_p\leqslant C_p\biggl(\frac{|\eta_1|}{\varrho_1}+\frac{|\eta_2-\tau\eta_1/\varrho_1|}{\varrho_2}+1\biggr)^{1/p}\|f\|_p,\quad f\in\mathcal{S}(\mathbb{R}^2), \tag{4.2}$$

*for every* $\psi$ *supported in* $U$ *and every* $R$ *with* $\|R\|_{C^M}<\delta$, *with a constant* $C_p$ *depending only on* $p$ *and the* $C^M$*-norm of* $\psi$ (*such constants will be called* admissible).

*Proof.* Consider the linear operator

$$Tf(x_1,x_2)=\frac{1}{(\varrho_1\varrho_2)^{1/p}} f\biggl(\frac{x_1}{\varrho_1},\frac{1}{\varrho_2}\biggl(x_2-\frac{\tau}{\varrho_1}x_1\biggr)\biggr).$$

Then $T$ is isometric on $L^p(\mathbb{R}^2)$, and one computes that $\tilde{A}_t:=T^{-1}A_tT$ is given by

$$\tilde{A}_t f(x)=\tilde{A}_t^\sigma f(x)=\int_{\mathbb{R}} f(x_1-t(s+\sigma_1), x_2-t(\sigma_2+g(s)))\psi(s)\,ds,$$

where $\sigma=(\sigma_1,\sigma_2)$ is given by

$$\sigma_1=\frac{\eta_1}{\varrho_1} \quad \text{and} \quad \sigma_2=\frac{\eta_2}{\varrho_2}-\frac{\tau\eta_1}{\varrho_1\varrho_2}.$$

Put

$$\widetilde{\mathcal{M}}f(x)=\sup_{t>0}|\tilde{A}_t f(x)|.$$

Then (4.2) is equivalent to the following estimate for $\widetilde{\mathcal{M}}$:

$$\|\widetilde{\mathcal{M}}f\|_p\leqslant C_p(|\sigma|+1)^{1/p}\|f\|_p, \quad f\in\mathcal{S}(\mathbb{R}^2), \tag{4.3}$$

for every $\sigma\in\mathbb{R}^2$, where $C_p$ is an admissible constant.

(a) We first consider the case $m=2$. By means of the Fourier inversion formula, we can write

$$\tilde{A}_t f(x)=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} e^{i(x-t\sigma)\cdot\xi}H(t\xi)\hat{f}(\xi)\,d\xi,$$

where

$$H(\xi_1,\xi_2):=\int_{\mathbb{R}} e^{-i(\xi_1 s+\xi_2 g(s))}\psi(s)\,ds.$$

A standard application of the method of stationary phase then yields that

$$H(\xi)=e^{iq(\xi)}\frac{\chi(\xi_1/\xi_2)A(\xi)}{(1+|\xi|)^{1/2}}+B(\xi),$$

where $\chi$ is a smooth function supported on a small neighborhood of the origin. Moreover, $q(\xi)=q(\xi,R)$ is a smooth function of $\xi$ and $R$ which is homogenous of degree 1 in $\xi$, and which can be considered as a small perturbation of $q(\xi,0)$, if $R$ is contained in a sufficiently small neighborhood of 0 in $C^\infty(I,\mathbb{R})$. The Hessian $D_\xi^2 q(\xi,0)$ has rank 1, so that the same applies to $D_\xi^2 q(\xi,R)$ for small perturbations $R$. Moreover, $A$ is a symbol of order zero such that $A(\xi)=0$, if $\|\xi\|\leqslant C$, and

$$|\xi^\alpha D_\xi^\alpha A(\xi)|\leqslant C_\alpha, \quad \alpha\in\mathbb{N}^2,\ |\alpha|\leqslant 3, \tag{4.4}$$

where the $C_\alpha$ are admissible constants. Finally, $B$ is a remainder term satisfying

$$|D_\xi^\alpha B(\xi)|\leqslant C_{\alpha,N}(1+|\xi|)^{-N}, \quad |\alpha|\leqslant 3,\ 0\leqslant N\leqslant 3, \tag{4.5}$$

again with admissible constants $C_{\alpha,N}$.

If we put

$$\tilde{A}_t^0 f(x):=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} e^{i(x-t\sigma)\cdot\xi}B(t\xi)\hat{f}(\xi)\,d\xi,$$

then, by (4.5), $\tilde{A}_t^0 f(x) = f * k_t^\sigma(x)$, where $k_t^\sigma(x) = t^{-2} k^\sigma(x/t)$ and $k^\sigma$ is the translate

$$k^\sigma(x) := k(x-\sigma) \tag{4.6}$$

of $k$ by the vector $\sigma$ of a fixed function $k$ satisfying an estimate of the form

$$|k(x)| \leqslant \frac{C}{(1+|x|)^3}. \tag{4.7}$$

Let

$$\widetilde{\mathcal{M}}^0 f(x) := \sup_{t>0} |\tilde{A}_t^0(x)|$$

denote the corresponding maximal operator. Now, (4.6) and (4.7) show that

$$\|\widetilde{\mathcal{M}}^0\|_{L^\infty \to L^\infty} \leqslant C,$$

with a constant $C$ which does not depend on $\sigma$. Moreover, scaling by the factor

$$\frac{1}{|\sigma|+1}$$

in the direction of the vector $\sigma$, we see that $\|\widetilde{\mathcal{M}}^0\|_{L^1 \to L^{1,\infty}} \leqslant C(|\sigma|+1)$, since we then can compare with $(|\sigma|+1)M$, where $M$ is the Hardy–Littlewood maximal operator. By interpolation, these estimates imply that (4.3) holds for $\widetilde{\mathcal{M}}^0$ in place of $\widetilde{\mathcal{M}}$, for every $p>1$.

The maximal operator $\widetilde{\mathcal{M}}^1$ corresponding to the family of averaging operators

$$\tilde{A}_t^1(x) := \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{i[\xi \cdot x - t(\sigma \cdot \xi + q(\xi))]} \frac{\chi(\xi_1/\xi_2) A(t\xi)}{(1+|t\xi|)^{1/2}} \hat{f}(\xi)\, d\xi$$

remains to be studied. Choose a non-negative function $\beta \in C_0^\infty(\mathbb{R})$ such that

$$\operatorname{supp} \beta \subset \left[\tfrac{1}{2}, 2\right] \quad \text{and} \quad \sum_{j=-\infty}^{\infty} \beta(2^{-j} r) = 1 \ \text{ for } r > 0,$$

and put

$$A_{j,t} f(x) := \int_{\mathbb{R}^2} e^{i[\xi \cdot x - t(\sigma \cdot \xi + q(\xi))]} \frac{\chi(\xi_1/\xi_2) A(t\xi)}{(1+t|\xi|)^{1/2}} \beta(2^{-j}|t\xi|) \hat{f}(\xi)\, d\xi.$$

Since we may assume that $A$ vanishes on a sufficiently large neighborhood of the origin, we have $A_{j,t} f = 0$, if $j \leqslant 0$, so that

$$\tilde{A}_t^1 f(x) = \sum_{j=1}^{\infty} A_{j,t} f(x). \tag{4.8}$$

Denote by $\mathcal{M}_j$ the maximal operator associated with the averages $A_{j,t}$, $t>0$.

Since $A_{j,t}$ is localized to frequencies $|\xi|\sim 2^j/t$, we can use Littlewood–Paley theory (see [39]) to see that

$$\|\mathcal{M}_j\|_{L^p\to L^p} \lesssim \|\mathcal{M}_{j,\mathrm{loc}}\|_{L^p\to L^p}, \tag{4.9}$$

where $\mathcal{M}_{j,\mathrm{loc}}f(x):=\sup_{1\leqslant t\leqslant 2}|A_{j,t}f(x)|$.

Choose a bump function $\varrho\in C_0^\infty(\mathbb{R})$ supported in $\left[\frac{1}{2},4\right]$ such that $\varrho(t)=1$, if $1\leqslant t\leqslant 2$. In order to estimate $\mathcal{M}_{j,\mathrm{loc}}$, we use the following well-known estimate (see, e.g., [21, Lemma 1.3]):

$$\sup_{t\in\mathbb{R}}|\varrho(t)A_{j,t}f(x)|^p \leqslant p\biggl(\int_{-\infty}^{\infty}|\varrho(t)A_{j,t}f(x)|^p\,dt\biggr)^{1/p'}\biggl(\int_{-\infty}^{\infty}\biggl|\frac{\partial}{\partial t}(\varrho(t)A_{j,t}f(x))\biggr|^p\,dt\biggr)^{1/p},$$

which follows by integration by parts. By Hölder's inequality, this implies that

$$\begin{aligned}\|\mathcal{M}_{j,\mathrm{loc}}f\|_p^p \leqslant C\biggl(\int_{\mathbb{R}^2}\int_{1/2}^4|A_{j,t}f(x)|^p\,dt\,dx\biggr)^{(p-1)/p}\biggl(\int_{\mathbb{R}^2}\int_{1/2}^4\biggl|\frac{\partial}{\partial t}A_{j,t}f(x)\biggr|^p\,dt\,dx\biggr)^{1/p}\\ +C\int_{\mathbb{R}}\int_{1/2}^4|A_{j,t}f(x)|^p\,dt\,dx.\end{aligned} \tag{4.10}$$

Moreover,

$$\frac{\partial}{\partial t}A_{j,t}f(x)=\int_{\mathbb{R}^2}e^{i[\xi\cdot x-t(\sigma\cdot\xi+q(\xi))]}\chi\biggl(\frac{\xi_1}{\xi_2}\biggr)h(t,j,\xi)\,d\xi,$$

where

$$\begin{aligned}h(t,j,\xi)=-i\frac{\sigma\cdot\xi+q(\xi)}{(1+t|\xi|)^{1/2}}A(t\xi)\beta(2^{-j}t|\xi|)+\frac{\partial}{\partial t}\biggl[\frac{A(t\xi)}{(1+t|\xi|)^{1/2}}\biggr]\beta(2^{-j}t|\xi|)\\ +\frac{A(t\xi)}{(1+t|\xi|)^{1/2}}2^{-j}|\xi|\beta'(2^{-j}t|\xi|).\end{aligned}$$

Now, if $t\sim 1$, since $A$ vanishes near the origin, it is easy to see that the amplitude of $A_{j,t}$ can be written as $2^{-j/2}a_{j,t}(\xi)$, where $a_{j,t}$ is a symbol of order zero localized where $|\xi|\sim 2^j$. Similarly, the amplitude of $\partial A_{j,t}/\partial t$ can be written as $2^{j/2}(|\sigma|+1)b_{j,t}$, where $b_{j,t}$ is a symbol of order zero localized where $|\xi|\sim 2^j$, and $a_{j,t}$ and $b_{j,t}$ satisfy estimates of the form

$$|(1+|\xi|)^{|\alpha|}(|D^\alpha a_{j,t}(\xi)|+|D^\alpha b_{j,t}(\xi)|)|\leqslant C_\alpha,$$

with admissible constants $C_\alpha$.

We can then apply the local smoothing estimates by Mockenhaupt, Seeger and Sogge from [28] and [29] for operators of the form

$$P_jf(x,t)=\int e^{i(\xi\cdot x-tq(\xi))}a(t,\xi)\beta(2^{-j}|\xi|)\hat{f}(\xi)\,d\xi,$$

where $a(t,\xi)$ is a symbol of order zero in $\xi$, and the Hessian matrix of $q$ has rank 1 everywhere. Their results imply in particular that, for $2<p<\infty$,

$$\left(\int_{1/2}^{4}\int_{\mathbb{R}^2}|P_jf(x)t|^p\,dx\,dt\right)^{1/p}\leqslant C_p2^{j(1/2-1/p-\delta(p))}\|f\|_{L^p(\mathbb{R}^2)} \tag{4.11}$$

for some $\delta(p)>0$.

Since

$$2^{j/2}A_{j,t}f(x)\quad\text{and}\quad 2^{-j/2}\frac{1}{|\sigma|+1}\frac{\partial}{\partial t}A_{j,t}f(x)$$

are of the form $P_jf(x-t\sigma)$, for suitable operators $P_j$ of this type, we can apply (4.10) and (4.11) to obtain, if $R=0$,

$$\|\mathcal{M}_{j,\mathrm{loc}}f\|_p\leqslant C_p2^{j(1/2-1/p-\delta(p))}2^{-(j/2(p-1)+j/2)/p}(|\sigma|+1)^{1/p}\|f\|_p,$$

that is,

$$\|\mathcal{M}_{j,\mathrm{loc}}f\|_p\leqslant C_p(|\sigma|+1)^{1/p}2^{-\delta(p)j}\|f\|_p, \tag{4.12}$$

if $2<p<\infty$, where $\delta(p)>0$.

However, as observed in [21], the estimate (4.11) remains valid under small, sufficiently smooth perturbations, and the constant $C_p$ depends only on a finite number of derivatives of the phase function and the symbol of $P_j$. Therefore, if $\delta$ is sufficiently small and $\|R\|_{C^M}<\delta$, then estimate (4.12) holds true also for $R\neq 0$, with an admissible constant $C_p$.

Summing over all $j\geqslant 1$ (cf. (4.8)), we thus get

$$\|\widetilde{\mathcal{M}}^1f\|_p\leqslant C_p(|\sigma|+1)^{1/p}\|f\|_p,\quad\text{if } p>2,$$

with an admissible constant $C_p$. This finishes the proof of the proposition in the case $m=2$.

(b) The case where $m\geqslant 3$ can easily be reduced to the case $m=2$ by means of a dyadic decomposition in the variable $s$ and rescaling of each of the dyadic pieces in a similar way as in [21]. We leave the details to the reader. □

Consider now a smooth function $a\colon I\to\mathbb{R}$, where $I$ is a compact interval of positive length. We say that $a$ is a function of *polynomial type* $m\geqslant 2$ ($m\in\mathbb{N}$), if there is a positive constant $c>0$ such that

$$c\leqslant\sum_{j=2}^{m}|a^{(j)}(s)|\quad\text{for every } s\in I, \tag{4.13}$$

and if $m$ is minimal with this property. Oscillatory integrals with phase functions $a$ of this type have been studied, e.g., by J. E. Björk (see [10]) and G. I. Arkhipov [1], and it is our goal here to estimate related maximal operators, allowing even for small perturbations of $a$. More precisely, consider averaging operators

$$A_t^\varepsilon f(x) := \int_{\mathbb{R}} f(x_1 - ts, x_2 - t(1+\varepsilon(a(s)+r(s))))\psi(s)\,ds, \quad f \in \mathcal{S}(\mathbb{R}^2),$$

along dilates by factors $t>0$ of the curve

$$\gamma(s) := (s, 1+\varepsilon(a(s)+r(s))), \quad s \in I,$$

where $\varepsilon>0$, $\psi\in C_0^\infty(I)$ is a smooth, non-negative density supported in $I$, and $r\in C^\infty(I)$ is a sufficiently small perturbation term. By $\mathcal{M}^\varepsilon$ we denote the corresponding maximal operator

$$\mathcal{M}^\varepsilon f(x) := \sup_{t>0} |A_t^\varepsilon f(x)|.$$

THEOREM 4.2. *Let $a$ be a function of polynomial type $m\geqslant 2$. Then there exist numbers $M\in\mathbb{N}$ and $\delta>0$, such that for every $r\in C^\infty(I,\mathbb{R})$ with $\|r\|_{C^M}<\delta$, $0<\varepsilon\ll 1$ and $p>m$, the following a priori estimate is satisfied*:

$$\|\mathcal{M}^\varepsilon f\|_p \leqslant C_p \varepsilon^{-1/p} \|f\|_p, \quad f \in \mathcal{S}(\mathbb{R}^2), \tag{4.14}$$

*with a constant $C_p$ depending only on $p$.*

*Proof.* By means of an induction argument (based on an idea of Duistermaat [11]), we shall reduce this theorem to Proposition 4.1. Let us fix a smooth function $a: I\to\mathbb{R}$ of polynomial type $m\geqslant 2$. We shall proceed by induction on the type $m$.

Observe first that it suffices to find for every fixed $s_0\in I$ a subinterval $I_0\subset I$ which is relatively open in $I$ and contains $s_0$ such that (4.14) holds for every $\psi$ supported in $I_0$. For, then we can cover $I$ by a finite number of such subintervals $I_j$, decompose $\psi$ by means of a subordinate smooth partition of unity into $\psi=\sum_j \psi_j$, where $\psi_j$ is supported in $I_j$, and apply the estimate (4.14) to each of the pieces.

So, fix $s_0\in I$. Extending the function $a$ in a suitable way to a $C^\infty$-function beyond the boundary points of $I$, we may assume that $s_0$ lies in the interior of $I$. Translating by $s_0$, we may furthermore assume that $s_0=0$. Then, by (4.13), there is some $k\in\mathbb{N}$, $2\leqslant k\leqslant m$, such that

$$a^{(j)}(0) = 0 \text{ for } 2 \leqslant j \leqslant k-1 \quad \text{and} \quad a^{(k)}(0) \neq 0. \tag{4.15}$$

Assume first that $k=2$. Then we may write

$$a(s) = \alpha_0 + \alpha_1 s + s^2 b(s) \quad \text{near } s = 0,$$

where $b\in C^\infty(I)$, $b(0)\neq 0$. Consequently, if $r\in C^\infty(I)$ has sufficiently small $C^{M+2}$-norm, then the estimate (4.14) follows from Proposition 4.1.

Now assume $k>2$.

LEMMA 4.3. *Assume that a satisfies* (4.15) *with* $k\geqslant 3$, *and let* $N\in\mathbb{N}$. *Then there is some* $\delta>0$, *and for every function* $r\in C^\infty(I)$ *with* $\|r\|_{C^{k+N}(I)}<\delta$ *a number* $\sigma(r)\in I$ *with* $\sigma(0)=0$ *and* $|\sigma(r)|\lesssim\delta$, *depending smoothly on* $r$, *such that*

$$(a+r)^{(k-1)}(\sigma(r))=0. \tag{4.16}$$

*In particular*, *if we put* $I_r:=-\sigma(r)+I$, *then*

$$(a+r)(s+\sigma(r))=(b(s)+R(s))s^k+\mu_0+\mu_1 s+\ldots+\mu_{k-2}s^{k-2}, \tag{4.17}$$

*where* $b\in C^\infty(I_r)$ *with* $b(0)\neq 0$, $R\in C^\infty(I_r)$ *with* $\|R\|_{C^N}\lesssim\delta$, *and where* $\mu:=(\mu_0,\ldots,\mu_{m-2})$ *satisfies* $|\mu|\lesssim\delta$.

*Proof.* (4.16) follows from the implicit function theorem, applied to the mapping $f\colon I\times C^{k+N}(I)\to\mathbb{R}$, $f(s,r):=(a+r)^{(k-1)}(s)$, and (4.17) is then a consequence of Taylor's formula. □

The case $k=3$ can now be treated by means of (4.17) in a similar way as the case $k=2$ (notice that $I_r$ and $I$ overlap in a neighborhood $U$ of 0 not depending on $r$, if $\delta$ is sufficiently small, so that we can again assume that $\psi$ is supported in a fixed interval contained in $U$).

We may thus from now on assume that $k\geqslant 4$. Since we have seen that the cases $m=2$ and $m=3$ of Theorem 4.2 are true, we may assume that $m\geqslant 4$, and, by the induction hypothesis, that the statement of Theorem 4.2 is true for all $m'\leqslant m-1$. Then, we may also assume that $k=m$ in (4.15), so that, by Lemma 4.3,

$$(a+r)(s+\sigma(r))=\tilde{b}(s)s^m+\mu_2 s^2+\ldots+\mu_{m-2}s^{m-2}$$

on $I_r$, where $m-2\geqslant 2$ (the affine linear term $\mu_0+\mu_1 s$ can again be omitted by means of a linear change of coordinates). Here we have set $\tilde{b}=b+R$, where, by Lemma 4.3,

$$\|R\|_{C^M}\lesssim\delta.$$

Let us now put $\mu=(\mu_2,\ldots,\mu_{m-2})$. The case $\mu=0$ can again be treated by means of Proposition 4.1, so assume that $\mu\neq 0$. If we scale in $s$ by a factor $\varrho^{1/m}$, $\varrho>0$, we obtain

$$(a+r)(\varrho^{1/m}s+\sigma(r))=\varrho\left[\tilde{b}(\varrho^{1/m}s)s^m+\frac{\mu_2}{\varrho^{(m-2)/m}}s^2+\ldots+\frac{\mu_{m-2}}{\varrho^{2/m}}s^{m-2}\right].$$

This suggests to introduce a quasi-norm

$$N(\mu):=[\mu_2^{m\nu/(m-2)}+\ldots+\mu_{m-2}^{m\nu/2}]^{1/\nu},$$

say with $\nu:=2(m-2)!$. For then $N$ is smooth away from the origin, and if we put $\varrho:=N(\mu)$, i.e., if we define $\xi=(\xi_2,\ldots,\xi_{m-2})$ by

$$\xi_j:=\frac{\mu_j}{N(\mu)^{(m-j)/m}},\quad j=2,\ldots,m-2,$$

then $N(\xi)=1$ and

$$g(s)=g(s,\varrho,\xi):=\frac{1}{\varrho}(a+r)(\varrho^{1/m}s+\sigma(r))=\tilde{b}(\varrho^{1/m}s)s^m+\xi_2s^2+\ldots+\xi_{m-2}s^{m-2}.$$

Then, putting $\eta:=\sigma(r)$, we have

$$A_tf(x)=\varrho^{1/m}\int_{\mathbb{R}}f(x_1-t(\varrho^{1/m}s+\eta),x_2-t(1+\varepsilon\varrho g(s))\psi(\varrho^{1/m}s+\eta)\,ds.$$

Recall at this point that $\eta\to0$ and $\varrho\to0$, as $\delta\to0$. In particular we may consider $g(s,\varrho,\xi)$ as a $C^\infty$-perturbation of $g(s,0,\xi)$, where

$$g(s,0,\xi)=\tilde{b}(0)s^m+\xi_2s^2+\ldots+\xi_{m-2}s^{m-2}.$$

Denote by $\Sigma$ the unit sphere $\Sigma:=\{\xi\in\mathbb{R}^{m-3}:N(\xi)=1\}$ with respect to the quasi-norm $N$, and choose $B>1$ so large that

$$|g''(s)|\geqslant c|s|^{m-2}\quad\text{whenever } |s|\geqslant B,\ \xi\in\Sigma \text{ and } \varrho<\delta, \tag{4.18}$$

where $c>0$. This is possible, since $b(0)\neq0$, provided $\delta$ is sufficiently small. We then choose $\chi_0,\chi\in C_0^\infty(\mathbb{R})$ such that $\operatorname{supp}\chi\subset\left(-2B,-\frac{1}{2}B\right)\cup\left(\frac{1}{2}B,2B\right)$ and

$$1=\chi_0(s)+\sum_{k=1}^\infty\chi\left(\frac{s}{2^k}\right):=\chi_0(s)+\sum_{k=1}^\infty\chi_k(s)\quad\text{for every } s\in\mathbb{R}.$$

Accordingly, we decompose $A_t^\varepsilon f=\sum_{k=0}^\infty A_t^{\varepsilon,k}f$, where

$$A_t^{\varepsilon,k}f(x):=\varrho^{1/m}\int_{\mathbb{R}}f(x_1-t(\varrho^{1/m}s+\eta),x_2-t(1+\varepsilon\varrho g(s)))\psi(\varrho^{1/m}s+\eta)\chi_k(s)\,ds.$$

Assume first that $k\geqslant1$. Then this can be rewritten as

$$A_t^{\varepsilon,k}f(x)=2^k\varrho^{1/m}\int_{\mathbb{R}}f(x_1-t(\varrho^{1/m}2^ks+\eta),x_2-t(1+\varepsilon\varrho2^{mk}g_k(s)))\psi(\varrho^{1/m}2^ks+\eta)\chi(s)\,ds,$$

where

$$g_k(s)=g_k(s,\varrho,\xi):=2^{-mk}g(2^k s,\varrho,\xi).$$

And, by (4.18),

$$|g_k''(s)|\geqslant c>0 \quad \text{for every } s\in\operatorname{supp}\chi,\ \xi\in\Sigma \text{ and } \varrho<\delta.$$

More precisely, since

$$g_k(s)=\tilde b(\varrho^{1/m}2^k s)s^m+\frac{\xi_2}{2^{(m-2)k}}s^2+\ldots+\frac{\xi_{m-2}}{2^{2k}}s^{m-2},$$

where $|s|\sim B$, and where $\varrho^{1/m}2^k\leqslant\delta$, unless $A_t^{\varepsilon,k}=0$, if we choose $\operatorname{supp}\psi$ sufficiently close to 0, we see that $g_k(s)$ is a small $\delta$-perturbation of $g_k(s,0,\xi)$.

Moreover, covering $\Sigma$ by a finite number of $\delta$-neighborhoods $\Sigma_j$ of points $\xi^{(j)}\in\Sigma$, for every $\xi\in\Sigma_j$ we may regard $g_k(s,0,\xi)$ as a $\delta$-perturbation of $g_k(s,0,\xi^{(j)})$. Thus, for $\xi\in\Sigma_j$, Proposition 4.1 can be applied for $m=2$, in a similar way as in our discussion of the case $k=2$, in order to estimate the maximal operator

$$\mathcal{M}^{\varepsilon,k}f(x)=\sup_{t>0}|A_t^{\varepsilon,k}f(x)|$$

by

$$\begin{aligned}\|\mathcal{M}^{\varepsilon,k}f\|_p&\leqslant C_p'2^k\varrho^{1/m}(|\eta|(2^k\varrho^{1/m})^{-1}+(\varepsilon\varrho2^{mk})^{-1}+1)^{1/p}\|f\|_p\\&\leqslant C_p((2^k\varrho^{1/m})^{1-1/p}+\varepsilon^{-1/p}(2^k\varrho^{1/m})^{1-m/p})\|f\|_p.\end{aligned}$$

Since $\mathcal{M}^{\varepsilon,k}=0$ if $2^k\varrho^{1/m}>\delta$, we then obtain, for $p>m$,

$$\sum_{k\geqslant1}\|\mathcal{M}^{\varepsilon,k}f\|_p=\sum_{\substack{k\geqslant1\\2^k\varrho^{1/m}\leqslant\delta}}\|\mathcal{M}^{\varepsilon,k}f\|_p\leqslant C_p\varepsilon^{-1/p}\|f\|_p.$$

There remains the operator $\mathcal{M}^{\varepsilon,0}$ to be considered. Conjugating $A_t^{\varepsilon,0}$ with the scaling operator

$$T_\varrho f(x_1,x_2):=\varrho^{-1/mp}f(\varrho^{-1/m}x_1,x_2),$$

which acts isometrically on $L^p(\mathbb{R}^2)$, we can reduce our considerations to the averaging operator

$$T_\varrho^{-1}A_t^{\varepsilon,0}T_\varrho f(x):=\varrho^{1/m}\int_{\mathbb{R}}f(x_1-t(s+\varrho^{-1/m}\eta),x_2-t(1+\varepsilon\varrho g(s)))\psi(\varrho^{1/m}s+\eta)\chi_0(s)\,ds.$$

Fixing again $\xi^0\in\Sigma$, for $\xi$ in a $\delta$-neighborhood $\Sigma_0$ of $\xi^0$, we can consider $g(s,\varrho,\xi)$ as a $\delta$-perturbation of the polynomial function

$$P(s):=g(s,0,\xi^0)=\tilde{b}(0)s^m+\xi_2^0s^2+...+\xi_{m-2}^0s^{m-2}.$$

Since there is no term $\xi_{m-1}^0s^{m-1}$ in $P(s)$, and since $\xi^0\neq 0$, it follows that for every $s_0$ one has

$$\sum_{j=2}^{m-1}\left|\left(\frac{\partial}{\partial s}\right)^j g(s_0,0,\xi^0)\right|\neq 0,$$

for otherwise we would have

$$P(s)-P(s_0)-P'(s_0)(s-s_0)=\tilde{b}(0)(s-s_0)^m=\tilde{b}(0)(s^m-ms_0s^{m-1}+...\,).$$

Hence $s_0=0$, and so $\xi^0=0$.

We can thus apply our induction hypothesis, and obtain, for $p>m-1$,

$$\|T_\varrho^{-1}\mathcal{M}^{\varepsilon,0}T_\varrho f\|_p\leqslant C_p\varrho^{1/m}(\varrho^{-1/m}|\eta|+(\varepsilon\varrho)^{-1})^{1/p}\|f\|_p,$$

and hence

$$\|\mathcal{M}^{\varepsilon,0}f\|_p\leqslant C_p\varepsilon^{-1/p}\varrho^{1/m-1/p}\|f\|_p,$$

first for $\xi\in\Sigma_0$, and then, by covering $\Sigma$ again by a finite number of $\delta$-neighborhoods of points $\xi_j$, for every $\xi\in\Sigma$. In particular, for $p>m$ we get the uniform estimate

$$\|\mathcal{M}^{\varepsilon,0}f\|_p\leqslant C_p\varepsilon^{-1/p}\|f\|_p,$$

which concludes the proof of Theorem 4.2. □

In the next subsection, we shall need a slight generalization of this theorem, namely for averaging operators of the form

$$A_t^{\varepsilon,\sigma_1}f(x):=\int_{\mathbb{R}}f(x_1-t(s+\sigma_1),x_2-t(1+\varepsilon(a(s)+r(s))))\psi(s)\,ds,\quad f\in\mathcal{S}(\mathbb{R}^2),$$

where $\sigma_1$ is a second real parameter which can be arbitrarily large. The corresponding maximal operator

$$\mathcal{M}^{\varepsilon,\sigma_1}f(x):=\sup_{t>0}|A_t^{\varepsilon,\sigma_1}f(x)|$$

can be estimated exactly as before, if we simply replace the shift term $\eta$ in the proof of Theorem 4.2 by $\eta+\sigma_1$, and one easily obtains the following consequence.

Corollary 4.4. *Let $a$ be a function of polynomial type $m\geqslant 2$. Then there exist numbers $M\in\mathbb{N}$ and $\delta>0$ such that, for every $r\in C^\infty(I,\mathbb{R})$ with $\|r\|_{C^M}<\delta$, $0<\varepsilon\ll 1$ and $p>m$, the following a priori estimate is satisfied*:

$$\|\mathcal{M}^{\varepsilon,\sigma_1}f\|_p\leqslant C_p(|\sigma_1|+\varepsilon^{-1})^{1/p}\|f\|_p,\quad f\in\mathcal{S}(\mathbb{R}^2),$$

*with a constant $C_p$ depending only on $p$.*

### 4.2. Related results for families of surfaces

By decomposing a given surface in $\mathbb{R}^3$ by means of a "fan" of hyperplanes into a family of curves, we can easily derive suitable estimates for certain families of surfaces from the maximal estimates in the previous subsection.

Let $U$ be an open neighborhood of the point $x^0 \in \mathbb{R}^2$, and let $\phi_{\mathrm{pr}} \in C^\infty(U,\mathbb{R})$ be such that

$$\partial_2^m \phi_{\mathrm{pr}}(x_1^0, x_2^0) \neq 0, \tag{4.19}$$

where $m \geqslant 2$. Let

$$\phi = \phi_{\mathrm{pr}} + \phi_r,$$

where $\phi_r \in C^\infty(U,\mathbb{R})$ is sufficiently small. Denote by $S_\varepsilon$ the surface in $\mathbb{R}^3$ given by

$$S_\varepsilon := \{(x_1, x_2, 1+\varepsilon\phi(x_1,x_2)) : (x_1,x_2) \in U\},$$

with $\varepsilon>0$, and consider the averaging operators

$$A_t f(x) = A_t^\varepsilon f(x) := \int_{S_\varepsilon} f(x-ty)\psi(y)\, d\sigma(y),$$

where $d\sigma$ denotes the surface measure and $\psi \in C_0^\infty(S_\varepsilon)$ is a non-negative cut-off function. Define the associated maximal operator by

$$\mathcal{M}^\varepsilon f(x) := \sup_{t>0} |A_t^\varepsilon f(x)|.$$

PROPOSITION 4.5. *Assume that $\phi_{\mathrm{pr}}$ satisfies* (4.19) *and that the neighborhood $U$ of the point $x^0$ is sufficiently small. Then there exist numbers $M \in \mathbb{N}$ and $\delta>0$, such that for every $\phi_r \in C^\infty(U,\mathbb{R})$ with $\|\phi_r\|_{C^M} < \delta$ and any $p>m$ there exists a positive constant $C_p$ such that for $\varepsilon>0$ sufficiently small the maximal operator $\mathcal{M}^\varepsilon$ satisfies the following a priori estimate*:

$$\|\mathcal{M}^\varepsilon f\|_p \leqslant C_p \varepsilon^{-1/p} \|f\|_p, \quad f \in \mathcal{S}(\mathbb{R}^3). \tag{4.20}$$

*Proof.* Let us write the averaging operator $A_t$ in the form

$$A_t f(y) = \int_{\mathbb{R}^2} f(y_1 - tx_1, y_2 - tx_2, y_3 - t(1+\varepsilon\phi(x_1,x_2)))\eta(x_1,x_2)\, dx, \tag{4.21}$$

where $\eta \in C_0^\infty(U)$.

Choose $\theta_0$ such that $\sin\theta_0 + x_1^0 \cos\theta_0 = 0$ (we may assume that $\cos\theta_0 > 0$). For small $\theta$, consider the equation

$$(1+\varepsilon\phi(x_1,x_2))\sin(\theta_0+\theta) + x_1\cos(\theta_0+\theta) = 0 \tag{4.22}$$

with respect to $x_1$. By the implicit function theorem, the last equation has a unique smooth solution $x_1(\theta,x_2,\varepsilon)$ for $|\theta|$, $|x_2-x_2^0|$ and $\varepsilon$ sufficiently small such that

$$x_1(0,x_2^0,0)=x_1^0.$$

Moreover,

$$\frac{\partial}{\partial\theta}x_1(0,x_2^0,0)\neq 0.$$

In the integral (4.21) we can thus use the change of variables $(\theta,x_2)\mapsto(x_1(\theta,x_2,\varepsilon),x_2)$ (assuming $U$ to be sufficiently small) and obtain

$$A_tf(y)=\int_{\mathbb{R}^2} f(y_1-tx_1(\theta,x_2,\varepsilon),y_2-tx_2,y_3-t(1+\varepsilon\phi(x_1(\theta,x_2,\varepsilon),x_2)))\psi(\theta,x_2,\varepsilon)\,d\theta\,dx_2, \tag{4.23}$$

where $\psi(\theta,x_2,\varepsilon):=\eta(x_1(\theta,x_2,\varepsilon),x_2)|J(\theta,x_2,\varepsilon)|$ and $J(\theta,x_2,\varepsilon)$ denotes the Jacobian of this change of coordinates. Let us write the integral (4.23) as an iterated integral

$$A_tf(y)=\int_{-b}^{b}A_t^\theta f(y)\,d\theta,$$

where $b$ is some positive number and $A_t^\theta$ denotes the following averaging operator along a curve:

$$A_t^\theta f(y):=\int_{\mathbb{R}^2} f(y_1-tx_1(\theta,s,\varepsilon),y_2-ts,y_3-t(1+\varepsilon\phi(x_1(\theta,s,\varepsilon),s)))\psi(\theta,s,\varepsilon)\,ds.$$

Now, we define the rotation operator

$$R^\theta f(x):=f(x_1\sin(\theta_0+\theta)-x_3\cos(\theta_0+\theta),x_2,x_1\cos(\theta_0+\theta)+x_3\sin(\theta_0+\theta)),$$

which acts isometrically on every $L^p(\mathbb{R}^3)$. Then we have

$$R^{-\theta}A_t^\theta R^\theta f(y)=\int_{\mathbb{R}^2} f\biggl(y_1+t\frac{1}{\cos(\theta_0+\theta)}(1+\varepsilon\phi(x_1(\theta,s,\varepsilon),s),y_2-ts,y_3)\biggr)\psi(\theta,s,\varepsilon)\,ds.$$

Observe that the last operator "acts" only on the first two variables. Moreover, for $\varepsilon=0$, by (4.22), we have $x_1(\theta,x_2,0)=-\tan(\theta_0+\theta)$, which is independent of $x_2$. This implies that

$$\frac{d^m}{ds^m}\phi_{\mathrm{pr}}(x_1(0,s,0),s)\bigg|_{s=x_2^0}=\partial_2^m\phi_{\mathrm{pr}}(x_1^0,x_2^0)\neq 0.$$

Notice also that for $\varepsilon$, $\delta$ and $U$ (and hence also $\theta$) sufficiently small, $\phi(x_1(\theta,s,\varepsilon),s)$ can be regarded as a small perturbation of $\phi_{\mathrm{pr}}(x_1(0,s,0),s)$. Therefore we can apply Theorem 4.2 (in the first two variables) and obtain that, for $p>m$,

$$\Bigl\|\sup_{t>0}|R^{-\theta}A_t^\theta R^\theta f|\Bigr\|_p\leqslant C_p\varepsilon^{-1/p}\|f\|_p,$$

and hence

$$\Big\|\sup_{t>0}|A_t^\theta f|\Big\|_p \leqslant C_p\varepsilon^{-1/p}\|f\|_p,$$

where $C_p$ is independent of $\theta$ and $\varepsilon$. Integrating finally in the $\theta$ variable we obtain the required estimate. □

In our later applications of this proposition, we shall also have to deal with functions $\phi$ which depend in fact also on the parameter $\varepsilon$ in such a way that they blow up as $\varepsilon\to 0$, however, in a particular way. More precisely, assume that $\tilde{\phi}=\tilde{\phi}_{\mathrm{pr}}+\tilde{\phi}_r$ has the same properties as $\phi$ in the proposition, so that in particular (4.19) is satisfied by $\tilde{\phi}$. We assume for simplicity that $\tilde{\phi}$ is defined on $\mathbb{R}^2$ and supported in the neighborhood $V$ of the point $x^0$. Let further $\psi_\varepsilon\in C^\infty(V_1)$ be a smooth function depending on the parameter $\varepsilon$ so that there is some $0\leqslant\delta<1$ such that

$$\psi_\varepsilon = O(\varepsilon^{-\delta}) \quad \text{in } C^\infty, \tag{4.24}$$

in the sense that $\|\psi_\varepsilon\|_{C^m(V_1)}=O(\varepsilon^{-\delta})$ for every $m\in\mathbb{N}$, where $V_1$ denotes the orthogonal projection of the neighborhood $V$ onto the $x_1$-axis. Put then

$$\phi_\varepsilon(x_1,x_2) := \tilde{\phi}(x_1, x_2-\psi_\varepsilon(x_1)). \tag{4.25}$$

Notice that

$$|\partial_1^j\partial_2^k\phi_\varepsilon(x)| = O(\varepsilon^{-j\delta}). \tag{4.26}$$

This means that we cannot directly apply Proposition 4.5 to $\phi_\varepsilon$. We shall see that nevertheless the proof of this proposition can be extended to $\phi_\varepsilon$. To this end, observe first that $|\nabla(\varepsilon\phi_\varepsilon)(x)|\leqslant C\varepsilon^{1-\delta}$, uniformly in $x$. Therefore, again by the implicit function theorem, we can solve the equation

$$(1+\varepsilon\phi_\varepsilon(x_1,x_2))\sin(\theta_0+\theta)+x_1\cos(\theta_0+\theta)=0$$

in $x_1$ near the point $(x_1^0, x_2^0+\psi_\varepsilon(x_1^0))$, and obtain a smooth solution $x_1(\theta,x_2,\varepsilon)$ for sufficiently small values of $|\theta|$, $|x_2-(x_2^0+\psi_\varepsilon(x_1^0))|$ and $\varepsilon>0$, satisfying

$$x_1(0,x_2^0+\psi_\varepsilon(x_1^0),0)=x_1^0.$$

Let us also define $x_1^0(\theta)$ as the solution of the equation

$$\sin(\theta_0+\theta)+x_1^0(\theta)\cos(\theta_0+\theta)=0,$$

and put $g(\theta,x_2,\varepsilon):=x_1(\theta,x_2,\varepsilon)-x_1^0(\theta)$. Then $g$ satisfies the equation

$$\varepsilon\phi_\varepsilon(x_1^0(\theta)+g(\theta,x_2,\varepsilon),x_2)\sin(\theta_0+\theta)+g(\theta,x_2,\varepsilon)\cos(\theta_0+\theta)=0.$$

Implicit differentiation shows that

$$g'_\varepsilon(x_2) = -\varepsilon \frac{\partial_2\phi_\varepsilon(x_1^0(\theta)+g_\varepsilon(x_2), x_2)\sin(\theta_0+\theta)}{\cos(\theta_0+\theta)+\varepsilon\partial_1\phi_\varepsilon(x_1^0(\theta)+g_\varepsilon(x_2), x_2)\sin(\theta_0+\theta)},$$

using the short-hand notation $g_\varepsilon(x_2)=g(\theta, x_2, \varepsilon)$. By (4.26), this implies that $|g'_\varepsilon(x_2)|=O(\varepsilon)$, and similarly $|g_\varepsilon^{(j)}(x_2)|=O(\varepsilon)$ for every $j\geqslant 1$, uniformly in $x_2$. But clearly this estimate is also true for $j=0$, so that

$$g_\varepsilon = O(\varepsilon) \quad \text{in } C^\infty. \tag{4.27}$$

If put

$$\Phi_\varepsilon(\theta, s) := \phi_\varepsilon(x_1^0(\theta)+g_\varepsilon(s), s),$$

then (4.26) and (4.27) show that $\Phi_\varepsilon(\theta, \cdot)=O(1)$ in $C^\infty$. The averaging operators associated with $\phi_\varepsilon$ will be of the form

$$A_t f(y) := \int_{\mathbb{R}^2} f(y_1-tx_1, y_2-tx_2, y_3-t(1+\varepsilon\phi_\varepsilon(x_1, x_2)))\eta(x_1, x_2)\, dx, \tag{4.28}$$

where $\eta(x_1, x_2)=\widetilde{\eta}(x_1, x_2-\psi_\varepsilon(x_1))$, with $\widetilde{\eta}\in C_0^\infty(\mathbb{R}^2)$ supported in a sufficiently small neighborhood $\widetilde{U}\subset V$ of $x^0$. The corresponding operators $R^{-\theta}A_t^\theta R^\theta$ are then given by

$$R^{-\theta}A_t^\theta R^\theta f(y) = \int_{\mathbb{R}^2} f\biggl(y_1+t\frac{1+\varepsilon\Phi_\varepsilon(\theta, s)}{\cos(\theta_0+\theta)}, y_2-ts, y_3\biggr)a(\theta, s, \varepsilon)\, ds,$$

with

$$a(\theta, s, \varepsilon) := \eta(x_1(\theta, s, \varepsilon), s)|J(\theta, s, \varepsilon)| = \tilde{\eta}(x_1^0(\theta)+g_\varepsilon(s), s-\psi_\varepsilon(x_1^0(\theta)+g_\varepsilon(s)))|J(\theta, s, \varepsilon)|.$$

The subsitution $s\mapsto s+\psi_\varepsilon(x_1^0(\theta))$ in the integral thus leads to

$$R^{-\theta}A_t^\theta R^\theta f(y) = \int_{\mathbb{R}^2} f\biggl(y_1+t\frac{1+\varepsilon\tilde{\Phi}_\varepsilon(\theta, s)}{\cos(\theta_0+\theta)}, y_2-t(s+\psi_\varepsilon(x_1^0(\theta))), y_3\biggr)\tilde{a}(\theta, s, \varepsilon)\, ds,$$

with $\tilde{\Phi}_\varepsilon(\theta, s):=\tilde{\phi}(x_1^0(\theta)+\tilde{g}_\varepsilon(s), s+\psi_\varepsilon(x_1^0(\theta))-\psi_\varepsilon(x_1^0(\theta)+\tilde{g}_\varepsilon(s)))$ and

$$\tilde{a}(\theta, s, \varepsilon) := \tilde{\eta}(x_1^0(\theta)+\tilde{g}_\varepsilon(s), s+\psi_\varepsilon(x_1^0(\theta))-\psi_\varepsilon(x_1^0(\theta)+\tilde{g}_\varepsilon(s)))|J(\theta, s+x_1^0(\theta), \varepsilon)|,$$

where we have set $\tilde{g}_\varepsilon(s):=g_\varepsilon(s+\psi_\varepsilon(x_1^0(\theta)))$. From (4.24) and (4.27), we have $\tilde{g}_\varepsilon=O(\varepsilon)$ in $C^\infty$ and $\psi_\varepsilon(x_1^0(\theta))-\psi_\varepsilon(x_1^0(\theta)+\tilde{g}_\varepsilon(s))=O(\varepsilon^{1-\delta})$ in $C^\infty$.

Consequently, $\tilde{a}$ is supported in $\widetilde{V}_1$, if $\varepsilon$ and $\theta$ are sufficiently small, and $\tilde{a}=O(1)$ in $C^\infty$. In a similar way, we see that

$$\tilde{\Phi}_\varepsilon(\theta, s) = \tilde{\phi}(x_1^0(\theta), s)+\tilde{\phi}_r(\theta, s, \varepsilon),$$

where the perturbation term $\tilde{\phi}_r(\theta, s, \varepsilon)$ can be made small in $C^\infty$ by choosing $\varepsilon$ and $\theta$ sufficiently small. Notice finally that for $\varepsilon<1$,

$$|\psi_\varepsilon(x_1^0(\theta))| \lesssim \varepsilon^{-\delta} \leqslant \varepsilon^{-1}.$$

We can therefore apply the maximal theorem for curves, Corollary 4.4, to each operator $R^{-\theta} A_t^\theta R^\theta$ and obtain the following consequence.

COROLLARY 4.6. *Let $V$ be an open neighborhood of $x^0 \in \mathbb{R}^2$, and let $\tilde{\phi}_{\mathrm{pr}} \in C^\infty(V, \mathbb{R})$ be such that*

$$\partial_2^m \tilde{\phi}_{\mathrm{pr}}(x_1^0, x_2^0) \neq 0,$$

*where $m \geqslant 2$. Let*

$$\tilde{\phi} := \tilde{\phi}_{\mathrm{pr}} + \tilde{\phi}_r,$$

*where $\tilde{\phi}_r \in C^\infty(V, \mathbb{R})$ is sufficiently small, and assume that $\psi_\varepsilon \in C^\infty(V_1)$ satisfies* (4.24) *for some $0 \leqslant \delta < 1$. Put $\phi_\varepsilon(x_1, x_2) := \tilde{\phi}(x_1, x_2 - \psi_\varepsilon(x_1))$ and $\eta(x_1, x_2) = \tilde{\eta}(x_1, x_2 - \psi_\varepsilon(x_1))$, with $\tilde{\eta} \in C_0^\infty(\mathbb{R}^2)$ supported in a sufficiently small neighborhood $\widetilde{U} \subset V$ of $x^0$, and consider the averaging operators $A_t$ given by* (4.28), *with associated maximal operator $\mathcal{M}^\varepsilon$.*

*Assume that the neighborhood $\widetilde{U}$ of the point $x^0$ is sufficiently small. Then there exist numbers $M \in \mathbb{N}$ and $\delta_1 > 0$ such that for every $\tilde{\phi}_r \in C^\infty(\widetilde{U}, \mathbb{R})$ with $\|\phi_r\|_{C^M} < \delta_1$ and any $p > m$ there exists a positive constant $C_p$ such that for $\varepsilon > 0$ sufficiently small the maximal operator $\mathcal{M}^\varepsilon$ satisfies the following a priori estimate*:

$$\|\mathcal{M}^\varepsilon f\|_p \leqslant C_p \varepsilon^{-1/p} \|f\|_p, \quad f \in \mathcal{S}(\mathbb{R}^3). \tag{4.29}$$

## 5. Estimates for oscillatory integrals with small parameters

In this section, we shall provide the crucial estimates for oscillatory integrals that will be needed in the subsequent sections. More precisely, we shall study oscillatory integrals

$$J(\lambda, \sigma, \delta) := \int_{\mathbb{R}^2} e^{i\lambda F(x,\sigma,\delta)} \psi(x, \delta)\, dx, \quad \lambda > 0,$$

with a phase function $F$ of the form

$$F(x_1, x_2, \sigma, \delta) := f_1(x_1, \delta) + \sigma f_2(x_1, x_2, \delta),$$

and an amplitude $\psi$ defined for $x$ in some open neighborhood of the origin in $\mathbb{R}^2$ with compact support in $x$. The functions $f_1$ and $f_2$ are assumed to be real-valued and will depend, like the function $\psi$, smoothly on $x$ and on small real parameters $\delta_1, ..., \delta_\nu$, which form the vector $\delta := (\delta_1, ..., \delta_\nu) \in \mathbb{R}^\nu$. Here $\sigma$ denotes a small real parameter.

With a slight abuse of language, we say that $\psi$ is *compactly supported in some open set* $U\subset\mathbb{R}^2$ if there is a compact subset $K\subset U$ such that $\operatorname{supp}\psi(\,\cdot\,,\delta)\subset K$ for every $\delta$.

Let us first recall the following lemma, which is a (not quite straightforward) consequence of van der Corput's lemma (cf. [39]) and whose formulation goes back to J. E. Björk (see [10]) and also G. I. Arkhipov [1].

LEMMA 5.1. *Assume that $f$ is a smooth real-valued function defined on an interval $I\subset\mathbb{R}$ which is of polynomial type $m\geqslant 2$, $m\in\mathbb{N}$, i.e., there are positive constants $c_1, c_2>0$ such that*

$$c_1 \leqslant \sum_{j=2}^{m} |f^{(j)}(s)| \leqslant c_2 \quad \textit{for every } s\in I.$$

*Then, for $\lambda\in\mathbb{R}$,*

$$\left|\int_I e^{i\lambda f(s)} g(s)\,ds\right| \leqslant C\|g\|_{C^1(I)}(1+|\lambda|)^{-1/m},$$

*where the constant $C$ depends only on the constants $c_1$ and $c_2$.*

### 5.1. Oscillatory integrals with non-degenerate critical points in $x_1$

PROPOSITION 5.2. *Assume that*

$$|\partial_1 f_1(0,0)|+|\partial_1^2 f_1(0,0)| \neq 0,$$

*and that there is some $m\geqslant 2$ such that*

$$\partial_2^m f_2(0,0,0)\neq 0.$$

*Then there exist a neighborhood $U\subset\mathbb{R}^2$ of the origin and some $\varepsilon>0$ such that for any $\psi$ which is compactly supported in $U$ the estimate*

$$|J(\lambda,\sigma,\delta)| \leqslant \frac{C\|\psi(\,\cdot\,,\delta)\|_{C^3}}{(1+\lambda)^{1/2}(1+|\lambda\sigma|)^{1/m}} \tag{5.1}$$

*holds true uniformly for $|\sigma|+|\delta|<\varepsilon$.*

*Proof.* If $\partial_1 f_1(0,0)\neq 0$, then we can integrate by parts in $x_1$ if $\lambda>1$, and obtain the stronger estimate

$$|J(\lambda,\sigma,\delta)| \leqslant \frac{C\|\psi(\,\cdot\,,\delta)\|_{C^1}}{1+\lambda}.$$

Assume therefore that $\partial_1 f_1(0,0)=0$, so that the mapping $x_1\mapsto f_1(x_1,0)$ has a non-degenerate critical point at $x_1=0$. Then, by the implicit function theorem, for $|\delta|$ sufficiently small there exists a unique critical point $x_1=x_1^0(\delta)$ depending smoothly on $\delta$ of the mapping $\xi\mapsto f_1(x_1,\delta)=0$, i.e., $\partial_1 f_1(x_1^0(\delta),\delta)\equiv 0$, where $x_1^0(0)=0$.

In a similar way, we see that there is a unique, smooth function $x_1^c(x_2,\sigma,\delta)$ for $|x_2|+|\sigma|+|\delta|$ sufficiently small such that

$$\partial_1 F(x_1^c(x_2,\sigma,\delta),x_2,\sigma,\delta)\equiv 0,$$

where $x_1^c(0,0,0)=0$. By comparison, we see that $x_1^c(x_2,0,\delta)=x_1^0(\delta)$, so that

$$x_1^c(x_2,\sigma,\delta)=x_1^0(\delta)+\sigma\gamma(x_2,\sigma,\delta)$$

for some smooth function $\gamma$. Applying the stationary phase formula with parameters to the integration in $x_1$, we thus obtain

$$J(\lambda,\sigma,\delta)=\int_{\mathbb{R}} e^{i\lambda\phi(x_2,\sigma,\delta)}a(\lambda,x_2,\sigma,\delta)\,dx_2, \tag{5.2}$$

where

$$\phi(x_2,\sigma,\delta):=F(x_1^0(\delta)+\sigma\gamma(x_2,\delta,\sigma),x_2,\sigma,\delta)$$

and where $a(\lambda,x_2,\sigma,\delta)$ is a symbol of order $-\frac{1}{2}$ in $\lambda$, so that in particular

$$|\partial_{x_2}^l a(\lambda,x_2,\sigma,\delta)|\leqslant C_l(1+|\lambda|)^{-1/2}, \tag{5.3}$$

with constants $C_l$ independent of $x_2$, $\sigma$ and $\delta$ (see, e.g., Sogge [35] or Hörmander [17]).

Moreover, a Taylor series expansion of $\phi$ with respect to $\sigma$ near $\sigma=0$ shows that

$$\phi(x_2,\sigma,\delta)=f_1(x_1^0(\delta),\delta)+\sigma(f_2(x_1^0(\delta),x_2,0,\delta)+O(\sigma))$$

in $C^\infty$. Since $\partial_2^m f_2(0,0,0)\neq 0$, for $|\sigma|$ sufficiently small we can thus apply van der Corput's lemma (or Lemma 5.1) to the integral (5.2) in $x_2$ and obtain the estimate (5.1). □

### 5.2. Oscillatory integrals of non-degenerate Airy type

PROPOSITION 5.3. *Assume that*

$$\partial_1^3 f_1(0,0)\neq 0 \quad \textit{and} \quad \partial_2^2 f_2(0,0,0)\neq 0.$$

*Then there exist a neighborhood $U\subset\mathbb{R}^2$ of the origin and some $\varepsilon>0$ such that for any $\psi$ which is compactly supported in $U$, the estimate*

$$|J(\lambda,\sigma,\delta)|\leqslant \frac{C\|\psi(\cdot,\delta)\|_{C^3}}{(1+\lambda)^{1/3}(1+|\lambda\sigma|)^{1/2}} \tag{5.4}$$

*holds true uniformly for $|\sigma|+|\delta|<\varepsilon$.*

*Proof.* Consider first the case where $\partial_2 f_2(0,0,0)\neq 0$. Then, if $|\lambda\sigma|\gg 1$, we first perform an integration by parts in $x_2$. Subsequently, we can apply van der Corput's lemma to the integration in $x_1$, provided $U$ and $\varepsilon$ are chosen sufficiently small, and obtain the stronger estimate

$$|J(\lambda,\sigma,\delta)|\leqslant \frac{C\|\psi(\,\cdot\,,\delta)\|_{C^2}}{(1+\lambda)^{1/3}(1+|\lambda\sigma|)}.$$

Now, assume that $\partial_2 f_2(0,0,0)=0$ but $\partial_2^2 f_2(0,0,0)\neq 0$. Then for $U$ and $\varepsilon$ chosen sufficiently small, by the implicit function theorem there exists a unique critical point $x_2^c(x_1,\delta)$ of the function $x_2\mapsto f_2(x_1,x_2,\delta)$. Then, by applying the stationary phase method with small parameters to the $x_2$-integration, we see that

$$J(\lambda,\sigma,\delta)=\int_{\mathbb{R}} e^{i\lambda\phi(x_1,\sigma,\delta)}a(\lambda\sigma,x_1,\delta)\,dx_1, \tag{5.5}$$

where

$$\phi(x_1,\sigma,\delta):=f_1(x_1,\delta)+\sigma f_2(x_1,x_2^c(x_1,\delta),\delta),$$

and where $a(\lambda,x_1,\delta)$ is a symbol of order $-\frac{1}{2}$ in $\lambda$, so that in particular

$$|\partial_{x_1}^l a(\lambda\sigma,x_1,\delta)|\leqslant C_l(1+|\lambda\sigma|)^{-1/2}, \tag{5.6}$$

with constants $C_l$ which are independent of $x_1$ and $\delta$.

We can now apply van der Corput's lemma to the integral (5.5) and obtain, in view of (5.6), the desired estimate (5.4). $\square$

### 5.3. Oscillatory integrals of degenerate Airy type

THEOREM 5.4. *Assume that*

$$|\partial_1 f_1(0,0)|+|\partial_1^2 f_1(0,0)|+|\partial_1^3 f_1(0,0)|\neq 0 \quad \textit{and} \quad \partial_1\partial_2 f_2(0,0,0)\neq 0, \tag{5.7}$$

*and that there is some $m\geqslant 2$ such that*

$$\partial_2^l f_2(0,0,0)=0 \ \ \textit{for}\ \ l=1,...,m-1 \quad \textit{and} \quad \partial_2^m f_2(0,0,0)\neq 0. \tag{5.8}$$

*Then there exist a neighborhood $U\subset\mathbb{R}^2$ of the origin and constants $\varepsilon,\varepsilon'>0$ such that for any $\psi$ which is compactly supported in $U$ the estimate*

$$|J(\lambda,\sigma,\delta)|\leqslant \frac{C\|\psi(\,\cdot\,,\delta)\|_{C^3}}{\lambda^{1/2+\varepsilon}|\sigma|^{l_m+c_m\varepsilon}} \tag{5.9}$$

*holds true uniformly for $|\sigma|+|\delta|<\varepsilon'$, where*

$$l_m:=\begin{cases}\frac{1}{6}, & \textit{for } m<6,\\ \dfrac{m-3}{2(2m-3)}, & \textit{for } m\geqslant 6\end{cases} \quad \textit{and} \quad c_m:=\begin{cases}1, & \textit{for } m<6,\\ 2, & \textit{for } m\geqslant 6.\end{cases}$$

*Remark* 5.5. If $|\partial_1 f_1(0,0)|+|\partial_1^2 f_1(0,0)|\neq 0$, then a stronger estimate than (5.9) follows from Proposition 5.2, since $\frac{1}{6}\leqslant l_m<\frac{1}{4}$. The full thrust of Theorem 5.4 therefore lies in the case where $\partial_1 f_1(0,0)=\partial_1^2 f_1(0,0)=0$ and $\partial_1^3 f_1(0,0)\neq 0$, on which we shall concentrate in the sequel.

The proof of Theorem 5.4 will be an immediate consequence of the following two lemmas. Our first lemma allows us to reduce the phase function $F$ to some normal form and is based on Theorem 7.5.13 in [17]. The latter result, which generalizes earlier work by C. Chester, B. Friedman, F. Ursell [7] and N. Levinson [26] in the analytic case, is derived by L. Hörmander by means of the Malgrange preparation theorem.

LEMMA 5.6. *Assume that $F$ satisfies the conditions of Theorem* 5.4, *and in addition that $\partial_1 f_1(0,0)=\partial_1^2 f_1(0,0)=0$. Then there exist smooth functions $X_1=X_1(x_1,\sigma,\delta)$ and $X_2=X_2(x_1,x_2,\delta)$ defined in a sufficiently small neighborhood $U\times V\subset\mathbb{R}^2\times\mathbb{R}^{\nu+1}$ of the origin such that the following hold true*:

(i) $X_1(0,0,0)=X_2(0,0,0)=0$, $\partial_1 X_1(0,0,0)\neq 0$ *and* $\partial_2 X_2(0,0,0)\neq 0$, *so that we can change coordinates from* $(x_1,x_2,\sigma,\delta)$ *to* $(X_1,X_2,\sigma,\delta)$ *near the origin*;

(ii) *In the new coordinates $X_1$ and $X_2$ for $\mathbb{R}^2$ near the origin*, *we can write*

$$F(x_1,x_2,\sigma,\delta)=g_1(X_1,\sigma,\delta)+\sigma g_2(X_1,X_2,\sigma,\delta),$$

*with*

$$g_1(X_1,\sigma,\delta)=X_1^3+a_1(\sigma,\delta)X_1+a_m(\sigma,\delta)$$

*and*

$$g_2(X_1,X_2,\sigma,\delta)=X_2^m+\sum_{j=2}^{m-2}a_j(\delta)X_2^{m-j}+(X_1-a_{m-1}(\sigma,\delta))X_2 b(X_1,X_2,\sigma,\delta),$$

*if $m\geqslant 3$, and $g_2(X_1,X_2,\sigma,\delta)=X_2^2$, if $m=2$, where $a_1,...,a_m$ are smooth functions of the variables $\sigma$ and $\delta$ such that $a_l(0,0)=0$, and where $b$ is a smooth function such that $b(0,0,0,0)\neq 0$.*

*Proof.* In a first step, we apply Theorem 7.5.13 in [17] to the function $f_2(x_1,x_2,\delta)$. Due to our assumption (5.8), there exists a smooth function $X_2=X_2(x_1,x_2,\delta)$ defined in a sufficiently small neighborhood of the origin with

$$X_2(0,0,0)=0 \quad\text{and}\quad \partial_2 X_2(0,0,0)\neq 0,$$

such that in the new coordinate $X_2$ for $\mathbb{R}$ near the origin $f_2$ assumes the form

$$f_2(x_1,x_2,\delta)=X_2^m+\tilde a_2(x_1,\delta)X_2^{m-2}+...+\tilde a_{m-1}(x_1,\delta)X_2+\tilde a_m(x_1,\delta),$$

where $\tilde{a}_2, ..., \tilde{a}_m$ are smooth functions satisfying $\tilde{a}_l(0,0)=0$, $l=2, ..., m-1$.

Notice that the case $m=2$ is special, since in this case

$$f_2(x_1, x_2, \delta) = X_2^2 + \tilde{a}_2(x_1, \delta)$$

contains no linear term in $X_2$.

If $m \geqslant 3$, then by assumption (5.7) we have

$$\frac{\partial \tilde{a}_{m-1}}{\partial x_1}(0,0) \neq 0, \tag{5.10}$$

since

$$0 \neq \frac{\partial^2 f_2}{\partial x_1 \partial x_2}(0,0,0) = \frac{\partial \tilde{a}_{m-1}}{\partial x_1}(0,0) \frac{\partial X_2}{\partial x_2}(0,0,0).$$

Consequently, any smooth function $\varphi=\varphi(x_1, \delta)$ defined in a sufficiently small neighborhood of the origin can be written in the form

$$\varphi(x_1, \delta) = \eta(x_1, \delta)\tilde{a}_{m-1}(x_1, \delta) + \widetilde{\varphi}(\delta),$$

with smooth functions $\eta=\eta(x_1, \delta)$ and $\widetilde{\varphi}(\delta)$. Applying this observation to the functions $\tilde{a}_l$, we can write

$$\tilde{a}_l(x_1, \delta) = \tilde{a}_{m-1}(x_1, \delta) b_l(x_1, \delta) + a_l(\delta), \quad l = 2, ..., m-2,$$

with smooth functions $b_l(x_1, \delta)$ and $a_l(\delta)$, where $a_l(0)=0$.

We can accordingly rewrite the function $F=f_1+\sigma f_2$ in the form $F=\tilde{f}_1+\sigma \tilde{f}_2$, where

$$\begin{aligned} \tilde{f}_1(x_1, \sigma, \delta) &= f_1(x_1, \delta) + \sigma \tilde{a}_m(x_1, \delta), \\ \tilde{f}_2(x_1, x_2, \sigma, \delta) &= X_2^m + \tilde{a}_{m-1}(x_1, \delta) X_2 \tilde{b}(x_1, X_2, \delta) + a_2(\delta) X_2^{m-2} + ... + a_{m-2}(\delta) X_2^2, \end{aligned} \tag{5.11}$$

with

$$\tilde{b}(x_1, X_2, \delta) := 1 + b_{m-2}(x_1, \delta) X_2 + ... + b_2(x_1, \delta) X_2^{m-2}.$$

In particular, $\tilde{b}(0,0,0) \neq 0$.

In a second step, we apply Theorem 7.5.13 in [17] to the function $\tilde{f}_1(x_1, \sigma, \delta)$. Since $\partial_1 f_1(0,0,0)=\partial_1^2 f_1(0,0,0)=0$ and $\partial_1^3 f_1(0,0,0)\neq 0$, we then see that there exists a smooth function $X_1=X_1(x_1, \sigma, \delta)$ defined in a sufficiently small neighborhood of the origin with

$$X_1(0,0,0) = 0 \quad \text{and} \quad \partial_1 X_1(0,0,0) \neq 0,$$

such that in the new coordinate $X_1$ for $\mathbb{R}$ near the origin $\tilde{f}_1$ assumes the form

$$\tilde{f}_1(x_1, \sigma, \delta) = X_1^3 + a_1(\sigma, \delta) X_1 + a_m(\sigma, \delta),$$

where $a_1$ and $a_m$ are smooth functions such that $a_1(0,0)=a_m(0,0)=0$.

Let us write $\tilde{a}_{m-1}(x_1,\delta)=\alpha(X_1(x_1,\sigma,\delta),\sigma,\delta)$, so that $\alpha$ expresses $\tilde{a}_{m-1}$ in the new coordinate $X_1$. By (5.10) and the chain rule, we have

$$\alpha(0,0,0)=0 \quad \text{and} \quad \frac{\partial\alpha}{\partial X_1}(0,0,0)\neq 0.$$

This implies that there is a unique, smooth function $a_{m-1}(\sigma,\delta)$ with $a_{m-1}(0,0)=0$, such that $\alpha(a_{m-1}(\sigma,\delta),\sigma,\delta)\equiv 0$. Taylor's formula then implies that $\alpha(X_1,\sigma,\delta)$ can be written in the form

$$\alpha(X_1,\sigma,\delta)=(X_1-a_{m-1}(\sigma,\delta))\tilde{g}(X_1,\sigma,\delta),$$

where $\tilde{g}(X_1,\sigma,\delta)$ is a smooth function with $\tilde{g}(0,0,0)\neq 0$. This shows that

$$\tilde{a}_{m-1}(x_1,\delta)X_2\tilde{b}(x_1,X_2,\delta)=(X_1-a_{m-1}(\sigma,\delta))X_2\tilde{g}(X_1,\sigma,\delta)\tilde{b}(x_1,X_2,\delta).$$

When expressed in the new variables $(X_1,X_2)$, we see that in combination with (5.11) we obtain the form of $F$ as described in (ii). $\square$

After changing coordinates, the previous lemma allows us to reduce Theorem 5.4 to the estimation of 2-dimensional oscillatory integrals with phase functions of the form $F(x_1,x_2,\delta,\delta)=f_1(x_1,\delta)+\sigma f_2(x_1,x_2,\delta)$, where

$$\begin{aligned} f_1(x_1,\delta)&=x_1^3+\delta_1 x_1,\\ f_2(x_1,x_2,\delta)&=x_2^m+\sum_{j=2}^{m-2}\delta_j x_2^{m-j}+(x_1-\delta_{m-1})x_2 b(x_1,x_2,\sigma,\delta), \end{aligned} \tag{5.12}$$

if $m\geqslant 3$, and $f_2(x_1,x_2,\delta)=x_2^2$, if $m=2$. Here, $\sigma$ and $\delta_1,...,\delta_\nu$ are small real parameters (where $\nu\geqslant m-1$), the latter forming the vector $\delta:=(\delta_1,...,\delta_\nu)\in\mathbb{R}^\nu$, and $b=b(x_1,x_2,\sigma,\delta)$ is a smooth function defined on a neighborhood of the origin with $b(0,0,0,0)\neq 0$.

LEMMA 5.7. *Assume that the phase function $F$ is given by* (5.12). *Then there exist a neighborhood $U\subset\mathbb{R}^2$ of the origin and constants $\varepsilon,\varepsilon'>0$ such that for any $\psi$ which is compactly supported in $U$ the estimate*

$$|J(\lambda,\sigma,\delta)|\leqslant\frac{C\|\psi(\cdot,\delta)\|_{C^3}}{\lambda^{1/2+\varepsilon}|\sigma|^{l_m+c_m\varepsilon}} \tag{5.13}$$

*holds true uniformly for $|\sigma|+|\sigma|<\varepsilon'$, where $l_m$ and $c_m$ are defined as in Theorem* 5.4.

*Proof of Lemma* 5.7 *and Theorem* 5.4. We shall prove Lemma 5.7 and Theorem 5.4 at the same time by induction over $m$.

If $m=2$ then the phase function (5.12) is reduced to the form

$$F(x_1,x_2)=x_1^3+\delta_1x_1+\sigma x_2^2,$$

and by applying the method of the stationary phase in $x_2$ and van der Corput's lemma in $x_1$ we easily obtain estimate (5.13), with $l_2=\frac{1}{6}$. This proves also Theorem 5.3 for $m=2$.

Assume that $m\geqslant 3$, and that the statement of Theorem 5.4 holds for every strictly smaller value of $m$. We shall apply again a Duistermaat-type argument, in a similar way as in §4, in order to prove the statement of Lemma 5.7, and hence also that of Theorem 5.4, for $m$. To this end, we introduce the mixed-homogeneous scalings $\Delta_\varrho(x_1,x_2):=(\varrho^{1/2}x_1,\varrho^{1/2(m-1)}x_2)$, $\varrho>0$. Notice that these are such that the principal part of $f_2$ with respect to these dilations is given by $x_2^m+x_1x_2b(0,0,\sigma,\delta)$. Then

$$F(\Delta_\varrho(x),\sigma,\delta)=\varrho^{3/2}F(x,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta),$$

where $F(x,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta)=f_1(x_1,\tilde\delta)+\tilde\sigma f_2(x,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta)$ is given by

$$\begin{aligned} f_1(x_1,\tilde\delta)&:=x_1^3+\tilde\delta_1x_1,\\ f_2(x,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta)&:=x_2^m+\sum_{j=2}^{m-2}\tilde\delta_jx_2^{m-j}+(x_1-\tilde\delta_{m-1})x_2b(\Delta_\varrho(x),\sigma,\delta),\end{aligned}\tag{5.14}$$

with $\tilde\sigma$ and $\tilde\delta$ defined by

$$\tilde\sigma:=\frac{\sigma}{\varrho^{(2m-3)2(m-1)}},\quad \tilde\delta_1:=\frac{\delta_1}{\varrho}\quad\text{and}\quad \tilde\delta_j:=\frac{\delta_j}{\varrho^{j/2(m-1)}},\quad j=2,...,m-1,$$

so that in particular $\tilde\delta_{m-1}=\delta_{m-1}/\varrho^{1/2}$. Thus, if we define "dual scalings" by

$$\Delta_\varrho^*(\sigma,\delta):=(\tilde\sigma,\tilde\delta),$$

we see that if $b$ is constant, then $F(\Delta_\varrho(x),\sigma,\delta)=\varrho^{3/2}F(x,\Delta_\varrho^*(\sigma,\delta))$.

It is then natural to introduce the quasi-norm

$$N(\sigma,\delta):=|\sigma|^{2(m-1)/(2m-3)}+|\delta_1|+|\delta_2|^{m-1}+...+|\delta_{m-2}|^{2(m-1)/(m-2)}+|\delta_{m-1}|^2,$$

which is $\Delta_\varrho^*$-homogeneous of degree $-1$, i.e., $N(\Delta_\varrho^*(\sigma,\delta))=\varrho^{-1}N(\sigma,\delta)$.

Given $\sigma$ and $\delta$, we now choose $\varrho$ such that $N(\tilde\sigma,\tilde\delta)=1$, i.e.,

$$\varrho:=N(\sigma,\delta).$$

Notice that $\varrho \ll 1$, and that $(\tilde{\sigma}, \tilde{\delta})$ lies in the "unit sphere"

$$\Sigma := \{(\sigma', \delta') \in \mathbb{R}^m : N(\sigma', \delta') = 1\}.$$

Then, after scaling, we may rewrite

$$J(\lambda, \sigma, \delta) = J(\lambda, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta) := \varrho^{m/2(m-1)} \int_{\mathbb{R}^2} e^{i\lambda\varrho^{3/2} F(x, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta)} \psi(\Delta_\varrho(x), \delta)\, dx,$$

where here $\varrho$, $\sigma$ and the $\delta_j$ are small parameters. For a while, it will be convenient to consider $\tilde{\sigma}$ and the $\tilde{\delta}_j$ as additional, independent real parameters, which may not be small, but bounded.

We shall apply a dyadic decomposition to this integral. To this end, we choose $\chi_0, \chi \in C_0^\infty(\mathbb{R}^2)$ with $\operatorname{supp} \chi \subset \{x : \frac{1}{2}B < |x| < 2B\}$ (where $B$ is a sufficiently large positive number to be fixed later) such that

$$\chi_0(x) + \sum_{k=1}^{\infty} \chi(\Delta_{2^{-k}}(x)) = 1 \quad \text{for every } x \in \mathbb{R}^2.$$

Accordingly, we decompose the oscillatory integral

$$J(\lambda, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta) = \sum_{k=0}^{\infty} J_k(\lambda, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta),$$

where

$$J_k(\lambda, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta) := \varrho^{m/2(m-1)} \int_{\mathbb{R}^2} e^{i\lambda\varrho^{3/2} F(x, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta)} \psi(\Delta_\varrho(x), \delta) \chi_k(x)\, dx$$

and $\chi_k(x) := \chi(\Delta_{2^{-k}}(x))$ for $k \geqslant 1$.

Assume first that $k \geqslant 1$. Then, by using the scaling $\Delta_{2^k}$, we get

$$J_k(\lambda, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta) = (2^k \varrho)^{m/2(m-1)} \int_{\mathbb{R}^2} e^{i\lambda(2^k\varrho)^{3/2} F_k(x)} \psi(\Delta_{2^k\varrho}(x), \delta) \chi(x)\, dx,$$

where $F_k(x) := g_1(x_1, \tilde{\sigma}_k) + \tilde{\sigma}_k g_2(x, \tilde{\sigma}_k, \tilde{\delta}_k, 2^k\varrho, \sigma, \delta)$ is given by

$$\begin{aligned} g_1(x_1, \tilde{\delta}_k) &:= x_1^3 + \tilde{\delta}_{1,k} x_1, \\ g_2(x, \tilde{\sigma}_k, \tilde{\delta}_k, 2^k\varrho, \sigma, \delta) &:= x_2^m + \sum_{j=2}^{m-2} \tilde{\delta}_{j,k} x_2^{m-j} + (x_1 - \tilde{\delta}_{m-1,k}) x_2 b(\Delta_{2^k\varrho}(x), \sigma, \delta), \end{aligned}$$

with

$$(\tilde{\sigma}_k, \tilde{\delta}_k) := (\tilde{\sigma}_k, \tilde{\delta}_{1,k}, ..., \tilde{\delta}_{m-1,k}) := \Delta_{2^k}^*(\tilde{\sigma}, \tilde{\delta}) = \Delta_{2^k\varrho}^*(\sigma, \delta).$$

Observe that we may restrict ourselves to those $k$ for which $2^k\varrho\lesssim 1/B$, since otherwise $J_k\equiv 0$. Consequently, if we choose $B$ in the definition of $\chi$ sufficiently large, then $2^k\varrho\ll 1$, and also $|\tilde\sigma_k|+|\tilde\delta_k|\ll 1$. We thus see that there is some positive constant $c>0$ such that if $x\in\operatorname{supp}\chi$, then either

$$|\partial_1 g_1(x_1,\tilde\delta_k)|\geqslant c \quad \text{or} \quad |\partial_2 g_2(x_1,x_2,\tilde\sigma_k,\tilde\delta_k,2^k\varrho,\sigma,\delta)|\geqslant c.$$

Fix a point $x^0=(x_1^0,x_2^0)\in\operatorname{supp}\chi$, let $\eta$ be a smooth cut-off function supported in a sufficiently small neighborhood of $x^0$, and consider the oscillatory integral $J_k^\eta$ defined by

$$J_k^\eta(\lambda,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta)=(2^k\varrho)^{m/2(m-1)}\int_{\mathbb{R}^2} e^{i\lambda(2^k\varrho)^{3/2}F_k(x)}\psi(\Delta_{2^k\varrho}(x),\delta)\chi(x)\eta(x)\,dx.$$

By using integration by parts in $x_1$ if $|\partial_1 g_1(x_1^0,\tilde\delta_k)|\geqslant c$, respectively in $x_2$ if

$$|\partial_2 g_2(x_1^0,x_2^0,\tilde\sigma_k,\tilde\delta_k,2^k\varrho,\sigma,\delta)|\geqslant c,$$

and subsequently applying van der Corput's lemma to the $x_1$-integration in the latter case, we then obtain

$$|J_k^\eta|\leqslant \frac{C(2^k\varrho)^{m/2(m-1)}\|\psi(\,\cdot\,,\delta)\|_{C^3}}{(1+\lambda(2^k\varrho)^{3/2})^{1/3}(1+\lambda(2^k\varrho)^{3/2}|\tilde\sigma_k|)^{2/3}}\leqslant \frac{C(2^k\varrho)^{m/2(m-1)}\|\psi(\,\cdot\,,\delta)\|_{C^3}}{|\lambda(2^k\varrho)^{3/2}|^{1/2+\varepsilon}|\tilde\sigma_k|^{1/6+\varepsilon}}.$$

By means of a partition of unity argument this implies the same type of estimate

$$|J_k|\leqslant \frac{C(2^k\varrho)^{m/2(m-1)}\|\psi(\,\cdot\,,\delta)\|_{C^3}}{|\lambda(2^k\varrho)^{3/2}|^{1/2+\varepsilon}|\tilde\sigma_k|^{1/6+\varepsilon}}=C(2^k\varrho)^{(6-m)/12(m-1)-\varepsilon m/2(m-1)}\frac{\|\psi(\,\cdot\,,\delta)\|_{C^3}}{\lambda^{1/2+\varepsilon}|\sigma|^{1/6+\varepsilon}} \tag{5.15}$$

for $J_k$.

Consider first the case where $m<6$. Then clearly

$$\sum_{k\geqslant 1}|J_k|=\sum_{2^k\varrho\lesssim 1}|J_k(\lambda,\tilde\sigma,\tilde\delta,\varrho,\sigma,\delta)|\leqslant \frac{C\|\psi(\,\cdot\,,\delta)\|_{C^3}}{\lambda^{1/2+\varepsilon}|\sigma|^{1/6+\varepsilon}}.$$

Assume next that $m\geqslant 6$. Then the infinite series

$$\sum_{k=1}^{\infty}(2^k)^{(6-m)/12(m-1)-\varepsilon m/2(m-1)}$$

converges. Note also that $\varrho\geqslant|\sigma|^{2(m-1)/2m-3}$. Summing therefore over all $k\geqslant 1$, we obtain from (5.15) that

$$\sum_{k\geqslant 1}|J_k|\leqslant \frac{c\|\psi(\,\cdot\,,\delta)\|_{C^3}}{|\lambda|^{1/2+\varepsilon}|\sigma|^{l_m+c_m\varepsilon}}.$$

We are thus left with the integral

$$J_0(\lambda,\tilde\sigma,\tilde{\tilde\delta},\varrho,\sigma,\delta):=\varrho^{m/2(m-1)}\int_{\mathbb{R}^2} e^{i\lambda\varrho^{3/2}F(x,\tilde\sigma,\tilde{\tilde\delta},\varrho,\sigma,\delta)}\psi(\Delta_\varrho(x),\delta)\chi_0(x)\,dx,$$

where $F(x,\tilde\sigma,\tilde{\tilde\delta},\varrho,\sigma,\delta)$ is given by (5.14).

Let us fix a point $(\tilde\sigma^0,\tilde{\tilde\delta}^0)\in\Sigma$, and a point $x^0=(x_1^0,x_2^0)\in\operatorname{supp}\chi_0$, and let again $\eta$ be a smooth cut-off function supported near $x^0$. $J_0^\eta$ will be defined by introducing $\eta$ into the amplitude of $J_0$ in the same way as before. We shall prove that the oscillatory integral $J_0^\eta$ satisfies the estimate

$$|J_0^\eta|\leqslant\frac{C\|\psi(\,\cdot\,,\delta)\|_{C^3}}{\lambda^{1/2+\varepsilon}|\sigma|^{(l_m+c_m\varepsilon)}}, \tag{5.16}$$

provided $\eta$ is supported in a sufficiently small neighborhood $U$ of $x^0$ and $(\tilde\sigma,\tilde{\tilde\delta},\varrho,\sigma,\delta)\in V$, where $V$ is a sufficiently small neighborhood of the point $(\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)$. By means of a partition of unity argument this will then imply the same type of estimate for $J_0$, and hence for $J$, which will conclude the proof of Lemma 5.7, and thus also of Theorem 5.4.

Now, if either $\partial_1F(x_1^0,x_2^0,\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)\neq0$ or $\partial_2f_2(x_1^0,x_2^0,\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)\neq0$, then we can estimate $J_0^\eta$ exactly like the $J_k^\eta$ and get the required estimate (5.16) for $J_0^\eta$.

Assume therefore next that

$$\partial_1F(x_1^0,x_2^0,\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)=0 \quad\text{and also}\quad \partial_2f_2(x_1^0,x_2^0,\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)=0. \tag{5.17}$$

We then distinguish the following four cases.

*Case* 1. $\tilde\sigma^0\neq0$ and $x_1^0\neq0$.

Then, since $x_1^0\neq0$, it is easy to see from (5.14) that $\partial_1^2F(x_1^0,x_2^0,\tilde\sigma^0,\tilde{\tilde\delta}^0,0,0,0)\neq0$ as well. Note here that if we write $b(x,\varrho,\sigma,\delta):=b(\Delta_\varrho(x),\sigma,\delta)$, then

$$b(x,0,0,0)\equiv b(0,0,0,0)\neq0. \tag{5.18}$$

We can then argue here in a similar way as in the proof of Proposition 5.2, so let us only briefly sketch the argument. Suppose that $x_1^c(x_2,\tilde\sigma,\tilde{\tilde\delta},\varrho,\sigma,\delta)$ is a critical point of $F$ with respect to $x_1$. Then it is a smooth function of its variables, and if $\varrho=\sigma=\delta=0$, then by (5.18),

$$x_1^c=x_1^c(x_2,\tilde\sigma,\tilde{\tilde\delta},0,0,0)=\frac{(-(\tilde{\tilde\delta}_1+\tilde\sigma x_2b(0,0,0,0)))^{1/2}}{\sqrt3}$$

and

$$\begin{aligned}&F(x_1^c(x_2,\tilde\sigma,\tilde{\tilde\delta},0,0,0),x_2,\tilde\sigma,\tilde{\tilde\delta},0,0,0)\\&\qquad=(x_1^c)^3+\tilde{\tilde\delta}_1x_1^c+\tilde\sigma\biggl(x_2^m+\sum_{j=2}^{m-2}\tilde{\tilde\delta}_jx_2^{m-j}+(x_1^c-\tilde{\tilde\delta}_{m-1})x_2b(0,0,0,0)\biggr).\end{aligned}$$

If $\phi$ denotes the phase function

$$\phi(x_2, , \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta) := F(x_1^c(x_2, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta), x_2, \tilde{\sigma}, \tilde{\delta}, \varrho, \sigma, \delta),$$

which arises after applying the method of stationary phase to the $x_1$-integration, then since $\tilde{\sigma}^0 \neq 0$, this easily shows that there exists a natural number $N$ such that

$$\partial_2^N \phi(x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) \neq 0.$$

Consequently, we can in a second step apply van der Corput's lemma to the $x_2$-integration and obtain the estimate

$$|J_0^\eta| \leqslant \frac{C\varrho^{m/2(m-1)} \|\psi(\,\cdot\,, \delta)\|_{C^3}}{|\lambda\varrho^{3/2}|^{1/2+\varepsilon} |\tilde{\sigma}|^{1/6+\varepsilon}}, \tag{5.19}$$

which implies (5.16) as before (just put $k=0$ in our previous argument).

*Case* 2. $\tilde{\sigma}^0 \neq 0$ and $x_1^0 = 0$.

Then, by (5.18), we have $\partial_1^2 F(x_1^0, x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) = 0$ as well. But, again by (5.18), we also have $\partial_1 \partial_2 F(x_1^0, x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) \neq 0$, so that $F$ has a non-degenerate critical point at $x^0$ as a function of two variables. If the neighborhoods $U$ and $V$ are chosen sufficiently small, we can therefore apply the stationary phase method in two variables, which leads to an even stronger estimate than the estimate (5.19), since here $|\tilde{\sigma}| \sim 1$.

*Case* 3. $\tilde{\sigma}^0 = 0$ and $\tilde{\delta}_1^0 \neq 0$.

In this case $x_1^0 \neq 0$, because of (5.17), and thus $\partial_1^2 F(x_1^0, x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) \neq 0$. Moreover, in this situation we consider $\tilde{\sigma}$ such that $|\tilde{\sigma}| \ll 1$. Since we can regard $\tilde{\sigma} - \tilde{\sigma}^0$ and $\tilde{\delta} - \tilde{\delta}^0$ as small perturbation parameters if the neighborhoods $U$ and $V$ are chosen sufficiently small, we can therefore apply Proposition 5.2, with $\sigma$ in this proposition replaced by $\tilde{\sigma}$, and obtain (5.19).

*Case* 4. $\tilde{\sigma}^0 = 0$ and $\tilde{\delta}_1^0 = 0$.

Then, by (5.17), $x_1^0 = 0$ as well. In this case we make use of our induction hypothesis. Indeed, let us consider the function

$$x_2 \longmapsto f_2(0, x_2, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) = x_2^m + \sum_{j=2}^{m-2} \tilde{\delta}_j^0 x_2^{m-j} - \tilde{\delta}_{m-1}^0 x_2 b(0, 0, 0, 0).$$

Now $x_2 = x_2^0$ is a critical point, say of multiplicity $\mu - 1$, of this function, i.e.,

$$\partial_2^l f_2(0, x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) = 0$$

for $l = 1, ..., \mu - 1$ and $\partial_2^\mu f_2(0, x_2^0, \tilde{\sigma}^0, \tilde{\delta}^0, 0, 0, 0) \neq 0$.

Then $\mu<m$, because at least one of the coefficients $\tilde{\delta}_j$, $j=2,...,m-1$, does not vanish and $b(0,0,0,0)\neq 0$. Moreover, at this critical point also the condition

$$\partial_1\partial_2 f_2(0,x_2^0,\tilde{\sigma}^0,\tilde{\delta}^0,0,0,0)\neq 0$$

is satisfied. Therefore, after translating coordinate $x_2$ by $x_2^0$, by our hypothesis we may apply the conclusion of Theorem 5.3 for $\mu$ in place of $m$ and obtain the estimate

$$|J_0^\eta|\leqslant \frac{C\varrho^{m/2(m-1)}\|\psi(\,\cdot\,,\delta)\|_{C^3}}{|\lambda\varrho^{3/2}|^{1/2+\varepsilon}|\tilde{\sigma}|^{l_\mu+c_\mu\varepsilon}},$$

provided again that $U$ and $V$ are small enough. Now, if $\mu<6$, then this estimate agrees with (5.19), and we are done.

So, assume finally that $\mu\geqslant 6$. Since $l_m$ is increasing in $m$, we may replace $l_\mu$ by $l_{m-1}$ in this estimate, and clearly we have $c_\mu=c_m=2$. Recall also that here $\tilde{\sigma}=\sigma\varrho^{(3-2m)/2(m-1)}$ and $\varrho\geqslant|\sigma|^{2(m-1)/(2m-3)}$. Then the total exponent of $\varrho$ in this estimate, except for the terms containing $\varepsilon$, is $-3/4(m-1)(2m-5)$, and

$$\varrho^{-3/4(m-1)(2m-5)}\leqslant|\sigma|^{-3/2(2m-5)(2m-3)}.$$

Moreover, one computes that

$$|\sigma|^{-3/2(2m-5)(2m-3)-l_{m-1}}=|\sigma|^{-l_m}.$$

In a similar way, if we replace $\varrho$ by $|\sigma|^{2(m-1)/(2m-3)}$ in the term $|\varrho^{3/2}|^{-\varepsilon}|\tilde{\sigma}|^{-c_{m-1}\varepsilon}$, we obtain the additional factor $|\sigma|^{-3(m-1)\varepsilon/(2m-3)}\leqslant|\sigma|^{-2\varepsilon}$ in the estimate for $J_0^\eta$. In combination, we obtain again the estimate (5.16).

This concludes the proof of the lemma as well as of Theorem 5.4. □

## 6. Maximal estimates when $\partial_2\phi_{\mathrm{pr}}\neq 0$

Before coming to the proof of our main result, Theorem 1.2, we need to provide a variant of Corollary 4.6 for the case $m=1$. Its proof is in fact more elementary, but it will already make use of 2-dimensional oscillatory integral techniques.

We shall consider averaging operators of the form

$$A_t^{\varrho_0}f(z):=\int_{\mathbb{R}^2} f(z_1-tx_1,z_2-tx_2,z_3-t(1+\phi(x_1,x_2)))\varrho\biggl(\frac{x_2-\psi(x_1)}{\varepsilon_0 x_1^a}\biggr)\tilde{\eta}(x)\,dx, \qquad (6.1)$$

with $t>0$ and suitable smooth functions $\phi$, $\psi$, $\varrho$, $\tilde{\eta}$ and a positive exponent $a>0$ whose properties will be specified soon. The functions $\tilde{\eta}$ and $\varrho$ are supposed to be smooth bump functions supported near the origin in $\mathbb{R}^2$ and $\mathbb{R}$, respectively, and $\varepsilon_0>0$ is a small parameter. The maximal operator associated with the averaging operators $A_t^{\varrho_0}$, $t>0$, will be denoted by $\mathcal{M}^{\varrho_0}$.

By splitting into two half-planes, we may assume that the integration takes place over the half-plane $\mathbb{R}^2_+:=\{(x_1,x_2)\in\mathbb{R}^2:x_1>0\}$ only. We consider the Fourier transforms of the convolution kernels of the averaging operators $A_t^{\varrho_0}$, i.e.,

$$\widehat{A_t^{\varrho_0}f}(\xi)=e^{it\xi_3}J^{\varrho_0}(t\xi)\hat{f}(\xi),$$

where

$$J^{\varrho_0}(\xi):=\int_{\mathbb{R}^2_+} e^{it(\xi_1x_1+\xi_2x_2+\xi_3\phi(x_1,x_2))}\varrho\biggl(\frac{x_2-\psi(x_1)}{\varepsilon_0x_1^a}\biggr)\tilde{\eta}(x)\,dx,\quad \xi\in\mathbb{R}^3.$$

Our goal will be to derive suitable estimates of the oscillatory integrals $J^{\varrho_0}(\xi)$ (cf. the method in [18]). If we change to the coordinates $y_1:=x_1$ and $y_2:=x_2-\psi(x_1)$ in the integral, we obtain

$$J^{\varrho_0}(\xi):=\int_{\mathbb{R}^2_+} e^{i(\xi_1y_1+\xi_2\psi(y_1)+\xi_2y_2+\xi_3\phi^a(y))}\varrho\biggl(\frac{y_2}{\varepsilon_0y_1^a}\biggr)\eta(y)\,dy,$$

where $\eta$ is again a smooth cut-off function supported in a sufficiently small neighborhood of the origin and where

$$\phi^a(y):=\phi(y_1,y_2+\psi(y_1)).$$

In our later applications, the coordinates $y$ will be adapted to $\phi^a$.

We are thus led to consider Fourier multipliers of the form $e^{i\xi_3}J(\xi)$, with

$$J(\xi):=\int_{\mathbb{R}^2_+} e^{i(\xi_1x_1+\xi_2\psi(x_1)+\xi_2x_2+\xi_3\phi^a(x))}\varrho\biggl(\frac{x_2}{\varepsilon_0x_1^a}\biggr)\eta(x)\,dx. \tag{6.2}$$

*Assumptions* 6.1. The functions $\phi^a$, $\psi$ and $\eta$ are smooth functions such that

(i) $\psi(x_1)=b_1x_1^{m_1}+x_1^{m_1+1}q(x_1)$, where $b_1\in\mathbb{R}\setminus\{0\}$ and where $q$ is smooth;

(ii) the principal face $\pi(\phi^a)$ is a compact edge, and the associated principal part $\phi^a_{\mathrm{pr}}$ of $\phi^a$ is $\varkappa$-homogeneous of degree 1, where $0<\varkappa_1<\varkappa_2<1$ and $a=\varkappa_2/\varkappa_1>m_1\geqslant 2$;

(iii) $d(\phi^a)>1$;

(iv) $\eta$ is a smooth bump function supported in a sufficiently small neighborhood $\Omega$ of the origin.

Note that assumption (iii) implies that $\phi^a$ is of finite type, $\phi^a(0)=0$ and $\nabla\phi^a(0)=0$.

In order to estimate the maximal operator $\mathcal{M}^{\varrho_0}$ associated with the Fourier multiplier $e^{i\xi_3}J(\xi)$, we shall further decompose it and estimate the corresponding constituents. If $\chi$ is a bounded measurable function, we shall use the notation

$$J^\chi(\xi) := \int_{\mathbb{R}^2_+} e^{i(\xi_1 x_1+\xi_2\psi(x_1)+\xi_2 x_2+\xi_3\phi^a(x))}\varrho\biggl(\frac{x_2}{\varepsilon_0 x_1^a}\biggr)\eta(x)\chi(x)\,dx.$$

The corresponding rescaled Fourier multiplier operators are the averaging operators $A_t^\chi$ given by

$$\widehat{A_t^\chi f}(\xi) = e^{it\xi_3}J^\chi(t\xi)\hat{f}(\xi), \quad t>0,$$

with associated maximal operator $\mathcal{M}^\chi$. Then we shall make use of the following essentially well-known result in order to estimate $\mathcal{M}^\chi$.

LEMMA 6.2. *Assume that, for some $n\in\mathbb{N}$ and $\varepsilon>0$, the estimate*

$$|J^\chi(\xi)| \leqslant A_\chi \|\eta\|_{C^n(\mathbb{R}^2)}(1+|\xi|)^{-(1/2+\varepsilon)}, \quad \xi\in\mathbb{R}^3, \tag{6.3}$$

*holds, where the constant $A_\chi$ is independent of $\eta$. Moreover, put*

$$B_\chi := \int_{\mathbb{R}^2_+} \biggl|\varrho\biggl(\frac{x_2}{\varepsilon_0 x_1^a}\biggr)\eta(x)\chi(x)\biggr|\,dx.$$

*Then, for $2\leqslant p\leqslant\infty$,*

$$\|\mathcal{M}^\chi f\|_p \leqslant C A_\chi^{2/p} B_\chi^{1-2/p}\|f\|_p,$$

*where the constant $C$ depends only on the $C^n$-norms of $\phi^a$, $\psi$ and $\eta$, but not on $\chi$.*

*Proof.* Observe that

$$\biggl|\frac{\partial}{\partial t}[e^{it\xi_3}J^\chi(t\xi)]\biggr| \leqslant |\xi|(|J^\chi(t\xi)|+|(\nabla J^\chi)(t\xi)|),$$

where, because of (6.3),

$$|J^\chi(\xi)|+|(\nabla J^\chi)(\xi)| \leqslant C A_\chi (1+|\xi|)^{-(1/2+\varepsilon)}.$$

The desired estimate of the maximal operator for $p=2$ follows then essentially from Littlewood–Paley theory and Sobolev's embedding theorem (for details, see, e.g., [39, §XI.1], or our discussion in §3.1). Moreover, since $B_\chi$ is just the $L^1$-norm of the convolution kernel of $A_t^\chi$, the estimate for $p=\infty$ is trivial. The general case $2\leqslant p\leqslant\infty$ then follows by interpolation. □

Let us write

$$\Phi(x,\xi) := \xi_1 x_1 + \xi_2 \psi(x_1) + \xi_2 x_2 + \xi_3 \phi^a(x)$$

for the complete phase function of $J$, and decompose

$$\phi^a = \phi^a_{\mathrm{pr}} + \phi^a_r.$$

Consider the dilations $\delta_r(x_1,x_2) := (r^{\varkappa_1}x_1, r^{\varkappa_2}x_2)$, $r>0$, associated with the weight $\varkappa$. We choose a smooth non-negative function $\chi$ supported in the annulus

$$D := \{x : 1 \leqslant |x| \leqslant R\}$$

satisfying

$$\sum_{k=k_0}^{\infty} \chi_k(x) = 1 \quad \text{for } 0 \neq x \in \Omega,$$

where $\chi_k(x) := \chi(\delta_{2^k} x)$. Notice that by choosing $\Omega$ small, we can choose $k_0 \in \mathbb{N}$ as large as we need. We can then decompose $J$ dyadically as

$$J = \sum_{k=k_0}^{\infty} J_k,$$

where

$$J_k(\xi) := J^{\chi_k}(\xi) = \int_{\mathbb{R}^2} e^{i\Phi(x,\xi)} \varrho\left(\frac{x_2}{\varepsilon_0 x_1^a}\right) \eta(x) \chi_k(x)\, dx.$$

Notice that then

$$|\mathcal{M}^{\varrho_0} f| \leqslant \sum_{k=k_0}^{\infty} |\mathcal{M}^{\chi_k} f|.$$

By a change of coordinates, we obtain

$$J_k(\xi) = 2^{-k|\varkappa|} \int_{\mathbb{R}^2} e^{i2^{-k}\lambda \Phi_k(x,s)} \varrho\left(\frac{x_2}{\varepsilon_0 x_1^a}\right) \eta(\delta_{2^{-k}} x) \chi(x)\, dx,$$

where we have put $\lambda := \xi_3$, $s = (s_1, s_2)$ and

$$\Phi_k(x,s) := s_1 x_1 + S_2 \psi_k(x_1) + s_2 x_2 + \phi^a_{\mathrm{pr}}(x_1,x_2) + \phi_{r,k}(x),$$

with

$$\psi_k(x_1) := 2^{\varkappa_1 m_1 k} \psi(2^{-\varkappa_1 k} x_1) = b_1 x_1^{m_1} + O(2^{-\delta_1 k}) \quad \text{in } C^\infty, \tag{6.4}$$

$$\phi_{r,k}(x) := 2^k \phi^a_r(\delta_{2^{-k}} x) = O(2^{-\delta_2 k}) \quad \text{in } C^\infty \tag{6.5}$$

and

$$s_1 := 2^{(1-\varkappa_1)k} \frac{\xi_1}{\lambda}, \quad s_2 := 2^{(1-\varkappa_2)k} \frac{\xi_2}{\lambda}, \quad S_2 := 2^{(1-\varkappa_1 m_1)k} \frac{\xi_2}{\lambda} = 2^{(\varkappa_2-\varkappa_1 m_1)k} s_2 \tag{6.6}$$

(assuming without loss of generality that $\xi_3 \neq 0$), where $\delta_1, \delta_2 > 0$.

We remark that indeed $\psi_k(x)$ and $\phi_{r,k}(x)$ can be viewed as smooth functions $\tilde{\psi}(x_1, \delta)$ and $\tilde{\phi}_r(x, \delta)$, respectively, depending on the small parameter $\delta = 2^{-k/r}$ for some suitable positive integer $r \geqslant 1$ such that

$$\tilde{\psi}(x_1, 0) = b_1 x_1^{m_1} \quad \text{and} \quad \tilde{\phi}_r(x, 0) \equiv 0.$$

Observe also that $1-\varkappa_1 m_1 > \varkappa_2 - \varkappa_1 m_1 > 0$ and $1-\varkappa_j > 0$, so that in particular

$$|S_2| \gg |s_2| \quad \text{and} \quad |\lambda s_j| \gg |\xi_j|. \tag{6.7}$$

Also notice that in our domain of integration we have

$$x_1 \sim 1 \quad \text{and} \quad |x_2| \lesssim \varepsilon_0.$$

PROPOSITION 6.3. *Assume that $\phi^a$, $\psi$ and $\eta$ satisfy Assumptions* 6.1, *and that $\partial_2 \phi^a_{\mathrm{pr}}(1,0) \neq 0$. If $\varepsilon_0$ in* (6.2) *and the neighborhood $\Omega$ of the origin are chosen sufficiently small, then the following estimate*

$$|J_k(\xi)| \leqslant C \|\eta\|_{C^3(\mathbb{R}^2)} \frac{2^{-k|\varkappa|}}{(1+|2^{-k}\xi|)^{1/2+\varepsilon}} \tag{6.8}$$

*holds true for some $\varepsilon > 0$, where the constant $C$ does not depend on $k$ and $\xi$.*

*Consequently, the maximal operator $\mathcal{M}^{\varrho_0}$ associated with the averaging operators $A_t^{\varrho_0}$, $t>0$, defined by $\widehat{A_t^{\varrho_0} f}(\xi) = e^{it\xi_3} J^{\varrho_0}(t\xi) \hat{f}(\xi)$, is bounded on $L^p(\mathbb{R}^3)$ for every $p > 1/|\varkappa|$.*

*Proof.* We shall distinguish several cases, assuming for simplicity that $\lambda > 0$.

*Case* 1. $|s_1| + |S_2| \leqslant C$ for some large constant $C \gg 1$.

In this case, if $k$ is sufficiently large, then we have $|s_2| \ll 1$, and since $\partial_2 \phi_{\mathrm{pr}}(1,0) \neq 0$, we can integrate by parts in $x_2$ and obtain

$$|J_k(\xi)| \leqslant C \frac{2^{-k|\varkappa|}}{1+2^{-k}\lambda},$$

and hence (6.8), since, by (6.7), $|\xi| \sim \lambda$ in this case.

*Case* 2. $|s_1| + |S_2| \geqslant C$, with $C$ as above, and either $|s_1| \ll |S_2|$ or $|s_1| \gg |S_2|$.

In this case we can integrate by parts in $x_1$ and obtain

$$|J_k(\xi)| \leqslant \frac{C2^{-k|\varkappa|}}{1+2^{-k}\lambda(|s_1|+|S_2|)},$$

which again implies (6.8), since here, by (6.7), $|\xi| \lesssim \lambda(|s_1|+|S_2|)$.

*Case* 3. $|s_1|+|S_2| \geqslant C$, with $C$ as above, and $|s_1| \sim |S_2|$.

Observe first that $|s_1| \sim |S_2|$ implies $|\xi_2| \sim 2^{\varkappa_1(m_1-1)k}|\xi_1|$, so that

$$|\xi_2| \gg |\xi_1|.$$

We then write

$$2^{-k}\lambda\Phi_k(x,s) = 2^{-k}\lambda S_2 F(x,\sigma,\delta),$$

where

$$F(x,\sigma,\delta) := \frac{s_1}{S_2}x_1 + \tilde{\psi}(x_1,\delta) + \sigma(\phi^a_{\mathrm{pr}}(x_1,x_2) + \tilde{\phi}_r(x_1,x_2,\delta) + s_2 x_2),$$

$\delta := 2^{-k/r} \ll 1$ and $\sigma := 1/S_2$, so that

$$\left|\frac{s_1}{S_2}\right| \sim 1 \quad \text{and} \quad |\sigma| \ll 1.$$

Observe that

$$\left|\partial^2_{x_1}\left(\frac{s_1}{S_2}x_1 + \tilde{\psi}(x_1,0)\right)\right| \sim 1$$

for $x_1 \sim 1$. We also claim that the polynomial $P(x_2) := \phi^a_{\mathrm{pr}}(x^0_1, x_2)$ has degree

$$m := \deg P \geqslant 2. \tag{6.9}$$

For, otherwise, by the homogeneity of $\phi^a_{\mathrm{pr}}$, the polynomial $\phi^a_{\mathrm{pr}}$ would be of the form $\phi^a_{\mathrm{pr}}(x) = c_1 x_1^n + c_2 x_1^l x_2$, where the point $(l,1)$ would have to lie in the closed half-space above the bisectrix, since $\phi^a_{\mathrm{pr}}$ is the principal part of $\phi^a$. Thus $l \leqslant 1$, so that $d(\phi^a) \leqslant 1$, in contradiction to our assumption $d(\phi^a) > 1$.

From (6.9) we conclude that there is some integer $m \geqslant 2$ so that

$$|\partial^m_{x_2}(\phi^a_{\mathrm{pr}}(x_1,x_2) + s_2 x_2)| \sim 1.$$

If we now fix $x^0_1 \sim 1$ and translate the $x_1$-coordinate by $x^0_1$, we see that we can apply Proposition 5.2 if we localize our oscillatory integral $J_k$ to a small neighborhood of $(x^0_1, 0)$ by introducing a suitable cut-off function into the amplitude, and obtain an estimate of order

$$O(2^{-k|\varkappa|}(1+2^{-k}\lambda(|S_2|))^{-1/2}(1+2^{-k}\lambda)^{-1/m})$$

for the corresponding localized integral, uniformly in $s_1$ and $s_2$, since Proposition 5.2 also gives uniform estimates for small perturbations of such parameters. Since we can decompose $J_k(\xi)$ by means of a suitable partition of unity into such localized oscillatory integrals, we see that

$$|J_k(\xi)| \leqslant C2^{-k|\varkappa|}(1+2^{-k}\lambda(|S_2|))^{-1/2}(1+2^{-k}\lambda)^{-1/m},$$

where $m\geqslant 2$.

(a) If we assume that $|s_2|\leqslant C$ for some fixed, large constant $C$, then we have

$$|\xi_1| \ll |\xi_2| \ll |s_2\lambda| \leqslant C|\lambda|,$$

and hence $|\xi|\sim\lambda$, so that this estimate implies (6.8).

(b) If $|s_2|\gg 1$, then we proceed in a slightly different way. We first perform one integration by parts in $x_2$, and then apply the method of stationary phase in $x_1$. This leads to the estimate

$$|J_k(\xi)| \leqslant C2^{-k|\varkappa|}(1+2^{-k}\lambda(|S_2|))^{-1/2}(1+2^{-k}|s_2|\lambda)^{-1}.$$

If now $|\xi_2|\leqslant\lambda$, then $|\xi|\sim\lambda$, and if $|\xi_2|\geqslant\lambda$, then $|\xi|\sim|\xi_2|\ll|s_2|\lambda$, so that again (6.8) follows.

In order to estimate the maximal operator $\mathcal{M}^{\varrho_0}$, we observe that (6.8) implies that

$$|J_k(\xi)| \leqslant C_\varepsilon 2^{-k|\varkappa|}2^{k(1/2+\varepsilon)}(1+|\xi|)^{-1/2-\varepsilon}$$

for every sufficiently small $\varepsilon>0$. We may therefore choose $A_{\chi_k}:=C_\varepsilon 2^{-k|\varkappa|}2^{k(1/2+\varepsilon)}$ for $\chi=\chi_k$ in Lemma 6.2. Moreover, clearly we can choose $B_{\chi_k}:=C2^{-k|\varkappa|}$, so that we have

$$\|\mathcal{M}^{\chi_k} f\|_p \leqslant C_\varepsilon 2^{-k(|\varkappa|-1/p-\varepsilon)},$$

with a constant $C_\varepsilon$ which is independent of $k$. If $p>1/|\varkappa|$, and if $\varepsilon$ is chosen small enough, these estimates sum in $k$, so that the maximal operator $\mathcal{M}^{\varrho_0}$ is bounded on $L^p$. □

We shall indeed need a slight extension of this result to the following situation. As before, we shall always assume that $x_1>0$.

*Definitions.* Let $q\in\mathbb{N}^\times$ be a fixed positive integer. Assume that $\phi$ is a smooth function of the variables $x_1^{1/q}$ and $x_2$ near the origin, i.e., that there exists a smooth function $\phi^{[q]}$ near the origin such that $\phi(x)=\phi^{[q]}(x_1^{1/q},x_2)$. If the Taylor series of $\phi^{[q]}$ is given by

$$\phi^{[q]}(x_1,x_2) \sim \sum_{j,k=0}^{\infty} c_{j,k}x_1^j x_2^k,$$

then $\phi$ has the formal Puiseux series expansion

$$\phi(x_1, x_2) \sim \sum_{j,k=0}^{\infty} c_{j,k} x_1^{j/q} x_2^k.$$

We therefore define the *Taylor–Puiseux support* of $\phi$ by

$$\mathcal{T}(\phi) := \left\{ \left( \frac{j}{q}, k \right) \in \mathbb{N}_q^2 : c_{jk} \neq 0 \right\},$$

where

$$\mathbb{N}_q^2 := \left( \frac{1}{q} \mathbb{N} \right) \times \mathbb{N}.$$

The *Newton–Puiseux polyhedron* $\mathcal{N}(\phi)$ of $\phi$ at the origin is then defined to be the convex hull of the union of all the quadrants $(j/q, k)+\mathbb{R}_+^2$ in $\mathbb{R}^2$, with $(j/q, k)\in\mathcal{T}(\phi)$. The associated *Newton–Puiseux diagram* $\mathcal{N}_d(\phi)$ is the union of all compact faces of the Newton–Puiseux polyhedron, and the notions of principal face, Newton distance and homogenous distance are defined as in the case of Newton diagrams. The principal part $\phi_{\mathrm{pr}}$ is analogously defined by

$$\phi_{\mathrm{pr}}(x) := \sum_{(j/q,k)\in\pi(\phi)} c_{j,k} x_1^{j/q} x_2^k,$$

where $\pi(\phi)$ denotes the principal face. We shall then again decompose $\phi=\phi_{\mathrm{pr}}+\phi_r$.

Corollary 6.4. *Proposition* 6.3 *remains true under the following assumptions on* $\psi$ *and* $\phi^a$ *in place of Assumptions* 6.1, *provided again that* $\partial_2\phi^a_{\mathrm{pr}}(1,0)\neq 0$:

(i) $\psi$ *is given by*

$$\psi(x_1) = \sum_{l=1}^{L} b_l x_1^{m_l},$$

*where* $b_l\neq 0$ *for* $l=1, ..., K$, *and where* $2\leqslant m_1<...<m_L$ *are positive real numbers*;

(ii) $\phi^a$ *is a smooth function of the variables* $x_1^{1/q}$ *and* $x_2$, *the principal face* $\pi(\phi^a)$ *of its Newton–Puiseux polyhedron is a compact edge*, *and the associated principal part* $\phi^a_{\mathrm{pr}}$ *of* $\phi^a$ *is* $\varkappa$*-homogeneous of degree* 1, *where* $0<\varkappa_1<\varkappa_2<1$ *and* $a=\varkappa_2/\varkappa_1>m_1$;

(iii) *the distance* $d(\phi^a)=1/|\varkappa|$ *satisfies* $d(\phi^a)>1$;

(iv) $\eta$ *is a smooth bump function supported in a sufficiently small neighborhood* $\Omega$ *of the origin.*

*Proof.* All of our arguments extend in a straightforward manner to this setting, except perhaps for the proof of (6.9) and the straightforward application of Lemma 6.2. However, if (6.9) was false in the present situation, then we could write

$$\phi^a_{\mathrm{pr}}(x) = c_1 x_1^{n/q} + c_2 x_1^{l/q} x_2.$$

The point $(l/q,1)$ had to lie above the bisectrix, since $\phi^a_{\mathrm{pr}}$ is the principal part of $\phi^a$. Thus $l<q$. Moreover, we would have $\varkappa_1=q/n$ and $\varkappa_2=1-l/n$, so that

$$|\varkappa| = 1+\frac{q-l}{n} > 1,$$

and hence $d(\phi^a)<1$, in contradiction to our assumption (iii).

As for Lemma 6.2, notice that when applying the gradient to $J_k(\xi)$, the function $\eta$ will be multiplied with terms like $\phi^a$ or $\psi$, which may not be smooth at $x_1=0$, so that the argument in the proof of the lemma fails to hold. However, if we look at the formula for $J_k(\xi)$ after scaling the coordinates $x$, we find for instance that the factor $\eta(\delta_{2^{-k}}x)$ will have to be replaced by $\phi^a(\delta_{2^{-k}}x)\eta(\delta_{2^{-k}}x)$, where we now are in the domain where $x_1\sim 1$ and $|x_2|\lesssim\varepsilon_0$. But, in this domain, the $C^n$-norms of such expressions are still uniformly bounded in $k$, so that we obtain the same type of estimate as for $J_k(\xi)$. $\square$

## 7. Estimation of the maximal operator $\mathcal{M}$ when the coordinates are adapted or the height is strictly less than 2

We now turn to the proof of our main result, Theorem 1.2. As observed in the introduction, we may assume that $S$ is locally the graph $S=\{(x_1,x_2,1+\phi(x_1,x_2)):(x_1,x_2)\in\Omega\}$ of a function $1+\phi$. Here and in the subsequent sections, $\phi\in C^\infty(\Omega)$ will be a smooth real-valued function of finite type defined on an open neighborhood $\Omega$ of the origin in $\mathbb{R}^2$ and satisfying

$$\phi(0,0)=0 \quad\text{and}\quad \nabla\phi(0,0)=0.$$

In this section we shall consider the easiest cases where the coordinates $x$ are adapted to $\phi$, or where $h(\phi)<2$.

Recall that $A_t$, $t>0$, denotes the corresponding family of averaging operators

$$A_tf(y) := \int_S f(y-tx)\varrho(x)\,d\sigma(x),$$

where $d\sigma$ denotes the surface measure on $S$ and $\varrho\in C_0^\infty(S)$ is a non-negative cut-off function. We shall assume that $\varrho$ is supported in an open neighborhood $U$ of the point $(0,0,1)$ which will be chosen sufficiently small. The associated maximal operator is given by

$$\mathcal{M}f(y) := \sup_{t>0}|A_tf(y)|, \quad y\in\mathbb{R}^3. \tag{7.1}$$

The averaging operator $A_t$ can be rewritten in the form

$$A_tf(y) := \int_{\mathbb{R}^2} f(y_1-tx_1, y_2-tx_2, y_3-t(1+\phi(x_1,x_2)))\eta(x_1,x_2)\,dx,$$

where $\eta$ is a smooth function supported in $\Omega$. If $\chi$ is any integrable function defined on $\Omega$, we shall denote by $A_t^\chi$ the correspondingly localized averaging operator

$$A_t^\chi f(y) := \int_{\mathbb{R}^2} f(y_1 - tx_1, y_2 - tx_2, y_3 - t(1 + \phi(x_1, x_2)))\chi(x)\eta(x)\,dx,$$

and by $\mathcal{M}^\chi$ the associated maximal operator

$$\mathcal{M}^\chi f(y) := \sup_{t>0} |A_t^\chi f(y)|, \quad y \in \mathbb{R}^3.$$

Proposition 7.1. *Let $\phi$ be as above, and assume that $\varkappa = (\varkappa_1, \varkappa_2)$ is a given weight such that $0 < \varkappa_1 \leqslant \varkappa_2 < 1$. As in* (2.2), *we decompose*

$$\phi = \phi_\varkappa + \phi_r$$

*into its $\varkappa$-principal part $\phi_\varkappa$ and the remainder term $\phi_r$ consisting of terms of $\varkappa$-degree greater than* 1. *Then, if the neighborhood $\Omega$ of the point* $(0,0)$ *is chosen sufficiently small, the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p > \max\{2, h(\phi_\varkappa)\}$.*

*Proof.* Let us modify our notation slightly and write points in $\mathbb{R}^3$ in the form $(x, x_3)$, with $x \in \mathbb{R}^2$ and $x_3 \in \mathbb{R}$. Recall from Corollary 2.2 the crucial fact that

$$h(\phi_\varkappa) = \max\{m(\phi_\varkappa), d_h(\phi_\varkappa)\}.$$

In particular, the multiplicity of every real root of the $\varkappa$-homogeneous polynomial $\phi_\varkappa$ is bounded by $h(\phi_\varkappa)$.

Choosing a dyadic decomposition

$$\sum_{k=k_0}^{\infty} \chi_k(x) = 1$$

for $0 \neq x \in \Omega$ as in §6, with $\chi_k(x) := \chi(\delta_{2^k} x)$, where the $\delta_r$ denote the dilations associated with the weight $\varkappa$, and where $\chi$ is supported in an annulus $1 \leqslant |x| \leqslant R$, we can write $A_t$ as a sum of averaging operators

$$A_t f(y, y_3) = \sum_{k=k_0}^{\infty} A_t^k f(y, y_3),$$

where $A_t^k := A_t^{\chi_k}$. Notice that by choosing $\Omega$ small, we can choose $k_0 \in \mathbb{N}$ to be large.

If we apply the change of variables $x \mapsto \delta_{2^{-k}}(x)$ in the integral above, we obtain

$$A_t^k f(y, y_3) = 2^{-k|\varkappa|} \int_{\mathbb{R}^2} f(y - t\delta_{2^{-k}}(x), y_3 - t(1 + 2^{-k}\phi^k(x)))\eta(\delta_{2^{-k}}(x))\chi(x)\,dx,$$

where

$$\phi^k(x) := \phi_\varkappa(x) + 2^k \phi_r(\delta_{2^{-k}}(x))$$

and where the perturbation term $2^k\phi_r(\delta_{2^{-k}}(\cdot))$ is of order $O(2^{-\varepsilon k})$ for some $\varepsilon>0$ in any $C^M$-norm. To express this fact, we shall in the sequel again use the short-hand notation

$$2^k \phi_r(\delta_{2^{-k}}(\cdot)) = O(2^{-\varepsilon k}).$$

By $\mathcal{M}^k$ we shall denote the maximal operator $\mathcal{M}^{\chi_k}$ associated with the averaging operators $A_t^k$.

Assume now that $p>\max\{2, h(\phi_\varkappa)\}$. We define the scaling operator $T^k$ by

$$T^k f(y, y_3) := 2^{k|\varkappa|/p} f(\delta_{2^k}(y), y_3).$$

Then $T^k$ acts isometrically on $L^p(\mathbb{R}^3)$, and

$$(T^{-k} A_t^k T^k) f(y, y_3) = 2^{-k|\varkappa|} \int_{\mathbb{R}^2} f(y - tx, y_3 - t(1 + 2^{-k}\phi^k(x)))\eta(\delta_{2^{-k}}(x))\chi(x)\, dx.$$

Assuming that $\Omega$ is a sufficiently small neighborhood of the origin, we need to consider only the case where $k$ is sufficiently large.

Let $x^0 \in D$ be a fixed point.

If $\nabla\phi_\varkappa(x^0)\neq 0$, then by Euler's homogeneity relation one can easily derives that $\operatorname{rank}(D^2\phi_\varkappa(x^0))\geqslant 1$ (see [18, Lemma 3.3]). Therefore, we can find a unit vector $e\in\mathbb{R}^2$ such that $\partial_e^2\phi_\varkappa(x^0)\neq 0$, where $\partial_e$ denotes the partial derivative in direction of $e$.

If $\nabla\phi_\varkappa(x^0)=0$, then by Euler's homogeneity relation we have $\phi_\varkappa(x^0)=0$ as well. Thus the function $\phi_\varkappa$ vanishes in $x^0$ at least of order 2, so that $m(\phi_\varkappa)\geqslant 2$, and hence $h(\phi_\varkappa)\geqslant 2$. On the other hand, by what we remarked earlier, it vanishes along the circle passing through $x^0$ and centered at the origin at most of order $h(\phi_\varkappa)$. Therefore, we can find a unit vector $e\in\mathbb{R}^2$ such that $\partial_e^m\phi_\varkappa(x^0)\neq 0$, for some $m$ with $2\leqslant m\leqslant h(\phi_\varkappa)$.

Thus, in both cases, after rotating coordinates so that $e=(0,1)$, we may apply Proposition 4.5 to conclude that for $p>\max\{2, h(\phi_\varkappa)\}$ and sufficiently large $k$,

$$\|T^{-k}\mathcal{M}^k T^k f\|_p \leqslant C 2^{k(1/p - |\varkappa|)} \|f\|_p, \quad f \in \mathcal{S}(\mathbb{R}^3),$$

if we replace $\chi$ in the definition of $A_t^k$ by $\chi\eta$, where $\eta$ is a bump function supported in a sufficiently small neighborhood of $x^0$. This is equivalent to

$$\|\mathcal{M}^k f\|_p \leqslant C 2^{k(1/p - |\varkappa|)} \|f\|_p.$$

Decomposing $\chi$ and correspondingly $A_t^k$ by means of a suitable partition of unity into a finite number of such pieces, we see that the same estimate holds for the original operators $\mathcal{M}^k$.

Since $1/|\varkappa| = d_h(\phi_\varkappa)\leqslant h(\phi_\varkappa)<p$, we can sum over all $k\geqslant k_0$ and obtain the desired estimate for $\mathcal{M}$. $\square$

Let us apply this result first to the case where the coordinates $x$ are adapted to $\phi$, possibly after a rotation of the coordinate system $(x_1, x_2)$. Observe first that a linear change of the coordinates $(x_1, x_2)$ induces a corresponding linear change of coordinates in $\mathbb{R}^3$ which fixes the coordinate $x_3$. This linear transformation is an automorphism of $\mathbb{R}^3$, so that it preserves the convolution product on $\mathbb{R}^3$ (up to a fixed factor), and hence the norm of the maximal operator $\mathcal{M}$. We may thus assume that the coordinates are adapted to $\phi$.

If we choose the weight $\varkappa$ as in Lemma 2.4 of §2.4, then Proposition 7.1 immediately implies the following result.

Corollary 7.2. *Let $\phi$ be as above, and assume that, possibly after a rotation of the coordinate system, the coordinates $x$ are adapted to $\phi$, i.e., that $h(\phi)=d(\phi)$. Then, if the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>\max\{2, h(\phi)\}$.*

*Remark* 7.3. One can easily extend Corollary 7.2 as follows: If the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, then the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>\max\{2, h(\phi_{\mathrm{pr}})\}$, no matter if the coordinates are adapted to $\phi$ or not.

*Proof.* Indeed, if the coordinates are adapted, then $h(\phi_{\mathrm{pr}})=h(\phi)$. So, assume that the coordinates $(x_1, x_2)$ are not adapted to $\phi$. Then the principal face of the Newton polyhedron is a compact edge, so that the principal part $\phi_{\mathrm{pr}}$ of $\phi$ is $\varkappa$-homogeneous, where $\varkappa$ satisfies the assumptions of Proposition 7.1. Since $\phi_{\mathrm{pr}}=\phi_\varkappa$, the result then follows from this proposition. □

The result above holds even when the coordinates are not adapted, but then it will in general not be sharp, since we have $h(\phi_{\mathrm{pr}})\geqslant h(\phi)$ (see [19, Corollary 4.3]), and in general strict inequality holds.

For example, let $\phi(x_1, x_2):=(x_2-x_1^2)^2+x_1^5$. Then we have $\phi_{\mathrm{pr}}(x)=(x_2-x_1^2)^2$. The coordinate system is not adapted to $\phi$, because $d(\phi)=\frac{4}{3}<2$, where 2 is the multiplicity of the root of $\phi_{\mathrm{pr}}$. A coordinate system which is adapted to $\phi$ is given by $y_2:=x_2-x_1^2$ and $y_1:=x_1$. It is then easy to see that $h(\phi_{\mathrm{pr}})=2>\frac{10}{7}=h(\phi)$.

Corollary 7.4. *If $h(\phi)<2$, and if the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, then the maximal operator $\mathcal{M}$ is bounded on $L^p(\mathbb{R}^3)$ for any $p>2$, also when the coordinates are not adapted to $\phi$.*

*Proof.* If $D^2\phi(0,0)\neq 0$ then we have at least one non-vanishing principal curvature at the origin, so that the result follows from C. D. Sogge's main theorem in [36].

Next, we consider the case where $D^2\phi(0,0)=0$. Then necessarily $D^3\phi(0,0)\neq 0$, for otherwise $h(\phi)\geqslant d(\phi)\geqslant 2$. In particular, $h(\phi)>1$. Denote by $P_3$ the Taylor polynomial of degree 3 with base point 0 of the function $\phi$, so that $P_3=\phi_\varkappa$, if we choose $\varkappa:=\left(\frac{1}{3},\frac{1}{3}\right)$. If $h(P_3)\leqslant 2$, then we obtain the desired estimate from Proposition 7.1. Assume therefore that $h(P_3)>2$. Then, by formula (2.6), $P_3$ must have a root of order 3. Thus, possibly after rotating the coordinate system, we may assume that $P_3(x_1,x_2)=cx_2^3$ with $c\neq 0$.

Now, we consider the Taylor support $\mathcal{T}(\phi)$ of $\phi$. As $\mathcal{T}(\phi)\subset\left\{(t_1,t_2):\frac{1}{3}t_1+\frac{1}{3}t_2\geqslant 1\right\}$, one easily checks that the subset

$$\left\{(t_1,t_2):\tfrac{1}{6}t_1+\tfrac{1}{3}t_2<1\right\}\cap\mathcal{T}(\phi)$$

of $\mathcal{T}(\phi)$ contains at most three points, namely $(4,0)$, $(5,0)$ and $(3,1)$, all of them lying below the bisectrix $t_1=t_2$.

Moreover, any line passing through the point $(0,3)\in\mathcal{T}(\phi)$ corresponding to $P_3=cx_2^3$ contains at most one of these points. Thus, if

$$\left\{(t_1,t_2):\tfrac{1}{6}t_1+\tfrac{1}{3}t_2<1\right\}\cap\mathcal{T}(\phi)\neq\varnothing,$$

then the principal part $\phi_{\mathrm{pr}}$ of $\phi$ contains only two monomials, one corresponding to the point $(0,3)$ above the bisectrix and the other one corresponding to one of the points listed above which lie below the bisectrix, i.e., $\phi_{\mathrm{pr}}$ is of the form $dx_1^4+cx_2^3$, $dx_1^5+cx_2^3$ or $dx_1^3x_2+cx_2^3$, with $d\neq 0$ (note that these all satisfy $d(\phi_{\mathrm{pr}})<2$). We remark that these correspond to the singularities of type $E_k$ with $k=6,7,8$ in Arnold's classification (see [2] and [11]). Therefore on the unit circle $\phi_{\mathrm{pr}}$ has no root of multiplicity bigger than 1, so that the coordinate system is adapted to $\phi$, and thus $h(\phi)<2$. The desired estimate for $\mathcal{M}$ follows therefore in this case from Corollary 7.2.

Finally assume that

$$\left\{(t_1,t_2):\tfrac{1}{6}t_1+\tfrac{1}{3}t_2<1\right\}\cap\mathcal{T}(\phi)=\varnothing.$$

Then $\mathcal{T}(\phi)\subset\left\{(t_1,t_2):\frac{1}{6}t_1+\frac{1}{3}t_2\geqslant 1\right\}$, and hence $h(\phi)\geqslant d(\phi)\geqslant 2$, which contradicts our assumption. $\square$

## 8. Non-adapted coordinates: Estimation of the maximal operator $\mathcal{M}$ away from the principal root jet

In view of the results in the previous section, we shall from now on assume that the coordinates $(x_1,x_2)$ are not adapted to $\phi$, and that $h:=h(\phi)\geqslant 2$.

Recall then from §2.5 that there exists a smooth function $\psi=\psi(x_1)$, the principal root jet, defined on a neighborhood of the origin such that an adapted coordinate system $(y_1,y_2)$ for $\phi$ is given locally near the origin by

$$y_1 := x_1 \quad \text{and} \quad y_2 := x_2-\psi(x_1). \tag{8.1}$$

In these coordinates, $\phi$ is given by

$$\phi^a(y) := \phi(y_1, y_2+\psi(y_1)). \tag{8.2}$$

The Taylor series expansion of $\psi$ is of the form

$$\psi(x_1)\sim\sum_{l=1}^{K} b_l x_1^{m_l}, \quad 1\leqslant K\leqslant\infty, \tag{8.3}$$

where we assume that all coefficients $b_l$ are non-trivial and where $1\leqslant m_1<m_2<\dots$. Recall that $K$ is finite in cases (a) and (b) listed below.

We first make the simple observation that, if $m_1=1$, the linear change of coordinates

$$y_1 := x_1 \quad \text{and} \quad y_2 := x_2-b_1x_1^{m_1}$$

allows us to reduce to the case $m_1\geqslant 2$ (cf. the corresponding discussion in the previous section). In the sequel, we shall therefore always assume that

$$2\leqslant m_1<m_2<\dots. \tag{8.4}$$

We shall also only consider the region where $x_1>0$, in order to simplify the notation. The remaining half-plane can be treated in the same way.

Recall also from §2.4 that one of the following cases applies:

(a) $\pi(\phi^a)$ is a compact edge, and either $\varkappa_2^a/\varkappa_1^a\notin\mathbb{N}$, or $\varkappa_2^a/\varkappa_1^a\in\mathbb{N}$ and $m(\phi^a_{\mathrm{pr}})\leqslant d(\phi)$;

(b) $\pi(\phi^a)$ consists of a vertex;

(c) $\pi(\phi^a)$ is unbounded.

Here $\varkappa^a=(\varkappa_1^a,\varkappa_2^a)=\varkappa^\lambda$ denotes the principal weight in case (a).

We recall from Corollary 3.2 that in case (a), if $a=\varkappa_2^\lambda/\varkappa_1^\lambda=\varkappa_2^a/\varkappa_1^a\in\mathbb{N}$, then there may be at most one point $y^m\in S_0^1$ for which $\partial_2^j\phi^a_{\varkappa^\lambda}(y^m)=0$ for $2\leqslant j\leqslant h$. If such a point $y^m$ exists, we shall perform the following convenient modification in our definition of $\psi$:

Denote by $y_2=c_py_1^a$ the corresponding real root of $\partial_2^j\phi^a_{\varkappa^\lambda}$, i.e., $y_2^m=c_p(y_1^m)^a$. We shall then define $\psi$ by

$$\psi(x_1) := \sum_{l=1}^{K} b_l x_1^{m_l}+c_px_1^a. \tag{8.5}$$

Notice that the corresponding change of coordinates leads again to adapted coordinates, and that the corresponding principal face $\pi(\phi^a)$ is a compact edge lying on the same line as in the previous coordinates, or a vertex.

This modification allows us to assume that in case (a), with $a \in \mathbb{N}$, for any point $y^0 \in S_0^1$ with $y_2^0 \neq 0$,

$$\partial_2^j \phi^a_{\varkappa^\lambda}(y^0) \neq 0 \quad \text{for some } 2 \leqslant j \leqslant h. \tag{8.6}$$

### 8.1. Preliminary reduction to a $\varkappa$-homogeneous neighborhood of the principal root $x_2 = b_1 x_1^{m_1}$ of $\phi_{\mathrm{pr}}$

Recall from the construction of adapted coordinates in §2.5 that since the coordinates $x$ are not adapted to $\phi$, the principal face $\pi(\phi)$ is a compact edge of the Newton polyhedron of $\phi$, so that it lies on a unique line $\varkappa_1 t_1 + \varkappa_2 t_2 = 1$, where $0 < \varkappa_1 \leqslant \varkappa_2$ and

$$\frac{\varkappa_2}{\varkappa_1} = m_1 \geqslant 2.$$

Moreover, if $\phi_{\mathrm{pr}} = \phi_\varkappa$ denotes the principal part of $\phi$, we have $m(\phi_{\mathrm{pr}}) > d(\phi_{\mathrm{pr}})$, and $m(\phi_{\mathrm{pr}})$ is just the multiplicity of the principal root $b_1 x_1^{m_1}$ of the $\varkappa$-homogeneous polynomial $\phi_{\mathrm{pr}}$. All other roots have multiplicity less than or equal to $d(\phi_{\mathrm{pr}})$.

This already indicates that the function $\phi$ will indeed be small of "highest order" (in some averaged sense) near the curve $x_2 = \psi(x_1)$ given by the principal root jet (even though $\phi$ need not vanish on this curve!), so that the region close to this curve should indeed give the main contribution to the maximal operator.

In order to localize to a $\varkappa$-homogeneous region away from the principal root jet, put, in a first step,

$$\varrho_1(x_1, x_2) := \varrho\left(\frac{x_2 - b_1 x_1^{m_1}}{\varepsilon_1 x_1^{m_1}}\right),$$

where $\varepsilon_1 > 0$ is a small parameter to be determined later, and set

$$A_t^{1-\varrho_1} f(y) := \int_{\mathbb{R}^2} f(y_1 - t x_1, y_2 - t x_2, y_3 - t(1 + \phi(x_1, x_2)))\left(1 - \varrho\left(\frac{x_2 - b_1 x_1^{m_1}}{\varepsilon_1 x_1^{m_1}}\right)\right)\eta(x)\, dx.$$

By $\mathcal{M}^{1-\varrho_1}$ we denote the associated maximal operator.

Lemma 8.1. *Let $\varepsilon_1 > 0$. If the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, then the maximal operator $\mathcal{M}^{1-\varrho_1}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p > h(\phi) =: h$.*

*Moreover, if $\mathcal{N}(\phi^a)$ is of the form $(\nu_1, h) + \mathbb{R}_+^2$, with $\nu_1 < h$, (case (c2) in §3), then the same statement holds true for the maximal operator $\mathcal{M}$ in place of $\mathcal{M}^{1-\varrho_1}$.*

*Proof.* We can argue exactly as in the proof of Proposition 7.1. Using the dilations $\delta_r(x_1,x_2)=\delta_r^{\varkappa}(x_1,x_2):=(r^{\varkappa_1}x_1,r^{\varkappa_2}x_2)$, $r>0$, we can dyadically decompose the operators $A_t^{1-\varrho_1}$ into the sum of operators $A_t^k$, which, after rescaling, are given by

$$(T^{-k}A_t^kT^k)f(y,y_3)=2^{-k|\varkappa|}\int_{\mathbb{R}^2}f(y_1-tx_1,y_2-tx_2,y_3-t(1+\phi^k(x_1,x_2)))\\ \times\left(1-\varrho\left(\frac{x_2-b_1x_1^{m_1}}{\varepsilon_1x_1^{m_1}}\right)\right)\eta(\delta_{2^{-k}}x)\chi(x)\,dx.$$

All roots of $\phi_{\rm pr}$ lying in the domain of integration have a positive distance to the principal root $b_1x_1^{m_1}$, and hence have multiplicities bounded by the distance $d(\phi_{\rm pr})$ (cf. Corollary 2.2), so that we can again estimate the associated maximal operators $\mathcal{M}^k$ by means of Proposition 4.5 (applied possibly in a rotated coordinate system) and obtain the first statement.

To prove the second statement, assume that $\mathcal{N}(\phi^a)=(\nu_1,h)+\mathbb{R}_+^2$, $\nu_1<h$. Then the Newton polyhedron of $\phi^a$ intersects the line $\varkappa_1t_1+\varkappa_2t_2=1$ in the single point $(\nu_1,h)$, so that $\phi_\varkappa^a(y)=cy_1^{\nu_1}y_2^h$, with $c\neq0$. Then $\phi_\varkappa(x)=cx_1^{\nu_1}(x_2-b_1x_1^{m_1})^h$, which implies that $h(\phi_\varkappa)=h(\phi_\varkappa^a)=h$, and we see that in this case we can argue as in the case of adapted coordinates by means of Proposition 7.1 to verify the second statement of the lemma. $\square$

### 8.2. Further domain decompositions

In view of Lemma 8.1, we may and shall from now on assume that the Newton polyhedron of $\phi^a$ has at least one compact edge "lying above" the principal face, i.e., that one of the cases (a), (b) or (c1) from §3 applies.

Furthermore, we have reduced considerations to a narrow $\varkappa$-homogeneous domain near the curve $x_2=b_1x_1^{m_1}$, of the form $|x_2-b_1x_1^{m_1}|\leqslant\varepsilon_1x_1^{m_1}$, where $\varepsilon_1>0$ can be chosen arbitrarily small. Since $\psi(x_1)=b_1x_1^{m_1}+$ terms of higher $\varkappa$-degree, choosing $\Omega$ sufficiently small we see that we are left to estimate the contribution to $\mathcal{M}$ of a domain of the form

$$|x_2-\psi(x_1)|\leqslant\varepsilon_1x_1^{m_1}. \tag{8.7}$$

In order to apply the results from §3, let us assume for the time being that $\phi$ is analytic. Recall then our choice of the index $\lambda$ and the number $a=a_\lambda=\varkappa_2^\lambda/\varkappa_1^\lambda>m_1$ (cf. (3.6)) from that section, so that the principal part of $\phi^a$ is $\varkappa^\lambda$-homogeneous of degree 1, in the case where the principal face is a compact edge $\gamma_\lambda$.

As a major step in the proof of Theorem 1.2 we shall in the sequel narrow down the domain (8.7) to a small $\varkappa^\lambda$-homogeneous neighborhood of the principal root jet of the form

$$|x_2-\psi(x_1)|\leqslant\varepsilon_\lambda x_1^{a_\lambda}. \tag{8.8}$$

More precisely, we fix a cut-off function $\varrho \in C_0^\infty(\mathbb{R})$ supported in a neighborhood of the origin such that $\varrho=1$ near the origin, and put

$$\varrho_0(x_1,x_2):=\varrho\left(\frac{x_2-\psi(x_1)}{\varepsilon_0 x_1^a}\right), \quad \varepsilon_0>0.$$

Recall that $\eta$ is a smooth function supported in the neighborhood $\Omega$ of the origin, define averaging operators

$$A_t^{1-\varrho_0}f(y):=\int_{\mathbb{R}^2} f(y_1-tx_1,y_2-tx_2,y_3-t(1+\phi(x_1,x_2)))\left(1-\varrho\left(\frac{x_2-\psi(x_1)}{\varepsilon_0 x_1^a}\right)\right)\eta(x)\,dx,$$

and consider the associated maximal operator $\mathcal{M}^{1-\varrho_0}$. Our goal will then be to prove the following result.

Proposition 8.2. *Let $\varepsilon_0>0$. If the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, then the maximal operator $\mathcal{M}^{1-\varrho_0}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>h(\phi)$.*

*Moreover, if the principal face $\pi(\phi^a)$ is a vertex or unbounded, then the same holds true for $\mathcal{M}$ in place of $\mathcal{M}^{1-\varrho_0}$.*

In order to prove Proposition 8.2, we shall decompose the difference set of the domains in (8.7) and (8.8) into domains $D_l$ of the form

$$D_l:=\{(x_1,x_2):\varepsilon_l x_1^{a_l}<|x_2-\psi(x_1)|\leqslant N_l x_1^{a_l}\}, \quad l=l_0,...,\lambda,$$

which, when expressed in terms of the coordinates $y$, are $\varkappa^l$-homogeneous, and the intermediate domains

$$E_l:=\{(x_1,x_2):N_{l+1}x_1^{a_{l+1}}<|x_2-\psi(x_1)|\leqslant \varepsilon_l x_1^{a_l}\}, \quad l=l_0,...,\lambda-1,$$

and

$$E_{l_0-1}:=\{(x_1,x_2):N_{l_0}x_1^{a_{l_0}}<|x_2-\psi(x_1)|\leqslant \varepsilon_1 x_1^{m_1}\}.$$

Here, the $\varepsilon_l>0$ are small and the $N_l>0$ are large parameters to be determined later, and $l_0\geqslant 1$ is chosen such that

$$a_l\leqslant m_1 \text{ for } l<l_0 \quad \text{and} \quad a_l>m_1 \text{ for } l\geqslant l_0. \tag{8.9}$$

To localize to domains of type $D_l$, we put

$$\varrho_l(x_1,x_2):=\varrho\left(\frac{x_2-\psi(x_1)}{N_l x_1^{a_l}}\right)-\varrho\left(\frac{x_2-\psi(x_1)}{\varepsilon_l x_1^{a_l}}\right), \quad l=l_0,...,\lambda, \tag{8.10}$$

and set

$$A_t^{\varrho_l} f(z) := \int_{\mathbb{R}^2} f(z_1 - tx_1, z_2 - tx_2, z_3 - t(1+\phi(x_1, x_2)))\varrho_l(x)\eta(x)\,dx,$$

with associated maximal operator $\mathcal{M}^{\varrho_l}$.

Similarly, in order to localize to domains of type $E_l$, we put

$$\begin{aligned}\tau_l(x_1, x_2) &:= \varrho\left(\frac{x_2 - \psi(x_1)}{\varepsilon_l x_1^{a_l}}\right)(1-\varrho)\left(\frac{x_2 - \psi(x_1)}{N_{l+1} x_1^{a_{l+1}}}\right), \quad l = l_0, ..., \lambda - 1, \\ \tau_{l_0-1}(x_1, x_2) &:= \varrho\left(\frac{x_2 - \psi(x_1)}{\varepsilon_1 x_1^{m_1}}\right)(1-\varrho)\left(\frac{x_2 - \psi(x_1)}{N_{l_0} x_1^{a_{l_0}}}\right),\end{aligned} \tag{8.11}$$

and set

$$A_t^{\tau_l} f(z) := \int_{\mathbb{R}^2} f(z_1 - tx_1, z_2 - tx_2, z_3 - t(1+\phi(x_1, x_2)))\tau_l(x)\eta(x)\,dx,$$

with associated maximal operator $\mathcal{M}^{\tau_l}$.

Notice that it suffices to control all the maximal operators defined in this way in order to prove the first statement in Proposition 8.2.

### 8.3. The maximal operators $\mathcal{M}^{\varrho_l}$

LEMMA 8.3. *If the neighborhood $\Omega$ of the point $(0,0)$ is chosen sufficiently small, then the maximal operator $\mathcal{M}^{\varrho_l}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p > h(\phi)$.*

*Proof.* In view of (8.10), the change of variables (8.1) transforms the integral for $A_t^{\varrho_l} f(z)$ into

$$A_t^{\varrho_l} f(z) = \int_{\mathbb{R}^2} f(z_1 - ty_1, z_2 - t(y_2 + \psi(y_1)), z_3 - t(1+\phi^a(y_1, y_2)))\varrho_l^a(y)\eta^a(y)\,dy,$$

with

$$\varrho_l^a(y) := \varrho\left(\frac{y_2}{N_l y_1^{a_l}}\right) - \varrho\left(\frac{y_2}{\varepsilon_l y_1^{a_l}}\right)$$

and

$$\eta^a(y) := \eta(y_1, y_2 + \psi(y_1)).$$

We recall that the $\varkappa^l$-principal part $\phi^a_{\varkappa^l}$ of $\phi^a$ is $\varkappa^l$-homogeneous of degree 1 and $\varrho_l^a$ is $\varkappa^l$-homogeneous of degree zero with respect to the dilations

$$\delta_r^l(y_1, y_2) := \delta_r^{\varkappa^l}(y_1, y_2) := (r^{\varkappa_1^l} y_1, r^{\varkappa_2^l} y_2), \quad r > 0.$$

Decomposing $\phi^a=\phi^a_{\varkappa^l}+\phi^a_r$ as in (2.2), where $\phi^a_r$ consists of terms of $\varkappa^l$-degree higher than 1, we dyadically decompose the operators $A_t^{\varrho_l}$ into the sum of operators $A_t^k$, with associated maximal operators $\mathcal{M}^k$, given by

$$\begin{aligned} A_t^k f(z) = 2^{-k|\varkappa^l|} \int_{\mathbb{R}^2} f(z_1 - t2^{-\varkappa_1^l k} y_1, z_2 - t(2^{-\varkappa_2^l k} y_2 + \psi(2^{-\varkappa_1^l k} y_1)), \\ z_3 - t(1+2^{-k}\phi^k(y_1,y_2)))\varrho_l^a(y)\eta^a(\delta^l_{2^{-k}} y)\chi(y)\, dy, \end{aligned}$$

with

$$\phi^k(y) := \phi^a_{\varkappa^l}(y) + 2^k \phi^a_r(\delta^l_{2^{-k}} y).$$

Notice that $2^k\phi^a_r(\delta^l_{2^{-k}}y)=O(2^{-\varepsilon k})$ in $C^\infty$, for some $\varepsilon>0$, so that this term can be considered as a perturbation term. Rescaling by means of the operators

$$T^k f(z, z_3) := 2^{k|\varkappa^l|/p} f(\delta^l_{2^k}(z), z_3),$$

we obtain

$$\begin{aligned} (T^{-k}A_t^k T^k) f(z) = 2^{-k|\varkappa^l|} \int_{\mathbb{R}^2} f(z_1 - ty_1, z_2 - t(y_2 + \psi^k(y_1)), z_3 - t(1+2^{-k}\phi^k(y_1,y_2))) \\ \times \varrho_l^a(y)\eta^a(\delta^l_{2^{-k}} y)\chi(y)\, dy. \end{aligned}$$

Here,

$$\psi^k(y_1) := 2^{\varkappa_2^l k}\psi(2^{-\varkappa_1^l k} y_1) = O(2^{(\varkappa_2^l - \varkappa_1^l m_1)k}) \quad \text{in } C^\infty,$$

since $\psi(x_1)=b_1 x_1^{m_1}+o(x_1^{m_1})$. Applying the change of variables $x_1:=y_1$, $x_2:=y_2+\psi^k(y_1)$ to this integral, we eventually arrive at

$$\begin{aligned} (T^{-k}A_t^k T^k) f(z) = 2^{-k|\varkappa^l|} \int_{\mathbb{R}^2} f(z_1 - tx_1, z_2 - tx_2, z_3 - t(1+2^{-k}\phi^k(x_1, x_2 - \psi^k(x_1)))) \\ \times \tilde{\eta}(x_1, x_2 - \psi^k(x_1))\, dx, \end{aligned} \tag{8.12}$$

with

$$\tilde{\eta}(y) := \chi(y)\varrho_l^a(y)\eta^a(\delta^l_{2^{-k}} y).$$

Since

$$1 > \varkappa_2^l - \varkappa_1^l m_1 = \varkappa_1^l(a_l - m_1) > 0$$

(cf. Lemma 3.1, (3.3) and (8.9)), we can no longer argue as in §7 by means of Proposition 4.5 in order to estimate the corresponding maximal operators. However, we shall see that we can make use of Corollary 4.6 in combination with Proposition 2.3, if we choose $\varepsilon:=2^{-k}$, $\psi_\varepsilon:=\psi^k$ and $\delta:=\varkappa_2^l-\varkappa_1^l m_1=\varkappa_1^l(a_l-m_1)$ in Corollary 4.6.

Indeed, assume first that for every point $y^0$ in the support of $\chi\varrho_l^a$ the following holds true:

$$\text{There exists some } j \text{ with } 2\leqslant j\leqslant d_h(\phi^a_{\varkappa^l}) \text{ such that } \partial_2^j\phi^a_{\varkappa^l}(y^0)\neq 0. \tag{8.13}$$

Then, by means of a partition of unity argument, we may reduce ourselves to a sufficiently small neighborhood of any such point $y^0$, and after applying Corollary 4.6 we may proceed exactly as in the proof of Proposition 7.1 in order to show that $\mathcal{M}^{\varrho_l}$ is bounded on $L^p$ for $p>d_h(\phi^a_{\varkappa^l})$.

Now, if $l<\lambda$, this assumption holds true, in view of Corollary 3.2 (i) , and $d_h(\phi^a_{\varkappa^l})<h$, so that we are done.

Thus, assume that $l=\lambda$. If either $\pi(\phi^a)$ is a vertex as in case (b), or an unbounded face as in case (c1), then we can argue as before in view of Corollary 3.2 (ii).

We may thus assume that $\pi(\phi^a)$ is a compact edge (case (a)). Recall that then $d_h(\phi^a_{\varkappa^\lambda})=1/|\varkappa^\lambda|=h$ and $\phi^a_{\varkappa^\lambda}=\phi^a_{\mathrm{pr}}$. If $a\in\mathbb{N}$, then (8.13) is again satisfied in view of (8.6), so we are left with the case where $a\notin\mathbb{N}$.

From Corollary 3.2, we know that in this case there is some $1\leqslant j\leqslant h$ such that $\partial_2^j\phi^a_{\varkappa^\lambda}(y^0)\neq 0$. If $j\geqslant 2$, we can argue as before.

We are left with those points $y^0$ in the support of $\chi\varrho^a_\lambda$ for which $\partial_2^j\phi^a_{\mathrm{pr}}(y^0)=0$ for every $2\leqslant j\leqslant h$ but $\partial_2\phi^a_{\mathrm{pr}}(y^0)\neq 0$. Given such a point $y^0$, we need to control the contribution of a sufficiently small $\varkappa^\lambda$-homogeneous neighborhood of it to the maximal operator, i.e., the maximal operator associated with averaging operators

$$A_t^{y^0}f(z):=\int_{\mathbb{R}^2} f(z_1-ty_1, z_2-t(y_2+\psi(y_1)), z_3-t(1+\phi^a(y_1,y_2)))\varrho_{y^0}(y)\eta^a(y)\,dy,$$

with

$$\varrho_{y^0}(y):=\varrho\biggl(\frac{y_2-cy_1^a}{\varepsilon_0 y_1^a}\biggr),$$

where $\varepsilon_0>0$ is sufficiently small and $c\in\mathbb{R}$ is such that $y_2^0=c(y_1^0)^a$. Changing coordinates, we find that

$$A_t^{y^0}f(z):=\int_{\mathbb{R}^2} f(z_1-tx_1, z_2-tx_2, z_3-t(1+\phi(x_1,x_2)))\varrho\biggl(\frac{x_2-\tilde\psi(x_1)}{\varepsilon_0 x_1^a}\biggr)\tilde\eta(x)\,dx,$$

where

$$\tilde\psi(x_1):=\psi(x_1)+cx_1^a=\sum_{l=1}^K b_l x_1^{m_l}+cx_1^a.$$

Notice that $2\leqslant m_1<...<m_K<a$, where $a$ is rational. Moreover, if we put

$$\tilde\phi^a(y):=\phi(y_1, y_2+\tilde\psi(y_1)),$$

then

$$(\tilde{\phi}^a)_{\mathrm{pr}}(y_1,y_2)=\phi^a_{\mathrm{pr}}(y_1,y_2+cy_1^a).$$

Therefore, the homogeneity of $(\tilde{\phi}^a)_{\mathrm{pr}}$ and our assumption $\partial_2\phi^a_{\mathrm{pr}}(y^0)\neq 0$ imply that

$$\partial_2(\tilde{\phi}^a)_{\mathrm{pr}}(1,0)\neq 0.$$

Corollary 6.4, applied to $\tilde{\psi}$ in place of $\psi$ and $\tilde{\phi}^a$ in place of $\phi^a$ then implies that the associated maximal operator is $L^p$-bounded, if $p>1/|\varkappa^\lambda|=h$.

This completes the proof of the lemma. $\square$

### 8.4. The maximal operators $\mathcal{M}^{\tau_l}$

LEMMA 8.4. *If the neighborhood $\Omega$ of the point $(0,0)$ and $\varepsilon_l$ are chosen sufficiently small and $N_{l+1}$ sufficiently large, then the maximal operator $\mathcal{M}^{\tau_l}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>h(\phi)$.*

*Proof.* (I) We begin with the case $l_0\leqslant l\leqslant\lambda-1$. Since the domain $E_l$, when viewed in $y$-coordinates, is a domain of transition between two different homogeneities, namely the ones given by the weights $\varkappa^l$ and $\varkappa^{l+1}$ (at least if $l\geqslant 1$), we shall apply an idea from Phong and Stein's article [31] and decompose it dyadically in each coordinate separately, and then rescale each of the bi-dyadic pieces obtained in this way.

By the change of variables (8.1), we can write

$$A_t^{\tau_l}f(z)=\int_{\mathbb{R}^2} f(z_1-ty_1,z_2-t(y_2+\psi(y_1)),z_3-t(1+\phi^a(y_1,y_2)))\tau_l^a(y)\eta^a(y)\,dy,$$

with

$$\tau_l^a(y):=\varrho\biggl(\frac{y_2}{\varepsilon_l y_1^{a_l}}\biggr)(1-\varrho)\biggl(\frac{y_2}{N_{l+1}y_1^{a_{l+1}}}\biggr),$$

and $\eta^a$ as before (cf. (8.11)).

Consider a dyadic partition of unity $\sum_{k=0}^\infty\chi_k(s)=1$ on the interval $0<s\leqslant 1$ with $\chi\in C_0^\infty(\mathbb{R})$ supported in the interval $\left[\frac12,4\right]$, where $\chi_k(s):=\chi(2^ks)$, and put

$$\chi_{j,k}(x):=\chi_j(x_1)\chi_k(x_2),\quad j,k\in\mathbb{N}.$$

We then decompose $A_t^{\tau_l}$ into the operators

$$A_t^{j,k}f(z):=\int_{\mathbb{R}^2} f(z_1-ty_1,z_2-t(y_2+\psi(y_1)),z_3-t(1+\phi^a(y_1,y_2)))\tau_l^a(y)\eta^a(y)\chi_{j,k}(y)\,dy,$$

with associated maximal operators $\mathcal{M}^{j,k}$.

Notice that by choosing the neighborhood $\Omega$ of the origin sufficiently small, we need only consider sufficiently large $j$ and $k$. Moreover, because of the localization imposed by $\tau_l^a$, it suffices to consider only pairs $(j,k)$ satisfying

$$a_l j+M \leqslant k \leqslant a_{l+1} j-M, \tag{8.14}$$

where $M$ can still be choosen sufficiently large, because we have the freedom to choose $\varepsilon_l$ sufficiently small and $N_{l+1}$ sufficiently large. In particular, we have $j\sim k$.

By rescaling in the integral, we have

$$\begin{aligned} A_t^{j,k} f(z) = 2^{-j-k} \int_{\mathbb{R}^2} f(z_1-t2^{-j}y_1, z_2-t(2^{-k}y_2+\psi(2^{-j}y_1)),\\ z_3-t(1+\phi^a(2^{-j}y_1, 2^{-k}y_2)))\tilde{\tau}^{j,k}(y)\tilde{\eta}^{j,k}(y)\chi(y_1)\chi(y_2)\,dy, \end{aligned}$$

with

$$\tilde{\tau}^{j,k}(y) := \varrho\left(\frac{y_2}{\varepsilon_l 2^{k-a_l j} y_1^{a_l}}\right)(1-\varrho)\left(\frac{y_2}{N_{l+1}2^{k-a_{l+1}j}y_1^{a_{l+1}}}\right), \quad \tilde{\eta}^{j,k}(y) := \eta^a(2^{-j}y_1, 2^{-k}y_2).$$

Notice that, by (8.14), all derivatives of $\tilde{\tau}^{j,k}$ are uniformly bounded in $j$ and $k$.

The scaling operators

$$T^{j,k} f(z) := 2^{(j+k)/p} f(2^j z_1, 2^k z_2, z_3)$$

then transform these operators into

$$\begin{aligned} (T^{-j,-k}A^{j,k}T^{j,k})f(z) = 2^{-j-k} \int_{\mathbb{R}^2} f(z_1-ty_1, z_2-t(y_2+\psi^{j,k}(y_1)),\\ z_3-t(1+\tilde{\phi}^{j,k}(y)))\tilde{\tau}^{j,k}(y)\tilde{\eta}^{j,k}(y)\chi(y_1)\chi(y_2)\,dy, \end{aligned}$$

where

$$\tilde{\phi}^{j,k}(y) := \phi^a(2^{-j}y_1, 2^{-k}y_2) \quad \text{and} \quad \psi^{j,k}(y_1) := 2^k\psi(2^{-j}y_1).$$

Notice that

$$\psi^{j,k} = O(2^{k-m_1 j}) \quad \text{in } C^\infty.$$

Applying the change of variables $x_1:=y_1$, $x_2:=y_2+\psi^{j,k}(y_1)$ to this integral, we eventually arrive at

$$\begin{aligned} (T^{-j,-k}A^{j,k}T^{j,k})f(z) = 2^{-j-k} \int_{\mathbb{R}^2} f(z_1-tx_1, z_2-tx_2,\\ z_3-t(1+\tilde{\phi}^{j,k}(x_1, x_2-\psi^{j,k}(x_1))))\tau^{j,k}(x)\eta^{j,k}(x)\chi^{j,k}(x)\,dx, \end{aligned} \tag{8.15}$$

where

$$\tau^{j,k}(x) := \tilde{\tau}^{j,k}(x_1, x_2 - \psi^{j,k}(x_1)), \quad \eta^{j,k}(x) := \eta(2^{-j}x_1, 2^{-k}x_2)$$

and

$$\chi^{j,k}(x) := \chi(x_1)\chi(x_2 - \psi^{j,k}(x_1)).$$

We next determine $\tilde{\phi}^{j,k}$, up to an error term. To this end, notice that if $y_1 \sim 1$ and $y_2 \sim 1$, and if $r \in \left[\begin{smallmatrix}\cdot\\ \mu\end{smallmatrix}\right]$, then $r(2^{-j}y_1) = c_\mu^\alpha 2^{-a_\mu j} y_1^{a_\mu} + O(2^{-\varepsilon(j+k)})$ in $C^\infty$, for some $\varepsilon > 0$. In view of (8.14), we thus get that

$$2^{-k}y_2 - r(2^{-j}y_1) = \begin{cases} -c_\mu^\alpha 2^{-a_\mu j}(y_1^{a_\mu} + O(2^{-\varepsilon(j+k)})), & \text{if } \mu < l, \\ -c_\mu^\alpha 2^{-a_\mu j}(y_1^{a_\mu} + O(2^{-M})), & \text{if } \mu = l, \\ 2^{-k}(y_2 + O(2^{-M})), & \text{if } \mu = l+1, \\ 2^{-k}(y_2 + O(2^{-\varepsilon(j+k)})), & \text{if } \mu > l+1, \end{cases}$$

with $M$ as in (8.14). Multiplying all these terms, we then see that

$$\tilde{\phi}^{j,k}(y) = 2^{-(A_l j + B_l k)}(c_l y_1^{A_l} y_2^{B_l} + O(2^{-CM})), \tag{8.16}$$

for some constant $C > 0$, where $A_l$ and $B_l$ are given by (3.2) and $M$ can still be chosen as large as we wish.

Observe that since $l \leqslant \lambda - 1$, we have $B_l \geqslant B_{\lambda-1} \geqslant \nu_2 + N\left[\begin{smallmatrix}\cdot\\ \lambda\end{smallmatrix}\right]$, and similarly as in the proof of Lemma 3.1, it is easy to see that we must have $\nu_2 + N\left[\begin{smallmatrix}\cdot\\ \lambda\end{smallmatrix}\right] \geqslant 2$, and hence $B_l \geqslant 2$. This implies that

$$\partial_2^2(y_1^{A_l} y_2^{B_l}) \sim 1,$$

and that $A_l j + B_l k \geqslant 2k$, so that $2^{k - m_1 j} \leqslant C 2^{(A_l j + B_l k)/2}$, and hence

$$\psi^{j,k} = O((2^{-(A_l j + B_l k)})^{-1/2})$$

in $C^\infty$. We can therefore argue in a similar way as in the previous subsection and apply Corollary 4.6, with $\varepsilon := 2^{-(A_l j + B_l k)}$ and $m = 2$, to obtain that

$$\|\mathcal{M}^{j,k} f\|_p \leqslant C 2^{(A_l j + B_l k)/p - j - k} \|f\|_p,$$

whenever $p > 2$, provided $j + k$ is sufficiently large.

Summing all these estimates, we thus have that

$$\|\mathcal{M}^{\tau_l} f\|_p \leqslant CJ \|f\|_p,$$

where

$$J := \sum_{\substack{(j,k) \\ a_l j + M \leqslant k \leqslant a_{l+1} j - M}} 2^{(A_l j + B_l k)/p - j - k}.$$

Assume now that $p>h(\phi)$. As $h(\phi)=d(\phi^a)\geqslant d_h(\phi^a_{\varkappa^\lambda})\geqslant d_h(\phi^a_{\varkappa^{l+1}})$, and since, by (3.4), $d_h(\phi^a_{\varkappa^{l+1}})=(A_l+a_{l+1}B_l)/(1+a_{l+1})$, we have

$$p>\frac{A_l+a_{l+1}B_l}{1+a_{l+1}}.$$

This condition is equivalent to

$$\left(1-\frac{A_l}{p}\right)+a_{l+1}\left(1-\frac{B_l}{p}\right)>0. \tag{8.17}$$

Similarly, since the mapping $a\mapsto(A_l+aB_l)/(1+a)$ is increasing, we may replace $a_{l+1}$ by $a_l$ in this estimate and also get

$$\left(1-\frac{A_l}{p}\right)+a_l\left(1-\frac{B_l}{p}\right)>0. \tag{8.18}$$

In order to estimate $J$, let us write $k$ in the form $k=\theta a_l j+(1-\theta)a_{l+1}j+\omega$, with $0\leqslant\theta\leqslant 1$ and $|\omega|\leqslant M$. Then

$$\begin{aligned} j+k-\frac{A_l j+B_l k}{p} &= \left(1-\frac{A_l}{p}\right)j+\left(1-\frac{B_l}{p}\right)k \\ &= \left(\theta\left[\left(1-\frac{A_l}{p}\right)+a_l\left(1-\frac{B_l}{p}\right)\right]+(1-\theta)\left[\left(1-\frac{A_l}{p}\right)+a_{l+1}\left(1-\frac{B_l}{p}\right)\right]\right)j+\left(1-\frac{B_l}{p}\right)\omega. \end{aligned}$$

In view of (8.17) and (8.18), this shows that there exists a positive constant $\varepsilon>0$ such that

$$j+k-\frac{A_l j+B_l k}{p}>\varepsilon j,$$

provided $j$ is sufficiently large. It is now clear that $J<\infty$, so that the maximal operator $\mathcal{M}^{\tau_l}$ is bounded on $L^p$ whenever $p>h(\phi)$.

(II) We now consider the case $l=l_0-1$. This case can be treated in a very similar way (formally, it is like the previous case, only with $a_{l_0-1}$ replaced by $m_1\geqslant a_{l_0-1}$). Indeed, in this case (8.14) must be replaced by the inequalities

$$m_1 j+M\leqslant k\leqslant a_{l_0}j-M,$$

from which one derives that (8.16) remains valid, with $l=l_0-1$. From here, we can proceed exactly as before. $\square$

In order to prove the second statement in Proposition 8.2, we assume that we are in case (b) or case (c1) from §3, so that the principal face of $\mathcal{N}(\phi^a)$ is either a vertex or unbounded. Recall from Corollary 3.2 (ii) that then $d_h(\phi^a_{\varkappa^\lambda})\leqslant h$ and $\partial_2^h\phi^a_{\varkappa^\lambda}(1,0)\neq 0$. Since, by the first part of Proposition 8.2, what remains to be controlled is the contribution to the maximal operator $\mathcal{M}$ given by a domain of the form $|x_2-\psi(x_1)|\leqslant\varepsilon_0 x_1^a$, where $\varepsilon_0$ can be chosen as small as needed, we can thus argue as in the estimation of the maximal operators $\mathcal{M}^{\varrho_\lambda}$ by means of Corollary 4.6, provided that $p>h$ and $p>d_h(\phi^a_{\varkappa^\lambda})$, i.e., that $p>h$. This concludes the proof of Proposition 8.2.

### 8.5. Reduction of the smooth case to the analytic setting

We shall now show how the estimates for the maximal operators in the preceding subsections can be established also for more general, smooth finite-type functions $\phi$.

To this end, let us first outline a small modification of our approach for analytic $\phi$:

Recall from §2.5 that the function $\psi$ is a polynomial, except for the case (c), in which the principal face of the Newton polyhedron of $\phi^a$ is an unbounded half-line $L$ of the form $t_1 \geqslant \nu_1$, $t_2 = h$, with $\nu_1 < h$, provided the algorithm described in that subsection does not terminate. Assume the latter is the case. Then we have seen that there is some $K_0$ such that for $k \geqslant K_0$, the principal part of the function $\phi_{(k)}$ constructed from $\phi$ in the $k$th step of the algorithm is of the form (2.17), i.e.,

$$(\phi_{(k)})_{\mathrm{pr}}(x) = c_k x_1^{\nu_1} (x_2 - b_{k+1} x_1^{m_{k+1}})^h \quad \text{for every } k \geqslant K_0.$$

Suppose now that we stop our algorithm at the $k$th step, where $k \geqslant K_0$ will be assumed to be sufficiently large, and write $\tilde{\phi} := \phi_{(k)}$. Let us then put

$$\tilde{\psi}(x_1) := \sum_{l=1}^{k} b_l x_1^{m_l}.$$

Then $\tilde{\phi}$ arises from $\phi$ by means of the change of coordinates $y_1 := x_1$, $y_2 := x_2 - \tilde{\psi}(x_1)$ in the same way as $\phi^a$ arises from $\phi$ by means of the change of coordinates $y_1 := x_1$, $y_2 := x_2 - \psi(x_1)$, only that $\tilde{\psi}$ and the corresponding coordinate transformation will be polynomial. We shall call such coordinates *almost adapted*. The Newton diagrams of $\phi^a$ and $\tilde{\phi}$ are the same, except that the unbounded horizontal edge $L$ of $\mathcal{N}_d(\phi^a)$ is replaced by the compact edge $L_k := [(\nu_1, h), (\nu_1 + h m_{k+1})]$ in $\mathcal{N}_d(\tilde{\phi})$, whose slope $1/m_{k+1}$ tends to zero as $k \to \infty$. Moreover, for $l \leqslant \lambda$, the $\varkappa^l$-principal parts of $\phi$ and $\tilde{\phi}$ associated with the edges "left to the principal faces" will also coincide (cf. the discussion in §3).

Therefore, if we work with $\tilde{\phi}$ in place of $\phi^a$, most of our previous arguments carry over verbatim. Indeed, a look at the proofs of Lemma 8.1 and Proposition 8.2 shows that only the second part of Lemma 8.1, which concerns the case (c2), requires some modification. However, it is still true that the Newton polyhedron of $\tilde{\phi}$ intersects the line $\varkappa_1 t_1 + \varkappa_2 t_2 = 1$ in the single point $(\nu_1, h)$, provided we choose $k$ so large that the line $L_k$ has a slope smaller than the slope of the line $\varkappa_1 t_1 + \varkappa_2 t_2 = 1$, which we may assume. From here on, the proof of Lemma 8.1 proceeds as before.

Assume now that $\phi$ is only smooth and of finite type in place of analytic. In the case where the principal root jet $\psi$ of $\phi$ is not polynomial, we work with almost adapted coordinates for $\phi$ as explained before, where we choose $k \geqslant K_0$ so large that the argument in the second part of Lemma 8.1 still works (notice that the proof of this lemma works for smooth functions $\phi$ as well). In all other cases, we use adapted coordinates, and write $\tilde{\phi}$ in place of $\phi^a$ and $\tilde{\psi}$ in place of $\psi$, too.

Denote by $P_n$ the Taylor polynomial of degree $n$ of $\phi$ centered at the origin. Since we are now working with a polynomial change of coordinates, we may choose the degree $n$ so large that the Newton polyhedra of $\phi$ and $P_n$ as well as of $\tilde{\phi}$ and $\widetilde{P}_n$ coincide, as do their faces and corresponding principal parts. Here, $\widetilde{P}_n$ denotes the polynomial obtained from $P_n$ by means of the change of coordinates $y_1{:=}x_1$, $y_2{:=}x_2-\tilde{\psi}(x_1)$, which is just what we get in the $k$th step from Varchenko's algorithm when applied to $P_n$, provided again that $n$ is chosen large enough.

Since $\phi$ is a small perturbation of $P_n$ near the origin when $n$ is large, it is then clear that the estimations of the operators $\mathcal{M}^{\varrho_l}$ in Proposition 8.2 work in the same way for $\phi$ as for $P_n$. Moreover, there exists a constant $c>0$ such that if $R_{j,k}$ is a dyadic rectangle on which $x_1\sim 2^{-j}$ and $x_2\sim 2^{-k}$, then the remainder term $\phi-P_n$ is of order $O(2^{-c(j+k)n})$ in $C^\infty(R_{j,k})$. We may thus apply our previous approach to the polynomial $P_n$ in place of $\phi$ and choose $n$ so large that the contributions of the remainder term $\phi-P_n$ can be considered as negligible errors for the estimations of the operators $\mathcal{M}^{\tau_l}$ associated with $\phi$ (compare the order $O(2^{-c(j+k)n})$ with the order of $\tilde{\phi}^{j,k}$ in formula (8.16)). In this way, also the proof of Proposition 8.2 extends to $\phi$.

## 9. Estimation of the maximal operator $\mathcal{M}$ near the principal root jet

In view of Proposition 8.2, we may and shall from now on assume that the principal face of $\mathcal{N}(\phi^a)$ is a compact edge (case (a)).

What remains to be controlled is the contribution to the maximal operator $\mathcal{M}$ given by a domain of the form

$$|x_2-\psi(x_1)|\leqslant \varepsilon_0 x_1^a, \quad \text{with } x_1>0, \tag{9.1}$$

where $\varepsilon_0>0$ can be chosen as small as we need. More precisely, in view of Proposition 8.2, what remains to be proven is that the maximal operator $\mathcal{M}^{\varrho_0}$ associated with this domain (defined already by (6.1) in §6) is bounded on $L^p(\mathbb{R}^3)$ for every $p>h=h(\phi)$.

Now, if there is some $1\leqslant j\leqslant h$ such that $\partial_2^j\phi_{\mathrm{pr}}^a(1,0)\neq 0$, then this can be proven in exactly the same way as we had proven the second statement of Proposition 8.2, provided $2\leqslant j\leqslant h$. And, if $j=1$, then we can argue as in the last part of the proof of Lemma 8.3 by means of Corollary 6.4.

The proof of Theorem 1.2 will therefore be completed once we have verified the following result.

PROPOSITION 9.1. *Assume that* $\pi(\phi^a)$ *is a compact edge and that*

$$\partial_2^j \phi^a_{\mathrm{pr}}(1,0)=0 \quad \textit{for every } 1\leqslant j\leqslant h.$$

*If the neighborhood* $\Omega$ *of the point* $(0,0)$ *and* $\varepsilon_0$ *are chosen sufficiently small, then the maximal operator* $\mathcal{M}^{\varrho_0}$ *is bounded on* $L^p(\mathbb{R}^3)$ *for every* $p>h$.

Since the proof of this proposition will be entirely based on the oscillatory integral estimates from §5, we recall some notation and observations from §6:

$\mathcal{M}^{\varrho_0}$ is the maximal operator associated with the family of Fourier multipliers $e^{it\xi_3}J(t\xi)$, $t>0$, with

$$J(\xi):=\int_{\mathbb{R}^2_+} e^{i(\xi_1x_1+\xi_2\psi(x_1)+\xi_2x_2+\xi_3\phi^a(x))}\varrho\biggl(\frac{x_2}{\varepsilon_0x_1^a}\biggr)\eta(x)\,dx.$$

At this point it will be convenient to *defray our notation* by writing $\phi$ in place of $\phi^a$, and $\varkappa$ in place of $\varkappa^a=\varkappa^\lambda$, and $x$ for the adapted coordinates $y$. Then

$$J(\xi):=\int_{\mathbb{R}^2_+} e^{i(\xi_1x_1+\xi_2\psi(x_1)+\xi_2x_2+\xi_3\phi(x))}\varrho\biggl(\frac{x_2}{\varepsilon_0x_1^a}\biggr)\eta(x)\,dx, \tag{9.2}$$

where $\varrho$ is a smooth bump function as before, and where the following assumptions are satisfied:

*Assumptions* 9.2. The functions $\phi$, $\psi$ and $\eta$ are smooth and have the following properties:

(i) $\psi(x_1)=b_1x_1^{m_1}+x_1^{m_1+1}q(x_1)$, where $b_1\in\mathbb{R}\setminus\{0\}$ and where $q$ is smooth;

(ii) $\phi$ is of finite type, $\phi(0)=0$ and $\nabla\phi(0)=0$;

(iii) the coordinates $x$ are adapted to $\phi$, and $h=h(\phi)=d(\phi)\geqslant 2$;

(iv) the principal face $\pi(\phi)$ is a compact edge, and the associated principal part $\phi_{\mathrm{pr}}$ of $\phi$ is $\varkappa$-homogeneous of degree 1, where $0<\varkappa_1<\varkappa_2<1$ and $a=\varkappa_2/\varkappa_1>m_1\geqslant 2$;

(v) $\eta$ is a smooth bump function supported in a sufficiently small neighborhood $\Omega$ of the origin.

Moreover,

$$\partial_2^j\phi_{\mathrm{pr}}(1,0)=0 \quad \text{for every } 1\leqslant j\leqslant h. \tag{9.3}$$

Notice that the domain (9.1) now corresponds to the domain

$$|x_2|\leqslant \varepsilon_0x_1^a. \tag{9.4}$$

**9.1. Further domain decompositions under hypothesis** (9.3)

We first observe that our assumptions imply that $\phi_{\mathrm{pr}}(1,0)\neq 0$. For, otherwise by (9.3) $x_2=0$ would be a root of multiplicity $N>h$. On the other hand, since the coordinates $x$ are adapted to $\phi$, we must have $N\leqslant h$, a contradiction.

We can thus write

$$\phi_{\mathrm{pr}}(x_1,x_2)=x_2^B Q(x_1,x_2)+cx_1^n, \quad \text{with } c\neq 0,$$

where $B\geqslant 1$, and where $Q$ is a $\varkappa$-homogeneous polynomial such that $Q(x_1,0)=bx_1^q$, $b\neq 0$, so that $Q(x_1,0)\neq 0$ for $x_1>0$. Without loss of generality, we shall assume that $c=1$. Notice that $B\geqslant 2$, since $\partial_2\phi_{\mathrm{pr}}(1,0)=0$, and then our assumption (9.3) implies that in fact

$$B>h\geqslant 2. \tag{9.5}$$

We also remark that $n=1/\varkappa_1>\varkappa_2/\varkappa_1>m_1$.

In order to understand the behavior of $\phi$ as a function of $x_2$, for $x_1$ fixed, we shall decompose

$$\phi(x_1,x_2)=\phi(x_1,0)+\theta(x_1,x_2), \tag{9.6}$$

and write the complete phase $\Phi$ for $J(\xi)$ in the form

$$\Phi(x,\xi)=(\xi_3\phi(x_1,0)+\xi_1x_1+\xi_2\psi(x_1))+(\xi_3\theta(x_1,x_2)+\xi_2x_2). \tag{9.7}$$

Notice that

$$\phi(x_1,0)=x_1^n(1+O(x_1)), \quad \psi(x_1)=b_1x_1^{m_1}(1+O(x_1)) \quad \text{and} \quad \theta_\varkappa(x_1,x_2)=x_2^BQ(x_1,x_2), \tag{9.8}$$

where $\theta_\varkappa$ denotes the $\varkappa$-principal part of $\theta$.

Now, by means of a dyadic decomposition and rescaling using the $\varkappa$-dilations $\{\delta_r\}_{r>0}$ we would like to reduce our considerations as before to the domain where $x_1\sim 1$. In this domain, $|x_2|\ll 1$, so that $\theta_\varkappa(x)\sim x_2^BQ(x_1,0)$. What leads to problems is that the "error term" $\theta_{\varkappa,r}:=\theta-\theta_\varkappa$, which consists of terms of higher $\varkappa$-degree than $\theta_\varkappa$, may nevertheless contain terms of lower $x_2$-degree $l_j<B$ of the form $c_jx_2^{l_j}x_1^{n_j}$, provided $n_j$ is sufficiently large. After scaling the $k$th dyadic piece in our decomposition by $\delta_{2^{-k}}$ in order to achieve that $x_1\sim 1$ and $|x_2|\lesssim\varepsilon_0$, such terms will have small coefficients compared to the one of $x_2^BQ(x_1,x_2)$, but for $|x_2|$ very small they may nevertheless become dominant and have to be taken into account.

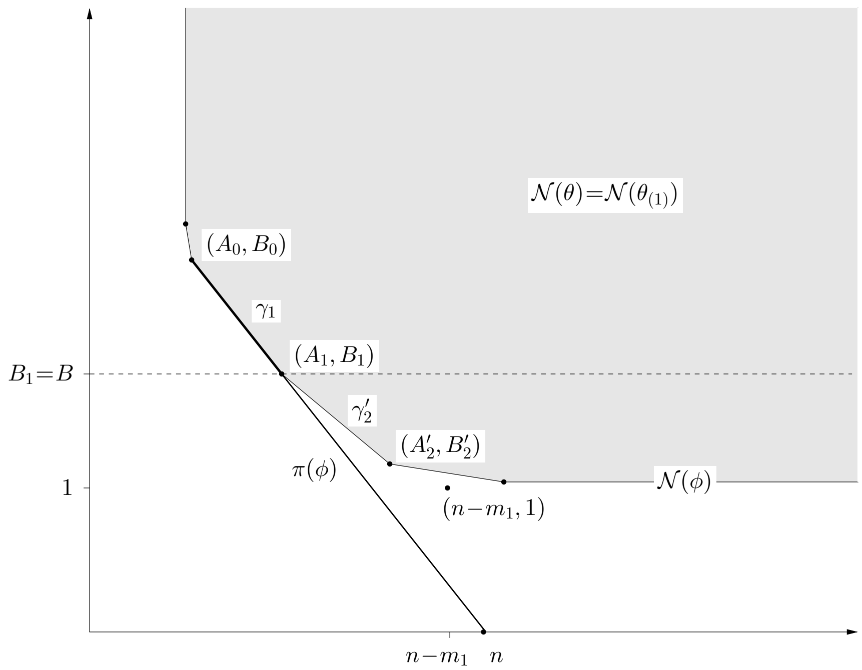


Figure 1.

**9.1.1. Outline of the stopping time algorithm**

To deal with this problem, consider the Newton polyhedron $\mathcal{N}(\theta)$. Since the Taylor support $\mathcal{T}(\theta)$ arises from $\mathcal{T}(\phi)$ by removing all points on the $t_1$-axis, we have

$$\mathcal{N}(\partial_2\theta) = (0, -1) + \mathcal{N}(\theta). \tag{9.9}$$

Let us put

$$\varkappa^1 := \varkappa \quad \text{and} \quad a_1 := a = \frac{\varkappa_2^1}{\varkappa_1^1}.$$

Then, by (9.8), the point $(q, B)$ is the right endpoint $(A_1, B_1)$ of the face

$$\gamma_1 := [(A_0, B_0), (A_1, B_1)]$$

of the Newton polyhedron $\mathcal{N}(\theta)$ of $\theta$ lying on the supporting line $\varkappa_1^1 t_1 + \varkappa_2^1 t_2 = 1$. Note that possibly $(A_0, B_0) = (A_1, B_1)$.

It is also clear from the construction of $\theta$ that

$$\mathcal{N}(\theta) \cap \{(t_1, t_2) : t_2 \geqslant B_1\} = \mathcal{N}(\phi) \cap \{(t_1, t_2) : t_2 \geqslant B_1\}. \tag{9.10}$$

We next describe a stopping time algorithm oriented at the level sets of $\partial_2\theta$ which will decompose our domain (9.4) in a finite number of steps into subdomains, whose contributions to our maximal operator will be treated in different ways in the subsequent subsections. Some features of this algorithm will resemble Varchenko's algorithm (see §2.5), and it will stop at latest when we have reached a domain containing only one root of $\partial_2\theta$ (up to multiplicity).

*Case* A. $\mathcal{N}(\theta)\subset\{(t_1,t_2):t_2\geqslant B_1\}$.

Then no term in $\theta$ has lower $x_2$-exponent than $B_1=B$, and we stop at this point.

*Case* B. $\mathcal{N}(\theta)$ contains a point below the line $t_2=B_1$.

Then the Newton diagram $\mathcal{N}_d(\theta)$ will contain a further edge

$$\gamma_2'=[(A_1,B_1),(A_2',B_2')]$$

below the line $t_2=B_1$, lying, say, on the line $\varkappa_1^2t_1+\varkappa_2^2t_2=1$ (see Figure 1). We then put

$$\varkappa^2:=(\varkappa_1^2,\varkappa_2^2)\quad\text{and}\quad a_2:=\frac{\varkappa_2^2}{\varkappa_1^2},\quad\text{where clearly } a_2>a_1.$$

Notice that $a_2\in\mathbb{Q}$. We then decompose the domain (9.4) into the domains

$$E_1:=\{(x_1,x_2):N_2x_1^{a_2}<|x_2|\leqslant\varepsilon_1x_1^{a_1}\}$$

(a domain of "type E") and

$$H_2:=\{(x_1,x_2):|x_2|\leqslant N_2x_1^{a_2}\},$$

where $N_2$ will be any sufficiently large constant and $\varepsilon_1:=\varepsilon_0$.

In the domain $E_1$, which is again a domain of transition between two different homogeneities, we stop our algorithm. It will be treated later be means of bi-dyadic decompositions.

The $\varkappa^2$-homogeneous domain $H_2$ will be further decomposed as follows: We first notice that the $\varkappa^2$-homogeneous part $(\partial_2\theta)_{\varkappa^2}$ of $\partial_2\theta$ will be associated with the edge

$$(0,-1)+\gamma_2'=[(A_1,B_1-1),(A_2',B_2'-1)]$$

of the Newton diagram of $\partial_2\theta$ and it is $\varkappa^2$-homogeneous of degree $1-\varkappa_2^2$. Observe also that, in view of (9.9), we have that $(\partial_2\theta)_{\varkappa^2}=\partial_2(\theta_{\varkappa^2})$. Decomposing the polynomial $t\mapsto(\partial_2\theta)_{\varkappa^2}(1,t)$ into linear factors and making use of the $\varkappa^2$-homogeneity of $(\partial_2\theta)_{\varkappa^2}$, we see that we can write

$$(\partial_2\theta)_{\varkappa^2}(x)=c_2x_1^{A_1}x_2^{B_2'-1}\prod_\alpha(x_2-c_2^\alpha x_1^{a_2})^{n_2^\alpha},$$

where

$$B_1 = B_2' + \sum_\alpha n_2^\alpha \quad \text{and} \quad A_2' = A_1 + a_2 \sum_\alpha n_2^\alpha,$$

with roots $c_2^\alpha \in \mathbb{C}\setminus\{0\}$ and multiplicities $n_2^\alpha \geqslant 1$. Let us assume in the sequel that

$$N_2 \gg \max_\alpha |c_2^\alpha|.$$

By $R_2$ we shall denote the set of all real roots $c_2^\alpha \in \mathbb{R}$ of $(\partial_2\theta)_{\varkappa^2}(1,t)$, where we include also the trivial root 0 in the case where $B_2'-1>0$.

We shall need to understand the behavior of the complete phase function $\Phi(x,\xi)$ in display (9.7) on the domain $H_2$. Now, after dyadic decomposition with respect to the $\varkappa^2$-dilations and rescaling, we have to look at $\Phi(2^{-\varkappa_1^2 k}x_1, 2^{-\varkappa_2^2 k}x_2, \xi)$ in the domain where $x_1 \sim 1$ and, say, $|x_2| \leqslant N_2$. We write

$$\Phi(2^{-\varkappa_1^2 k}x_1, 2^{-\varkappa_2^2 k}x_2, \xi) = 2^{-\varkappa_1^2 nk}\lambda \Phi_k(x,s),$$

where

$$\begin{aligned}\Phi_k(x,s) := {} & x_1^n(1+v_k(x_1)) + s_1x_1 + S_2 b_1 x_1^{m_1}(1+w_k(x_1)) \\ & + 2^{(\varkappa_1^2 n-1)k}(\theta_{\varkappa^2}(x_1,x_2) + \theta_{r,k}(x_1,x_2) + s_2x_2)\end{aligned}$$

and again $\lambda := \xi_3$ (assumed to be positive),

$$s_1 := 2^{\varkappa_1^2(n-1)k}\frac{\xi_1}{\lambda}, \quad s_2 := 2^{(1-\varkappa_2^2)k}\frac{\xi_2}{\lambda} \quad \text{and} \quad S_2 := 2^{\varkappa_1^2(n-m_1)k}\frac{\xi_2}{\lambda} = 2^{(\varkappa_1^2(n-m_1)+\varkappa_2^2-1)k}s_2.$$

The functions $v_k, w_k$ and $\theta_{r,k}$ are of order $O(2^{-\delta k})$ in $C^\infty$ for some $\delta>0$.

In the estimation of the corresponding oscillatory integral, the worst possible case arises for the $x_1$-integration when $|s_1| \sim |S_2| \sim 1$, so that

$$|s_2| \sim 2^{-(\varkappa_1^2(n-m_1)+\varkappa_2^2-1)k}. \tag{9.11}$$

Fix now an arbitrary $\varepsilon_2>0$. For any point $d$ in the interval $[-N_2, N_2]$ denote by $D_2(d)$ the $\varkappa^2$-homogeneous domain (inside the half-plane $x_1>0$)

$$D_2(d) := \{(x_1,x_2) : |x_2 - dx_1^{a_2}| \leqslant \varepsilon_2 x_1^{a_2}\}$$

(a domain of "type D").

Since we can cover the domain $H_2$ by a finite number of such domains $D_2(d)$, it will be sufficient to examine the contribution of each of the domains $D_2(d)$.

*Case* B (a). If $\varkappa_1^2(n-m_1)+\varkappa_2^2\leqslant 1$, then we have $|s_2|\geqslant c>0$ in (9.11). In this case, it will be possible to control the corresponding oscillatory integrals if $\varepsilon_2$ is chosen sufficiently small, and we shall stop our algorithm with the domains $D_2(d)$.

Indeed, if $\varkappa_1^2(n-m_1)+\varkappa_2^2<1$, then $|s_2|\gg 1$, which will allow for an integration by parts with respect to $x_2$ in a similar way as we argued in the first case of the proof of Proposition 6.3.

The worst possible case will actually arise when $\varkappa_1^2(n-m_1)+\varkappa_2^2=1$ and when in addition $d\notin R_2$, i.e., when $(\partial_2\theta)_{\varkappa^2}(1,d)\neq 0$, which will indeed lead the *"degenerate Airy-type"* integrals of §5.3.

*Case* B (b). Assume that $\varkappa_1^2(n-m_1)+\varkappa_2^2>1$, so that $|s_2|\ll 1$ in (9.11).

(i) If $d\notin R_2$, then $(\partial_2\theta)_{\varkappa^2}(1,d)\neq 0$ and $|s_2|\ll 1$, so that again one can integrate by parts with respect to $x_2$, and again the algorithm will stop.

(ii) Finally assume that $d\in R_2$, so that $(\partial_2\theta)_{\varkappa^2}(1,d)=0$ and $|s_2|\ll 1$. In this case, we introduce new coordinates

$$y_1:=x_1,\quad y_2:=x_2-dx_1^{a_2},$$

and denote our original functions, when expressed in the new coordinates $y$, by a subscript "(2)", e.g.,

$$\phi_{(2)}(y):=\phi(y_1,y_2+dy_1^{a_2}).$$

Correspondingly, we define $\theta_{(2)}$ by

$$\phi_{(2)}(y_1,y_2)=\phi_{(2)}(y_1,0)+\theta_{(2)}(y),$$

and so on. Observe that in general we will not have $\theta_{(2)}(y)=\theta(y_1,y_2+dy_1^{a_2})$, but it is true that

$$\partial_2\theta_{(2)}(y)=\partial_2\theta(y_1,y_2+dy_1^{a_2}).$$

Notice that this $\varkappa^2$-homogeneous change of coordinates will have the effect on the Newton(–Puiseux) polyhedron that the edge $\gamma_2'=[(A_1,B_1),(A_2',B_2')]$ of $\mathcal{N}(\theta)$ on the line $\varkappa_1^2t_1+\varkappa_2^2t_2=1$ will be turned into a face

$$\gamma_2=[(A_1,B_1),(A_2,B_2)]$$

of $\mathcal{N}(\theta_{(2)})$ on the same line, with the same left endpoint $(A_1,B_1)$ but possibly different right endpoint $(A_2,B_2)$ (which may even agree with the left endpoint), where still $B_2\geqslant 1$.

Notice that $B_1\geqslant B_2$, and that the domain $D_2(d)$ corresponds to the domain where $|y_2|\leqslant\varepsilon_2 y_1^{a_2}$ in the new coordinates $y$.

In case B (b) (ii), which is the only one where our algorithm did not stop, we see that by passing from $\phi=:\phi_{(1)}$ to $\phi_{(2)}$ and denoting the new coordinates $y$ again by $x$, we have thus reduced ourselves to the smaller $\varkappa^2$-homogeneous domain

$$|x_2| \leqslant \varepsilon_2 x_1^{a_2}$$

in place of (9.4).

We observe also that since the $\varkappa=\varkappa^1$-homogenous part of our change of coordinates $y_1=x_1$, $y_2=x_2-dx_1^{a_2}$ is given by $x_1$, $x_2$, i.e., by the identity mapping, the Newton polyhedra of $\theta_{(1)}$ and $\theta_{(2)}$ will have the same $\varkappa^1$-principal faces and corresponding principal parts. This implies in particular that still

$$\phi_{(2)}(x_1,0)=x_1^n(1+O(x_1^{1/r}))$$

for some rational exponent $r>0$. Moreover, since $a_2>a_1>m_1$, also the new function $\psi_{(2)}(x_1):=\psi(x_1)+dx_1^{a_2}$, which corresponds to $\psi$ in the new coordinates, will still satisfy

$$\psi_{(2)}(x_1)=b_1x_1^{m_1}(1+O(x_1^{1/r})).$$

Replacing $\phi=\phi_{(1)}$ by $\phi_{(2)}$, $\varkappa^1$ by $\varkappa^2$ and $B_1$ by $B_2$, we can now iterate this procedure. Notice that already the function $\phi_{(2)}$ will in general be a smooth function of $x_2$ and some fractional power of $x_1$ only, so that from here on we shall have to work with Newton–Puiseux polyhedra in place of Newton polyhedra, etc.

*Example* 9.3. Let $\phi(x_1,x_2):=x_1^n+x_2^B+x_2x_1^{n-m_1}$ and $\psi(x_1):=x_1^{m_1}$, where we assume that $B>2$ and $n/B>m_1\geqslant 2$. Then the coordinates $(x_1,x_2)$ are adapted to $\phi$. Here we have

$$\phi(x_1,0)=x_1^n, \quad \theta(x)=x_2^B+x_2x_1^{n-m_1} \quad \text{and} \quad \theta_{\varkappa^1}(x)=\theta_{\varkappa}(x)=x_2^B,$$

whereas

$$\theta_{\varkappa^2}(x)=x_2^B+x_2x_1^{n-m_1}.$$

Since obviously $\varkappa_1^2(n-m_1)+\varkappa_2^2=1$ and $(\partial_2\theta)_{\varkappa^2}(1,0)=\partial_2(\theta_{\varkappa^2})(1,0)\neq 0$, we see that the "degenerate Airy-type" situation described in case B (a) applies to $d:=0$.

### 9.1.2. Details on and modification of the algorithm

If we have applied this procedure $L-1$ times (where $L\geqslant 2$), we are left with a finite number of domains of type D within the half-plane $x_1>0$ of the form

$$|x_2-r(x_1)| \leqslant \varepsilon_L x_1^{a_{j_0\cdots j_{L-2}}} \tag{9.12}$$

(in our original coordinates from Assumption 9.2), where $\varepsilon_L$ can be chosen small and where $r$ is of the form

$$r(x_1)=d_{j_1}x_1^{a_{j_0}}+d_{j_1j_2}x_1^{a_{j_0j_1}}+...+d_{j_1...j_{L-1}}x_1^{a_{j_0...j_{L-2}}},$$

with $j_0:=2$. Here, the exponents are rational and satisfy $a_2=a_{j_0}<a_{j_0j_1}<...<a_{j_0...j_{L-2}}$, and each $d_{j_1...j_l}$ is a real root which has been chosen in the $l$th step.

The complement of the union of these domains has been decomposed into domains of type E and D on which the algorithm had stopped in a previous step.

In order to defray the notation, let us fix one of these functions $r$ and its corresponding domain (9.12), and then write

$$r(x_1)=\sum_{j=2}^{L}d_jx_1^{a_j},$$

with $a_2<a_3...<a_L$. We are thus looking at the domain

$$|x_2-r(x_1)|\leqslant\varepsilon_Lx_1^{a_L}. \tag{9.13}$$

Correspondingly, in this way we will have recursively constructed a sequence

$$\phi=\phi_{(1)},\quad \phi_{(2)},\quad ...,\quad \phi_{(L)}$$

of functions, where $\phi_{(l)}$ is obtained from $\phi_{(l-1)}$, for $l\geqslant 2$, by means of a change of coordinates $y_1:=x_1$, $y_2:=x_2-d_lx_1^{a_l}$. Then $\phi_{(l)}$ arises from $\phi$ by the total change of coordinates $x=s_{(l)}(y)$, where

$$y_1:=x_1 \quad\text{and}\quad y_2:=x_2-\sum_{j=2}^{l}d_jx_1^{a_j},$$

i.e., $\phi_{(l)}=\phi\circ s_{(l)}$, and correspondingly $\theta_{(l)}$ is defined by

$$\phi_{(l)}(y_1,y_2)=\phi_{(l)}(y_1,0)+\theta_{(l)}(y),$$

and so on. Notice that in general we do not have $\theta_{(l)}=\theta\circ s_{(l)}$, but

$$\partial_2\theta_{(l)}=\partial_2\theta\circ s_{(l)}=\partial_2\phi\circ s_{(l)}.$$

For the functions $\phi_{(l)}(x_1,0)$ and $\psi_{(l)}(x_1)=\psi(x_1)+\sum_{j=2}^{l}d_jx_1^{a_j}$ we then still have

$$\phi_{(l)}(x_1,0)=x_1^n(1+O(x_1^{1/r}))\quad\text{and}\quad\psi_{(l)}(x_1)=b_1x_1^{m_1}(1+O(x_1^{1/r})) \tag{9.14}$$

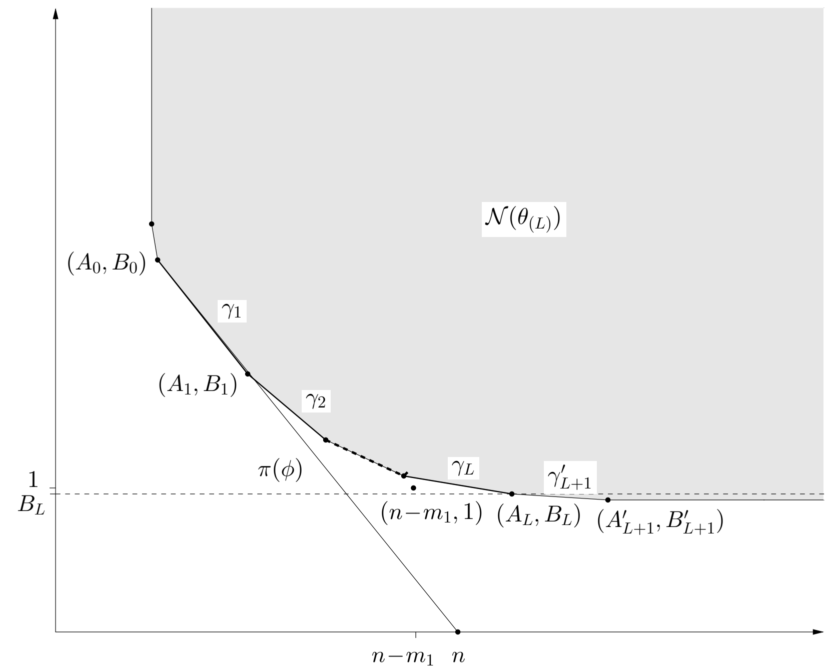


Figure 2.

for some rational exponent $r>0$.

To be more precise, for each choice of $r(x_1)$ our algorithm will produce in the $(l-1)$-st step domains of type D of the form

$$D_l(d) := \left\{ (x_1, x_2) : \left| x_2 - \sum_{j=2}^{l-1} d_j x_1^{a_j} - d x_1^{a_l} \right| \leqslant \varepsilon_l x_1^{a_l} \right\},$$

and a transition domain of type E of the form

$$E_{l-1} := \left\{ (x_1, x_2) : N_l x_1^{a_l} < \left| x_2 - \sum_{j=2}^{l-1} d_j x_1^{a_j} \right| \leqslant \varepsilon_{l-1} x_1^{a_{l-1}} \right\},$$

when expressed in the coordinates $x$ of Assumption 9.2. It will stop on $E_{l-1}$.

As for the domains $D_l(d)$, notice that like in passing from $\theta=\theta_{(1)}$ to $\theta_{(2)}$, in each step when passing from $\theta_{(l-1)}$ to $\theta_{(l)}$ we replace a face $\gamma'_l = [(A_{l-1}, B_{l-1}), (A'_l, B'_l)]$ of the Newton–Puiseux diagram $\mathcal{N}_d(\theta_{(l-1)})$ by a new face $\gamma_l = [(A_{l-1}, B_{l-1}), (A_l, B_l)]$ lying on the same line as $\gamma'_l$ (possibly degenerating to a single point) of $\mathcal{N}(\theta_{(l)})$, so that the Newton–Puiseux diagram $\mathcal{N}_d(\theta_{(l)})$ will in particular possess the faces

$$\gamma_1 = [(A_0, B_0), (A_1, B_1)], \quad \ldots, \quad \gamma_l = [(A_{l-1}, B_{l-1}), (A_l, B_l)],$$

where $B_l \geqslant 1$ (see Figure 2). Indeed, the Newton–Puiseux diagram $\mathcal{N}_d(\theta_{(l-1)})$ will have, besides the edges $\gamma_1, ..., \gamma_{l-1}$, in addition an edge $\gamma_l' = [(A_{l-1}, B_{l-1}), (A_l', B_l')]$ lying on a unique line

$$\varkappa_1^l t_1 + \varkappa_2^l t_2 = 1,$$

where $a_l = \varkappa_2^l / \varkappa_1^l$. Otherwise, $\mathcal{N}(\theta_{(l-1)})$ would be contained in the half-plane $t_2 \geqslant B_{l-1}$ and the algorithm would have stopped earlier.

Finally, if $\partial_2(\theta_{(l-1)})_{\varkappa^l}$ denotes the $\varkappa^l$-homogeneous principal part of $\partial_2(\theta_{(l-1)})$ corresponding to the edge $\gamma_l'$, then the algorithm will also stop on $D_l(d)$, unless $d$ is a real root $d_l$ of $\partial_2(\theta_{(l-1)})_{\varkappa^l}(1, \cdot)$ and $\varkappa_1^l (n - m_1) + \varkappa_2^l > 1$.

If the latter is the case, then we are left with the domain $D_l(d_l)$ at the end of step $l-1$ (where $d_l$ is one of possibly several roots). When expressed in the coordinates $y$ defined by the change of coordinates $s_{(l)}$, it is $\varkappa^l$-homogenous and given by $|y_2| \leqslant \varepsilon_l x_1^{a_l}$.

We note that

$$B_1 \geqslant B_2 \geqslant ... \geqslant B_l \geqslant 1 \quad \text{and} \quad m_1 < a = a_1 < a_2 < ... < a_l. \tag{9.15}$$

In particular, the descending sequence $\{B_l\}_l$ must eventually become constant (unless our algorithm stops already earlier).

Our algorithm will always stop after a finite number of steps, since eventually we will have $\varkappa_1^l (n - m_1) + \varkappa_2^l \leqslant 1$ because $a_l \to \infty$.

This is rather evident from the geometry of the Newton–Puiseux polyhedra $\mathcal{N}(\theta_{(l)})$, but let us give a precise argument:

*We first claim that there is some fixed rational number $1/r$ such that every $a_l$ is a multiple of $1/r$.*

In the analytic case, this follows easily from the Puiseux series expansions of roots of $\partial_2 \theta$, but we can give a more direct argument.

The polynomial $\partial_2(\theta_{(l-1)})_{\varkappa^l}$ is $\varkappa^l$-homogeneous of degree 1, and so we may write it as

$$\partial_2(\theta_{(l-1)})_{\varkappa^l}(x) = c x_1^{A_{l-1}} x_2^{B_l' - 1} \prod_\alpha (x_2 - c_l^\alpha x_1^{a_l})^{n_l^\alpha}.$$

Since the $x_2$-degree of this polynomial is bounded by $B_0$, putting $N := \sum_\alpha n_l^\alpha$, this implies that $a_l \in (1/N)\mathbb{N}$, where $N \leqslant B_0$, which implies the claim. In combination with (9.15), this proves that $a_l \to \infty$.

Assume now that our algorithm did not terminate. Then we could find some minimal $L \geqslant 1$ such that $B_l = B_L$ for every $l \geqslant L$. This implies that $B_L \geqslant 2$, since for $B_L = 1$ we had $\mathcal{N}_d(\theta_{(L)}) \subset \{(t_1, t_2) : t_2 \geqslant B_L\}$, and we would stop. Moreover, from $1 = \varkappa_1^l A_l + \varkappa_2^l B_l \geqslant 2\varkappa_2^l$

we conclude that $\varkappa_2^l \leqslant \frac{1}{2}$. But then clearly

$$\varkappa_1^l(n-m_1)+\varkappa_2^l=\varkappa_2^l\left(1+\frac{n-m_1}{a_l}\right)\leqslant\frac{1}{2}\left(1+\frac{n-m_1}{a_l}\right)\leqslant 1$$

for $l$ sufficiently large, and so our algorithm would eventually stop, contradicting our assumption.

This argument also shows that the number of steps $L_0-1$ after which the algorithm will stop everywhere in $\Omega$ can be chosen to be independent of the choice of roots $d_j$ along the way. The domains of type E and D that we shall produce by this algorithm up to step $L_0-1$ will then cover $\Omega$, so that it will suffice to study the contributions to our maximal operator of each of these domains.

Assume from now on that we have fixed a choice of roots $d_j$ when passing along our algorithm. We then choose $L\geqslant 2$ with $L\leqslant L_0$ so that the algorithm will stop for this choice of roots at step $L-1$.

*Next, in the case where* $B_l=B_{l+1}=...=B_{l+j}$ *for some* $j\geqslant 1$, *we will modify our stopping time argument as follows*:

We shall skip the intermediate steps and pass from $\phi_{(l)}$ to $\phi_{(l+j)}$ directly, decomposing in the passage from $\phi_{(l)}$ to $\phi_{(l+j)}$ the domain $\{(x_1,x_2):|x_2|\leqslant\varepsilon_l x_1^{a_l}\}$ into the bigger transition domain

$$E_l':=\{(x_1,x_2):N_{l+j}x_1^{a_{l+j}}<|x_2|\leqslant\varepsilon_l x_1^{a_l}\}$$

and the $\varkappa^{l+j}$-homogeneous domain

$$H_{l+j}':=\{(x_1,x_2):|x_2|\leqslant N_{l+j}x_1^{a_{l+j}}\},$$

where $N_{l+j}$ will be any sufficiently large constant.

We may and shall therefore assume that the sequence $\{B_l\}_l$ is strictly decreasing.

Now, if the domain on which we stop our algorithm is of type E, then it is a transition domain

$$E_l:=\left\{(x_1,x_2):N_{l+1}x_1^{a_{l+1}}<\left|x_2-\sum_{j=2}^{l}d_jx_1^{a_j}\right|\leqslant\varepsilon_l x_1^{a_l}\right\},\quad 1\leqslant l\leqslant L,$$

when expressed in the coordinates $x$ of Assumption 9.2, where the case $l=L$ arises only if $\mathcal{N}(\theta_{(L)})$ is not contained in $\{(t_1,t_2):t_2\geqslant B_L\}$.

When $\mathcal{N}(\theta_{(L)})\subset\{(t_1,t_2):t_2\geqslant B_L\}$, so that our algorithm stops because of this inclusion (cf. the discussion of case A in §9.1.1, only with $\theta=\theta_{(1)}$ replaced by $\theta_{(L)}$), then the algorithm will stop on the whole domain where

$$\left|x_2-\sum_{j=2}^{L}d_jx_1^{a_j}\right|\leqslant\varepsilon_L x_1^{a_L}$$

which remained after the previous step. We shall then replace $E_L$ by the "generalized" transition domain (which is at the same time $\varkappa^L$-homogeneous)

$$E'_L := \left\{ (x_1, x_2) : \left| x_2 - \sum_{j=2}^{L} d_j x_1^{a_j} \right| \leqslant \varepsilon_L x_1^{a_L} \right\},$$

where formally $a_{L+1}=\infty$.

And, if we stop on a domain of type D, then it is of the form

$$D_{l+1}(d) := \left\{ (x_1, x_2) : \left| x_2 - \sum_{j=2}^{l} d_j x_1^{a_j} - d x_1^{a_{l+1}} \right| \leqslant \varepsilon_{l+1} x_1^{a_{l+1}} \right\}, \quad 1 \leqslant l \leqslant L,$$

which is $\varkappa^{l+1}$-homogeneous after applying the change of coordinates $x=s_{(l)}(y)$, where $|d|\leqslant N_{l+1}$, and where

$$\varkappa_1^{l+1}(n-m_1)+\varkappa_2^{l+1} \leqslant 1,$$

in the case where $d=d_{l+1}$ is a real root of $\partial_2(\theta_{(l)})_{\varkappa^{l+1}}(1,\cdot)$ (cf. the discussion of cases B (a) and (b) in §9.1.1).

The case $l=L$ can here only arise if $\mathcal{N}(\theta_{(L-1)})$ is not contained in $\{(t_1,t_2):t_2\geqslant B_L\}$; moreover, in this case there is no real root of $\partial_2(\theta_{(L)})_{\varkappa^{L+1}}(1,\cdot)$ if $\varkappa_1^{L+1}(n-m_1)+\varkappa_2^{L+1}>1$, since otherwise the algorithm would not stop at step $L-1$.

The contribution to the oscillatory integral $J(\xi)$ of a domain $E_l$, after changing to the coordinates $y$ given by $s_{(l)}$ in the integral, can be put into the form

$$J^{\tau_l}(\xi) := \int_{\mathbb{R}_+^2} e^{i\Phi_{(l)}(y,\xi)} \tilde{\eta}(y) \tau_l(y)\, dy,$$

where $\tilde{\eta}$ is again a smooth bump function supported near the origin,

$$\Phi_{(l)}(y,\xi) := (\xi_3 \phi_{(l)}(y_1,0)+\xi_1 y_1+\xi_2 \psi_{(l)}(y_1))+(\xi_3 \theta_{(l)}(y_1,y_2)+\xi_2 y_2),$$

and where

$$\tau_l(y) := \varrho\left(\frac{y_2}{\varepsilon_l y_1^{a_l}}\right)(1-\varrho)\left(\frac{y_2}{N_{l+1} y_1^{a_{l+1}}}\right),$$

if $\mathcal{N}(\theta_{(l)})$ is not contained in $\{(t_1,t_2):t_2\geqslant B_l\}$, and

$$\tau_l(y) := \varrho\left(\frac{y_2}{\varepsilon_l y_1^{a_l}}\right),$$

if $\mathcal{N}(\theta_{(l)})\subset\{(t_1,t_2):t_2\geqslant B_l\}$ and $l=L$. As we have seen, the latter case corresponds to the domain $E'_L$.

Similarly, the contribution of a domain $D_{l+1}(d)$ is of the form

$$J^{\varrho_{l+1}}(\xi) := \int_{\mathbb{R}_+^2} e^{i\Phi_{(l)}(y,\xi)}\tilde{\eta}(y)\varrho_{l+1}(y_1, y_2 - dy_1^{a_{l+1}})\,dy,$$

where

$$\varrho_{l+1}(y) := \varrho\biggl(\frac{y_2}{\varepsilon_{l+1}y_1^{a_{l+1}}}\biggr).$$

At this point, it will again be helpful to defray the notation by writing $\phi$ in place of $\phi_{(l)}$, $\theta$ in place of $\theta_{(l)}$, $\psi$ in place of $\psi_{(l)}$, etc., and assuming that $\phi$, $\psi$ and $\theta$ satisfy the following assumptions on $\mathbb{R}_+^2$:

*Assumptions* 9.4. The functions $\phi$ and $\eta$ are smooth functions of $x_1^{1/r}$ and $x_2$, and $\psi$ is a smooth function of $x_1^{1/r}$, where $r$ is a positive integer. If we write

$$\phi(x_1, x_2) = \phi(x_1, 0) + \theta(x_1, x_2),$$

then the following hold true:

(i) the Newton–Puiseux diagram $\mathcal{N}_d(\theta)$ contains at least the faces

$$\gamma_1 = [(A_0, B_0), (A_1, B_1)], \quad ..., \quad \gamma_l = [(A_{l-1}, B_{l-1}), (A_l, B_l)],$$

where $B_1 > B_2 > ... > B_l$, so that $\gamma_j$ is an edge, if $j > 1$, and $B_1 > h \geqslant 2$, and in the case where $\mathcal{N}(\theta)$ is not contained in $\{(t_1, t_2) : t_2 \geqslant B_l\}$, it contains the additional edge

$$\gamma'_{l+1} = [(A_l, B_l), (A'_{l+1}, B'_{l+1})].$$

The face $\gamma_j$ lies on the line $\varkappa_1^j t_1 + \varkappa_2^j t_2 = 1$, where $\varkappa^1 = \varkappa$. Putting $a_j := \varkappa_2^j / \varkappa_1^j$, we have

$$a = a_1 < ... < a_j < a_{j+1} < ... .$$

(ii) We have

$$\phi(x_1, 0) = x_1^n (1 + O(x_1^{1/r})) \quad \text{and} \quad \psi(x_1) = b_1 x_1^{m_1}(1 + O(x_1^{1/r})),$$

where $n = 1/\varkappa_1 > \varkappa_2/\varkappa_1 = a > m_1 \geqslant 2$.

With these data, we define the phase function

$$\Phi(x, \xi) := (\xi_3\phi(x_1, 0) + \xi_1 x_1 + \xi_2\psi(x_1)) + (\xi_3\theta(x_1, x_2) + \xi_2 x_2),$$

and the oscillatory integrals

$$J^{\tau_l}(\xi) := \int_{\mathbb{R}_+^2} e^{i\Phi(x,\xi)}\eta(x)\tau_l(x_1, x_2)\,dx$$

and

$$J^{\varrho_{l+1}}(\xi):=\int_{\mathbb{R}^2_+} e^{i\Phi(x,\xi)}\eta(x)\varrho_{l+1}(x_1,x_2-dx_1^{a_{l+1}})\,dx,$$

where again $\eta$ denotes a smooth bump function supported in a sufficiently small neighborhood $\Omega$ of the origin, and $\tau_l$ and $\varrho_{l+1}$ are defined as before.

The maximal operators corresponding to the Fourier multipliers $e^{i\xi_3}J^{\tau_l}$ and $e^{i\xi_3}J^{\varrho_{l+1}}$ will again be denoted by $\mathcal{M}^{\tau_l}$ and $\mathcal{M}^{\varrho_{l+1}}$, respectively.

In view of our previous discussion, and since we have $h(\phi_{(l)})=1/|\varkappa|=h$ for every $l$, what remains to be proven is the following result.

PROPOSITION 9.5. *Assume that the neighborhood $\Omega$ of the point $(0,0)$, and $\varepsilon_l$ and $\varepsilon_{l+1}$ are chosen sufficiently small and $N_{l+1}$ sufficiently large. Then the following hold true*:

(a) *The maximal operator $\mathcal{M}^{\tau_l}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>1/|\varkappa|$;*

(b) *The maximal operator $\mathcal{M}^{\varrho_{l+1}}$ is bounded on $L^p(\mathbb{R}^3)$ for every $p>1/|\varkappa|$, provided $\varkappa_1^{l+1}(n-m_1)+\varkappa_2^{l+1}\leqslant 1$ in the case where $d=d_{l+1}$ is a real root of $\partial_2\theta_{\varkappa^{l+1}}(1,\cdot)$.*

## 10. Proof of Proposition 9.5

### 10.1. Estimation of $J^{\tau_l}$

Let us first assume that $\mathcal{N}(\theta)$ is not contained in $\{(t_1,t_2):t_2\geqslant B_l\}$, so that

$$\tau_l(x):=\varrho\left(\frac{x_2}{\varepsilon_l x_1^{a_l}}\right)(1-\varrho)\left(\frac{x_2}{N_{l+1}x_1^{a_{l+1}}}\right).$$

Arguing in a similar way as in §8.4, we consider a dyadic partition of unity $\sum_{k=0}^{\infty}\chi_k(s)=1$ on the interval $0<s\leqslant 1$ with $\chi\in C_0^\infty(\mathbb{R})$ supported in the interval $\left[\frac{1}{2},4\right]$, where

$$\chi_k(s):=\chi(2^k s),$$

and put again

$$\chi_{j,k}(x):=\chi_j(x_1)\chi_k(x_2),\quad j,k\in\mathbb{N}.$$

Then

$$J^{\tau_l}=\sum_{j,k}J_{j,k}, \tag{10.1}$$

where

$$J_{j,k}(\xi):=\int_{\mathbb{R}^2_+} e^{i\Phi(x,\xi)}\eta(x)\tau_l(x)\chi_{j,k}(x)\,dx=2^{-j-k}\int_{\mathbb{R}^2_+} e^{i\Phi_{j,k}(x,\xi)}\eta_{j,k}(x)\chi\otimes\chi(x)\,dx,$$

with $\Phi_{j,k}(x,\xi):=\Phi(2^{-j}x_1, 2^{-k}x_2, \xi)$, and where the functions $\eta_{j,k}$ are uniformly bounded in $C^\infty$. The summation in (10.1) takes place over pairs $(j,k)$ satisfying

$$a_l j + M \leqslant k \leqslant a_{l+1} j - M, \tag{10.2}$$

where $M$ can still be choosen sufficiently large, because we have the freedom to choose $\varepsilon_l$ sufficiently small and $N_{l+1}$ sufficiently large. In particular, we have $j \sim k$.

Moreover, our Assumptions 9.4 on the Newton–Puiseux diagram of $\theta$ imply exactly as in §8.4 that

$$\theta_{j,k}(x) = 2^{-(A_l j + B_l k)}(c_l x_1^{A_l} x_2^{B_l} + O(2^{-CM}))$$

for some constants $c_l \neq 0$ and $C>0$. Notice also that $B_l > B_{l+1} \geqslant 1$ here, so that $B_l \geqslant 2$, and that we are here only interested in the domain where

$$x_1 \sim 1 \sim x_2.$$

In combination with our further requirements in Assumptions 9.4 we then obtain

$$\begin{aligned}\Phi_{j,k}(x,\xi) = {} & 2^{-jn}\xi_3 x_1^n(1+v_{j,k}(x_1)) + 2^{-jm_1}\xi_2 b_1 x_1^{m_1}(1+w_{j,k}(x_1)) + 2^{-j}\xi_1 x_1 \\ & + 2^{-(A_l j + B_l k)}\xi_3(c_l x_1^{A_l} x_2^{B_l} + u_{j,k}(x_1,x_2)) + 2^{-k}\xi_2 x_2,\end{aligned}$$

where the functions $v_{j,k}$, $w_{j,k}$ and $u_{j,k}$ are of order $O(2^{-\delta(j+k)})$, respectively $O(2^{-\delta M})$, in $C^\infty$ for some $\delta>0$.

*Remark* 10.1. More precisely, the functions $v_{j,k}$, $w_{j,k}$ and $u_{j,k}$ depend smoothly on the small parameters $\delta_1 := 2^{-j/r}$ and $\delta_2 := 2^{-k}$, respectively $\delta_3 := 2^{-M}$, and vanish identically for $\delta_1 = \delta_2 = 0$, respectively $\delta_3 = 0$.

Assuming again without loss of generality that $\lambda := \xi_3 > 0$, we may thus write

$$\Phi_{j,k}(x,\xi) = 2^{-jn}\lambda F_{j,k}(x,s,\sigma),$$

with

$$\begin{aligned}F_{j,k}(x,s,\sigma) := {} & x_1^n(1+v_{j,k}(x_1)) + S_2 x_1^{m_1}(1+w_{j,k}(x_1)) + s_1 x_1 \\ & + \sigma(c_l x_1^{A_l} x_2^{B_l} + u_{j,k}(x_1,x_2) + s_2 x_2),\end{aligned}$$

and

$$s_1 := 2^{(n-1)j}\frac{\xi_1}{\lambda}, \quad s_2 := 2^{A_l j + (B_l - 1)k}\frac{\xi_2}{\lambda}, \quad S_2 := 2^{(n-m_1)j} b_1 \frac{\xi_2}{\lambda}, \quad \sigma = \sigma_{j,k} := 2^{nj - A_l j - B_l k}. \tag{10.3}$$

Lemma 10.2. *Under the Assumptions* 9.4, *the following hold true*:

(a) *The sequence* $\{1/\varkappa_1^m\}_m$ *is increasing and the sequence* $\{1/\varkappa_2^m\}_m$ *is decreasing*;

(b) *Given any constant* $N>0$, *we can choose the constant* $M$ *in* (10.2) *so large that the following holds for all* $j$ *and* $k$ *satisfying* (10.2):

$$\frac{j}{\varkappa_1^l}+N\leqslant A_l j+B_l k\leqslant\frac{k}{\varkappa_2^l}-N.$$

*In particular,*

$$\frac{j}{\varkappa_1}+N=nj+N\leqslant A_l j+B_l k\leqslant\frac{k}{\varkappa_2}-N.$$

*Proof.* Statement (a) is evident from the geometry of the Newton diagram of $\theta$. It follows also from the identity (4.4) in [19], according to which

$$\begin{aligned}\frac{1}{\varkappa_2^m}&=\frac{A_m}{a_m}+B_m=\frac{A_{m-1}}{a_m}+B_{m-1},\\ \frac{1}{\varkappa_1^m}&=A_m+a_mB_m=A_{m-1}+a_mB_{m-1},\end{aligned}$$

since the sequence $\{a_m\}_m$ is increasing.

Statement (b) is a consequence of (a) and the identities above. $\square$

Since $B_l\geqslant 1$ and $n>m_1\geqslant 2$, in combination with Lemma 10.2 we see that

$$\sigma\ll 1,\quad |\xi_1|\ll\lambda|s_1|,\quad |\xi_2|\ll\lambda|s_2|,\quad\text{and also }\ |\xi_2|\ll\lambda|S_2|. \tag{10.4}$$

Proposition 10.3. *If* $M$ *in* (10.2) *is chosen sufficiently large*, *then the following estimate*

$$|J_{j,k}(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}(1+2^{-nj}|\xi|)^{-1/3}(1+2^{-nj}\sigma_{j,k}|\xi|)^{-1/2} \tag{10.5}$$

*holds true*, *where the constant* $C$ *does not depend on* $j$, $k$ *and* $\xi$.

*Consequently*, *the maximal operator* $\mathcal{M}^{\tau_l}$ *is bounded on* $L^p(\mathbb{R}^3)$ *for every* $p>1/|\varkappa|$.

*Proof.* We first notice that $B_l\geqslant 2$, so that $\partial_2^2(x_1^{A_l}x_2^{B_l})\sim 1$.

As in the proof of Proposition 6.3 we shall distinguish several cases.

*Case* 1. $|s_1|+|S_2|\ll 1$, or $|s_1|+|S_2|\gg 1$ and $|s_1|\ll|S_2|$ or $|s_1|\gg|S_2|$.

Here, an integration by parts in $x_1$ yields

$$J_{j,k}(\xi)=O(2^{-j-k}(1+2^{-nj}\lambda(1+|s_1|+|S_2|))^{-1}),$$

which implies (10.5) because of (10.4).

*Case* 2. $|s_1|+|S_2|\gg 1$ and $|s_1|\sim|S_2|$.

Since $m_1\geqslant 2$, we have $\partial_1^2(x_1^{m_1})\sim 1$. Therefore, if $s_2$ is fixed, with $|s_2|\lesssim 1$, in view of Remark 10.1 we can apply Proposition 5.2 in a similar way as in the proof of Proposition 6.3, with $\lambda$ replaced by $2^{-nj}\lambda(|s_1|+|S_2|)$, and obtain

$$|J_{j,k}(\xi)|\leqslant C2^{-j-k}(1+2^{-nj}\lambda(1+|s_1|+|S_2|))^{-1/2}(1+2^{-nj}\sigma\lambda(1+|s_2|))^{-1/2}. \tag{10.6}$$

In fact, the proposition even shows that this estimate remains valid under small perturbations of $s_2$, so that we can choose the constant $C$ uniformly for $s_2$ in a fixed, compact interval.

On the other hand, if $|s_2|\gg 1$, we can obtain the even stronger estimate where the second exponent $-\frac{1}{2}$ is replaced by $-1$ by first integrating by parts in $x_2$ and then applying the method of stationary phase in $x_1$.

Observe at this point that if $|\xi_1|+|\xi_2|\geqslant\lambda$, so that $|\xi|\sim|\xi_1|+|\xi_2|$, then by (10.4),

$$|s_1|+|S_2|\gg 1.$$

Notice also that $|s_1|\sim|S_2|$ implies, by (10.3), that $1\sim 2^{-(m_1-1)j}|\xi_2|/|\xi_1|$, and hence

$$|\xi_1|\ll|\xi_2|.$$

Thus, if $|\xi_1|+|\xi_2|\geqslant\lambda$ and $|s_1|\sim|S_2|$, then $|\xi|\sim|\xi_2|$, and since $|s_2|\lambda\gg|\xi_2|$, we see that (10.6) implies (10.5) in this case, as well as of course in the case where $|\xi_1|+|\xi_2|\leqslant\lambda$. We are thus left with the next case.

*Case* 3. $|s_1|+|S_2|\sim 1$ and $|\xi_1|+|\xi_2|\leqslant\lambda$, and hence $|\xi|\sim\lambda$.

Since $n>m_1$, it is easy to see that in this case the polynomial

$$p(x_1):=x_1^n+S_2b_1x_1^{m_1}+s_1x_1$$

satisfies $|p''(x_1)|+|p'''(x_1)|\neq 0$ for every $x_1\sim 1$. Therefore, if we fix some point $x_1^0\sim 1$, then we can either apply Proposition 5.2 or Proposition 5.3 if we localize the oscillatory integral $J_{j,k}(\xi)$ by means of a suitable cut-off function to a small neighborhood of $x_1^0$ and translate coordinates, and finally obtain, by means of a suitable partition of unity in a similar way as in the previous case, that

$$|J_{j,k}(\xi)|\leqslant C2^{-j-k}(1+2^{-nj}\lambda)^{-1/3}(1+2^{-nj}\sigma\lambda(1+|s_2|))^{-1/2},$$

and hence (10.5). Note again that this argument first applies for fixed $s_1$, $s_2$ and $S_2$, but since Propositions 5.2 and 5.3 allow for small perturbations of parameters, the estimate above will hold uniformly in $s_1$, $s_2$ and $S_2$.

Next, observe that we may replace the factor $(1+2^{-nj}\sigma_{j,k}|\xi|)^{1/2}$ in (10.5) by

$$(1+2^{-nj}\sigma_{j,k}|\xi|)^{1/6+\varepsilon},$$

for any sufficiently small $\varepsilon>0$, which leads to

$$\begin{aligned}|J_{j,k}(\xi)| &\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}2^{nj/3}2^{(A_lj+B_lk)(1/6+\varepsilon)}(1+|\xi|)^{-1/2-\varepsilon}\\ &\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}2^{j/3\varkappa_1}2^{(k/\varkappa_2)(1/6+\varepsilon)}(1+|\xi|)^{-1/2-\varepsilon},\end{aligned}$$

since Lemma 10.1 shows that $A_lj+B_lk<k/\varkappa_2^l\leqslant k/\varkappa_2$.

Lemma 6.2 then implies that the maximal operators $\mathcal{M}^{j,k}$ associated with the multipliers $J_{j,k}$ can be estimated by

$$\|\mathcal{M}^{j,k}f\|_p\leqslant C2^{-j-k}2^{2j/3\varkappa_1p}2^{(k/\varkappa_2p)(1/3+\varepsilon)}\|f\|_p$$

for every sufficiently small $\varepsilon>0$ and $p\geqslant2$.

Observe that for $p=1/|\varkappa|$, we have

$$\frac{2}{3\varkappa_1p}=\frac{2}{3}\frac{\varkappa_1+\varkappa_2}{\varkappa_1}=\frac{2}{3}(1+a)>1,$$

so that for $p>1/|\varkappa|$ sufficiently close to $1/|\varkappa|$, we have

$$\begin{aligned}\sum_{a_lj+M\leqslant k}2^{-j-k}2^{2j/3\varkappa_1p}2^{(k/\varkappa_2p)(1/3+\varepsilon)} &\leqslant \sum_{\substack{j\leqslant k/a\\ k\geqslant M}}2^{-j-k}2^{2j/3\varkappa_1p}2^{k/3\varkappa_2p+\varepsilon}\\ &\leqslant \sum_{k\geqslant M}2^{(2(1+a)/3-\delta-1)k/a-k+(\varkappa_1+\varkappa_2)/3\varkappa_2k+\varepsilon k}\\ &=\sum_{k\geqslant M}2^{(\varepsilon-\delta/a)k},\end{aligned}\tag{10.7}$$

where $\delta>0$ depends on $p$. Choosing $\varepsilon$ sufficiently small, this series converges, so that $\mathcal{M}^{\tau_l}$ is bounded on $L^p$. For $p=\infty$, the series converges as well. By real interpolation, we thus find that $\mathcal{M}^{\tau_l}$ is $L^p$-bounded for every $p>1/|\varkappa|$. □

The case where $\mathcal{N}(\theta)\subset\{(t_1,t_2):t_2\geqslant B_l\}$ can be treated in a very similar way, if we formally replace $a_{l+1}$ by $\infty$. Indeed, in this case we have $\tau_l(x):=\varrho(x_2/\varepsilon_lx_1^{a_l})$, so that condition (10.2) has to be replaced by

$$a_lj+M\leqslant k.\tag{10.8}$$

Moreover, in this case we obviously have

$$\theta_{j,k}(x)=2^{-(A_l j+B_l k)}x_2^{B_l}(c_l x_1^{A_l}+O(2^{-\delta(j+k)}))$$

for some $\delta>0$. Therefore, if $B_l\geqslant 2$, we can argue exactly as before and see that Proposition 10.3 remains valid (notice that in (10.7) we only made use of (10.8)).

What remains open at this stage is the case where $B_l=1$. It turns out that here the oscillatory integrals $J_{j,k}(\xi)$ may possibly be of degenerate Airy type. We shall then need more detailed information, which we shall obtain be regarding $\mathcal{M}^{\tau_l}$ rather as a maximal operator of type $\mathcal{M}^{\varrho_l}$, which will be treated in the next subsection.

### 10.2. Estimation of $J^{\varrho_{l+1}}$

We now consider the maximal operators $\mathcal{M}^{\varrho_{l+1}}$ in Proposition 9.5 (b). It will here be convenient to change to the $\varkappa^{l+1}$-homogeneous coordinates

$$y_1:=x_1,\quad y_2:=x_2-dx_1^{a_{l+1}}.$$

This change of coordinates has the effect that we can assume that $d=0$. The Newton–Puiseux diagram of $\theta$ in the new coordinates will still contain the edges $\gamma_1,...,\gamma_l$, but the edge $\gamma'_{l+1}=[(A_l,B_l),(A'_{l+1},B'_{l+1})]$ may change to an interval $[(A_l,B_l),(A_{l+1},B_{l+1})]$ on the same line $\varkappa_1^{l+1}t_1+\varkappa_2^{l+1}t_2=1$, possibly with a different right endpoint $(A_{l+1},B_{l+1})$, which may even coincide with the left endpoint $(A_l,B_l)$, so that this face may even degenerate to becoming a single point.

Simplifying the notation by writing $\varkappa':=\varkappa^{l+1}$ and $a':=\varkappa'_2/\varkappa'_1=a_{l+1}$, we shall then have to estimate the oscillatory integral $J(\xi)=J^{\varrho_{l+1}}(\xi)$, with

$$J(\xi):=\int_{\mathbb{R}^2_+}e^{i\Phi(x,\xi)}\eta(x)\varrho\left(\frac{x_2}{\varepsilon' x_1^{a'}}\right)dx, \tag{10.9}$$

corresponding to the domain

$$|x_2|\leqslant\varepsilon' x_1^{a'},$$

where $\varepsilon'=\varepsilon_{l+1}>0$ can still be chosen as small as we like, under one of the following assumptions:

(i) $\partial_2\theta_{\varkappa'}(1,0)=0$, i.e., $B_{l+1}\geqslant 2$, and $\varkappa'_1(n-m_1)+\varkappa'_2\leqslant 1$;
(ii) $\partial_2\theta_{\varkappa'}(1,0)\neq 0$, i.e, $B_{l+1}=1$, and $\varkappa'(n-m_1)+\varkappa'_2\neq 1$;
(iii) $\partial_2\theta_{\varkappa'}(1,0)\neq 0$, i.e, $B_{l+1}=1$, and $\varkappa'(n-m_1)+\varkappa'_2=1$.

The most delicate case is (iii), which will lead to degenerate Airy-type integrals. Notice that the second condition in (iii) just means that the point $(n-m_1, 1)=(A_{l+1}, B_{l+1})$ belongs to $\mathcal{N}_d(\theta)$.

We shall denote the maximal operator associated with the Fourier multiplier $e^{i\xi_3}J(\xi)$ by $\mathcal{M}'$.

Observe at this point that the oscillatory integral $J^{\tau_l}$ for the still open case where $B_l=1$ can be written in the form (10.9) too, with $\varkappa':=\varkappa^l$, and hence $a'=a_l$ and $(A_l, B_l)=(A_{l+1}, B_{l+1})$, and since $B_l=1$, it will satisfy assumption (ii) or (iii). Notice that here necessarily $l>1$.

We shall therefore in the sequel relax the condition $a'>a_l$ and assume only that $a'\geqslant a_l$ in the case where $(A_l, B_l)=(A_{l+1}, B_{l+1})$ and $B_l=1$. Then, as in the proof of Proposition 6.3, we can decompose

$$J=\sum_{k=k_0}^{\infty} J_k \tag{10.10}$$

by means of a dyadic decomposition based on the $\varkappa'$-dilations $\delta'_r(x_1, x_2):=(r^{\varkappa'_1}x_1, r^{\varkappa'_2}x_2)$, where the dyadic constituent $J_k$ of $J$ is given, after rescaling, by

$$J_k(\xi)=2^{-k|\varkappa'|}\int_{\mathbb{R}^2} e^{i2^{-\varkappa'_1 nk}\lambda\Phi_k(x,s)}\varrho\biggl(\frac{x_2}{\varepsilon' x_1^a}\biggr)\eta(\delta'_{2^{-k}}x)\chi(x)\,dx,$$

where again $\lambda:=\xi_3$ is assumed to be positive, and where

$$\begin{aligned}\Phi_k(x, s, \sigma):=&\, x_1^n(1+v_k(x_1))+s_1x_1+S_2b_1x_1^{m_1}(1+w_k(x_1))\\ &+\sigma(\theta_{\varkappa'}(x_1, x_2)+\theta_{r,k}(x_1, x_2)+s_2x_2),\end{aligned}$$

with

$$s_1:=2^{\varkappa'_1(n-1)k}\frac{\xi_1}{\lambda}, \quad s_2:=2^{(1-\varkappa'_2)k}\frac{\xi_2}{\lambda}, \quad S_2:=2^{\varkappa'_1(n-m_1)k}\frac{\xi_2}{\lambda}, \quad \sigma=\sigma_k:=2^{(\varkappa'_1 n-1)k}. \tag{10.11}$$

In particular, we have

$$S_2=2^{(\varkappa'_1(n-m_1)+\varkappa'_2-1)k}s_2. \tag{10.12}$$

Moreover, since $\varkappa'_1<\varkappa_1=1/n$, we have $\varkappa'_1(n-1)>0$ and $\varkappa'_1 n-1<0$, and as

$$1=\varkappa'_1A_{l+1}+\varkappa'_2B_{l+1}\geqslant\varkappa'_2,$$

we have $1-\varkappa'_2>0$. We see that if $\Omega$ is chosen sufficiently small so that $k_0\gg 1$ in (10.10), then

$$|\sigma|\ll 1, \quad |\xi_1|\ll\lambda|s_1|, \quad |\xi_2|\ll\lambda|s_2| \quad \text{and also} \quad |\xi_2|\ll\lambda|S_2|. \tag{10.13}$$

Recall that $\theta_{\varkappa'}$ denotes the $\varkappa'$-homogeneous part of $\theta$. The functions $v_k$, $w_k$ and $\theta_{r,k}$ are of order $O(2^{-\varepsilon k})$ in $C^\infty$ for some $\varepsilon>0$, and can in fact be viewed as smooth functions $v(x_1,\delta)$, $w(x_1,\delta)$ and $\theta_r(x,\delta)$, respectively, depending also on the small parameter $\delta=2^{-k/r}$ for some positive integer $r>0$, which vanish identically when $\delta=0$.

Notice again that in our domain of integration for $J_k(\xi)$, we have

$$x_1\sim 1 \quad \text{and} \quad |x_2|\lesssim\varepsilon',$$

and clearly $|\mathcal{M}'f|\leqslant\sum_{k=k_0}^{\infty}|\mathcal{M}^kf|$, if $\mathcal{M}^k$ denotes the maximal operator associated with the Fourier multiplier $e^{i\xi_3}J_k(\xi)$.

The following proposition will then cover Proposition 9.5 (b) as well as the remaining case of Proposition 9.5 (a). The constants $l_m$ and $c_m$ will be as in Theorem 5.3. We remark at this point that clearly

$$\tfrac{1}{6}\leqslant l_m<\tfrac{1}{4}. \tag{10.14}$$

PROPOSITION 10.4. *If* $k_0$ *in* (10.10) *is chosen sufficiently large and* $\varepsilon'$ *sufficiently small, then*

$$|J_k(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-|\varkappa'|k}\sigma_k^{-(l_m+c\varepsilon)}(2^{-\varkappa_1'nk}|\xi|)^{-1/2-\varepsilon} \tag{10.15}$$

*for some* $m\in\mathbb{N}$ *with* $2\leqslant m\leqslant B_l$, *some constant* $c>0$ *and every sufficiently small* $\varepsilon>0$, *where the constant* $C$ *does not depend on* $k$ *and* $\xi$.

*Consequently, the maximal operator* $\mathcal{M}'$ *is bounded on* $L^p(\mathbb{R}^3)$ *for every* $p>1/|\varkappa|$.

*Proof.* We proceed in a similar way as in the proof of Proposition 10.3.

*Case* 1. $|s_1|+|S_2|\ll 1$, or $|s_1|+|S_2|\gg 1$ and $|s_1|\ll|S_2|$ or $|s_1|\gg|S_2|$.

Here, an integration by parts in $x_1$ yields

$$|J_k(\xi)|\leqslant C2^{-|\varkappa'|k}(1+2^{-\varkappa_1'nk}\lambda(1+|s_1|+|S_2|))^{-1},$$

which implies (10.15) because of (10.13).

*Case* 2. $|s_1|+|S_2|\gg 1$ and $|s_1|\sim|S_2|$.

Observe first that for any $x_1^0\sim 1$, the polynomial $P(x_2):=\theta_{\varkappa'}(x_1^0,x_2)$ has degree $\deg P\geqslant 2$. Indeed, this is clear under assumption (i), since $B_{l+1}\geqslant 2$, and under the assumptions (ii) and (iii) it follows from $B_l\geqslant 2$, respectively $B_{l-1}\geqslant 2$ in the case where $B_l=B_{l+1}=1$. Clearly also $\deg P\leqslant B_l$.

Therefore, if $|s_2|\lesssim 1$, we can argue in a similar way as in case 3 of the proof of Proposition 6.3, and obtain by means of Proposition 5.2 that

$$|J_k(\xi)|\leqslant C2^{-|\varkappa'|k}(1+2^{-\varkappa_1'nk}\lambda(1+|s_1|+|S_2|))^{-1/2}(1+2^{-\varkappa_1'nk}\sigma\lambda(1+|s_2|))^{-1/m} \tag{10.16}$$

for some $m$ with $2\leqslant m\leqslant B_l$, provided $\varepsilon'$ is chosen sufficiently small.

On the other hand, if $|s_2|\gg 1$, we can obtain the even stronger estimate where the second exponent $-1/m$ is replaced by $-1$ by first integrating by parts in $x_2$ and then integrating in $x_1$.

Now from (10.13) we deduce as in the proof of Proposition 10.3 that if $|\xi_1|+|\xi_2|\geqslant\lambda$, so that $|\xi|\sim|\xi_1|+|\xi_2|$, then we have $|s_1|+|S_2|\gg 1$, and $|s_1|\sim|S_2|$ implies that $|\xi_1|\ll|\xi_2|$.

Thus, if $|\xi_1|+|\xi_2|\geqslant\lambda$ and $|s_1|\sim|S_2|$, then $|\xi|\sim|\xi_2|$, and since $|s_2|\lambda\gg|\xi_2|$, we see that (10.16) implies (10.15) in this case, as well as of course in the case where $|\xi_1|+|\xi_2|\leqslant\lambda$, provided $\varepsilon$ is chosen small enough. We are thus left with the next case.

*Case* 3. $|s_1|+|S_2|\sim 1$ and $|\xi_1|+|\xi_2|\leqslant\lambda$, and hence $|\xi|\sim\lambda$.

Since $n>m_1$, the polynomial $p(x_1):=x_1^n+S_2b_1x_1^{m_1}+s_1x_1$ satisfies

$$|p''(x_1)|+|p'''(x_1)|\neq 0 \quad \text{for every } x_1\sim 1.$$

But, if either $|s_1|\ll|S_2|$ or $|s_1|\gg|S_2|$, then all critical points of the polynomial

$$x_1^n+S_2b_1x_1^{m_1}+s_1x_1$$

will be non-degenerate, so that we can argue exactly as in case 2. We shall therefore assume that

$$|s_1|\sim|S_2|\sim 1.$$

Now, under assumption (i), we have $\partial_2\theta_{\varkappa'}(x_1^0,0)=0$ whenever $x_1^0\sim 1$, whereas $|s_2|\gtrsim 1$, by (10.12), so that

$$\partial_2(\theta_{\varkappa'}+s_2x_2)(x_1^0,0)=\partial_2\theta_{\varkappa'}(x_1^0,0)+s_2\neq 0. \tag{10.17}$$

The same is true also under assumption (ii), for then either $|s_2|\gg 1$ or $|s_2|\ll 1$ (by (10.12)), whereas $\partial_2\theta_{\varkappa'}(x_1^0,0)\neq 0$, and it also applies in case (iii), provided $|s_2|\gg 1$ or $|s_2|\ll 1$.

In these cases, we shall first integrate by parts in $x_2$ and then apply Lemma 5.1 from §11, which is a useful variant of van der Corput's lemma, to the $x_1$-integration, which leads to the estimate

$$|J_k(\xi)|\leqslant C2^{-|\varkappa'|k}(1+2^{-\varkappa_1'nk}\lambda)^{-1/3}(1+2^{-\varkappa_1'nk}\sigma\lambda(1+|s_2|))^{-1}.$$

By replacing the second exponent $-1$ by $-\frac{1}{6}-\varepsilon$, we see in view of (10.14) that this implies (10.15).

We are thus left with the case where assumption (iii) holds true, and where $|s_2|\sim 1$. Fix $x_1^0\sim 1$. Then $\partial_1\partial_2(\theta_{\varkappa'}+s_2x_2)(x_1^0,0)\neq 0$, since $\theta_{\varkappa'}(x_1,x_2)=c_0x_1^{n-m_1}x_2+O(x_2^2)$, where $c_0\neq 0$.

Assume first that (10.17) holds true. Then we can again argue as before, provided we introduce in our formula for $J_k(\xi)$ an additional smooth cut-off function $a(x_1)$ supported in a sufficiently small neighborhood of $x_1^0$.

So, assume next that $\partial_2(\theta_{\varkappa'}+s_2x_2)(x_1^0,0)=0$. Since the degree of the polynomial $P(x_2):=\theta_{\varkappa'}(x_1^0,x_2)$ satisfies $B_l\geqslant\deg P\geqslant 2$, after shifting the $x_1$ coordinates by $x_1^0$, we can apply Theorem 5.4 and obtain estimate (10.15) for some $m$ with $2\leqslant m\leqslant B_l$, if we again introduce a cut-off function $a(x_1)$ supported in a sufficiently small neighborhood of $x_1^0$ into $J_k(\xi)$. Recall here that the functions $v_k$, $w_k$ and $\theta_{r,k}$ are smooth functions $v(x_1,\delta)$, $w(x_1,\delta)$ and $\theta_r(x,\delta)$, respectively, depending also on the small parameter $\delta=2^{-k/r}$ for some positive integer $r>0$, which vanish identically when $\delta=0$.

The estimate (10.15) then follows by decomposing $J_k(\xi)$ into a finite number of such "localized" integrals by means of a partition of unity.

Next, in order to estimate the maximal operator $\mathcal{M}'$, observe that (10.15) implies that for any sufficiently small $\varepsilon>0$ we have

$$|J_k(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-|\varkappa'|k}2^{\varkappa_1'nk(1/2+\varepsilon)}2^{(1-\varkappa_1'n)k(l_m+c\varepsilon)}(1+|\xi|)^{-1/2-\varepsilon}.$$

Recalling that $1-\varkappa_1'n>0$ and $l_m<\frac{1}{4}$ by (10.14), we thus see that there is some $\delta>0$ such that

$$|J_k(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-\delta k}2^{-|\varkappa'|k}2^{(1+\varkappa_1'n)k/4}(1+|\xi|)^{-1/2-\varepsilon},$$

provided $\varepsilon$ is sufficiently small. Lemma 6.2 then implies that

$$\|\mathcal{M}^kf\|_p\leqslant C2^{-\delta k}2^{-|\varkappa'|k}2^{(1+\varkappa_1'n)/2p}\|f\|_p$$

for every $p\geqslant 2$. Notice that

$$\frac{1+\varkappa_1'n}{2|\varkappa'|}\leqslant\frac{1}{|\varkappa|}. \tag{10.18}$$

Indeed, we have $t:=\varkappa_1'n=\varkappa_1'/\varkappa_1\leqslant 1$ and $\varkappa_2'\geqslant\varkappa_2$ by Lemma 10.2, so that

$$\frac{1+\varkappa_1'n}{2|\varkappa'|}=\frac{1+t}{2(\varkappa_1t+\varkappa_2')}\leqslant\frac{1+t}{2(\varkappa_1t+\varkappa_2)}.$$

The latter function is increasing in $t$, so that we may replace $t$ by 1 and obtain (10.18).

The estimate (10.18) shows that the norms of the maximal operators $\mathcal{M}^k$ sum in $k$ when $p\geqslant 1/|\varkappa|$, which concludes the proof of Proposition 10.4, and hence also the proof of our main result, Theorem 1.2. □

## 11. Uniform estimates for oscillatory integrals with finite-type phase functions of two variables

In this section we shall provide a proof of Theorem 1.9. We shall closely follow the proof of Theorem 1.3, which did already provide uniform estimates for the Fourier transforms of surface carried measures $\widehat{\varrho\, d\sigma}(\xi)$ for the contribution by the region near the principal root jet. Notice that the assumption $\varrho\geqslant 0$ that we had made for the estimation of the maximal operator $\mathcal{M}$ had only been introduced for convenience and was not needed for the estimations of oscillatory integrals. Without further mentioning, we shall use the same notation as in the various parts of the proof of Theorem 1.3.

We may assume that $x^0=0$ and that $S$ is the graph

$$S=\{(x_1,x_2,\phi(x_1,x_2)):(x_1,x_2)\in\Omega\}$$

of a smooth real-valued function $\phi\in C^\infty(\Omega)$ of finite type defined on an open neighborhood $\Omega$ of the origin in $\mathbb{R}^2$ and satisfying

$$\phi(0,0)=0 \quad \text{and} \quad \nabla\phi(0,0)=0.$$

We then have to prove the following result.

Theorem 11.1. *There exist a neighborhood $\Omega\subset\mathbb{R}^2$ of the origin and a constant $C$ such that for every $\eta\in C_0^\infty(\Omega)$ the following estimate holds true for every $\xi\in\mathbb{R}^3$:*

$$\left|\int_{\mathbb{R}^2} e^{i(\xi_3\phi(x_1,x_2)+\xi_1x_1+\xi_2x_2)}\eta(x_1,x_2)\,dx\right|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}(1+|\xi|)^{-1/h(\phi)}\log(2+|\xi|). \quad (11.1)$$

We note that the van der Corput type Lemma 5.1 will play a similar role for the proof of Theorem 11.1 as Corollary 4.6 did for the proof of Theorem 1.2.

By decomposing $\mathbb{R}^2$ into its four quadrants, we may reduce ourselves to the estimation of oscillatory integrals of the form

$$J(\xi):=\int_{\mathbb{R}^2_+} e^{i(\xi_3\phi(x_1,x_2)+\xi_1x_1+\xi_2x_2)}\eta(x_1,x_2)\,dx.$$

Notice also that we may assume in the sequel that

$$|\xi_1|+|\xi_2|\leqslant\delta|\xi_3|, \quad \text{and hence } |\xi|\sim|\xi_3|, \quad (11.2)$$

where $0<\delta\ll 1$ is a sufficiently small constant, since for $|\xi_1|+|\xi_2|>\delta|\xi_3|$ the estimate (11.1) follows by an integration by parts, if $\Omega$ is chosen small enough. Of course, we may in addition always assume that $|\xi|\geqslant 2$.

If $\chi$ is any integrable function defined on $\Omega$, we shall put

$$J^\chi(\xi) := \int_{(\mathbb{R}_+)^2} e^{i(\xi_3\phi(x_1,x_2)+\xi_1x_1+\xi_2x_2)}\eta(x_1,x_2)\chi(x)\,dx.$$

The case $h(\phi)<2$ is contained in [11] (here, estimate (11.1) holds true even without the logarithmic term $\log(2+|\xi|)$), so let us assume from now on that

$$h=h(\phi)\geqslant 2.$$

Following §7 we shall begin with the easiest case where the coordinates $x$ are adapted to $\phi$. In analogy with the proof of Proposition 7.1 we then decompose $J(\xi)=\sum_{k=k_0}^\infty J_k(\xi)$, where

$$\begin{aligned} J_k(\xi) &:= \int_{(\mathbb{R}_+)^2} e^{i(\xi_3\phi(x)+\xi_1x_1+\xi_2x_2)}\eta(x)\chi_k(x)\,dx \\ &= 2^{-k|\varkappa|}\int_{(\mathbb{R}_+)^2} e^{i(2^{-k}\xi_3\phi^k(x)+2^{-k\varkappa_1}\xi_1x_1+2^{-k\varkappa_2}\xi_2x_2)}\eta(\delta_{2^{-k}}(x))\chi(x)\,dx, \end{aligned}$$

and where $\chi$ is supported in an annulus $D$. Moreover, as in the proof of Corollary 7.2 we can choose the weight $\varkappa$ according to Lemma 2.4 such that $0<\varkappa_1\leqslant\varkappa_2<1$ and

$$\frac{1}{|\varkappa|}=d_h(\phi_\varkappa)\leqslant h(\phi_\varkappa)=h.$$

Then, as in the proof of Proposition 7.1, given any point $x^0\in D$, we can find a unit vector $e\in\mathbb{R}^2$ and some $m\in\mathbb{N}$ with $2\leqslant m\leqslant h(\phi_\varkappa)=h$ such that $\partial_e^m\phi_\varkappa(x^0)\neq 0$. For $k\geqslant k_0$ sufficiently large we can thus apply Lemma 5.1 to the $x_2$-integration in $J_k(\xi)$ near the point $x^0$. By means of a partition of unity argument, we then get that

$$\begin{aligned} |J_k(\xi)| &\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-k|\varkappa|}(1+2^{-k}|\xi_3|)^{-1/m} \\ &\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-k/h(\phi)}(1+2^{-k}|\xi|)^{-1/h}. \end{aligned}$$

The estimate (11.1) then follows by summation in $k$.

Assume next that the coordinates $x$ are not adapted to $\phi$. In a first step, we then decompose $J(\xi)=J^{1-\varrho_1}(\xi)+J^{\varrho_1}(\xi)$, where $\varrho_1$ is the cut-off function introduced in §8.1 which localizes to a narrow $\varkappa$-homogeneous neighborhood

$$|x_2-b_1x_1^{m_1}|\leqslant\varepsilon_1x_1^{m_1}$$

of the curve $x_2=b_1x_1^{m_1}$.

The oscillatory integral $J^{1-\varrho_1}(\xi)$ can be estimated in a similar way as in the case of adapted coordinates by means of Lemma 5.1 (cf. also the proof of Lemma 8.1).

Moreover, if $\mathcal{N}(\phi^a)=(\nu_1,h)+\mathbb{R}^2_+$, with $\nu_1<h$ (case (c2) in §3), we recall from the proof of Lemma 8.1 that $\phi_\varkappa(x)=cx_1^{\nu_1}(x_2-b_1x_1^{m_1})^h$, which implies that $h(\phi_\varkappa)=h(\phi^a_\varkappa)=h$, and we see that in this case we can again apply Lemma 5.1 to the $x_2$-integration in order to see that also $J^{\varrho_1}(\xi)$ satisfies (11.1).

We may and shall therefore from now on assume that the Newton polyhedron of $\phi^a$ has at least one compact edge "lying above" the principal face, i.e., that one of the cases (a), (b) or (c1) from §3 applies. Finally $J^{\varrho_1}(\xi)$ remains to be considered.

In analogy with Proposition 8.2, in the next step we prove the following result.

PROPOSITION 11.2. *Let* $\varepsilon_0>0$. *If the neighborhood* $\Omega$ *of the point* $(0,0)$ *is chosen sufficiently small*, *then the oscillatory integral* $J^{1-\varrho_0}(\xi)$ *satisfies estimate* (11.1).

*Moreover*, *if the principal face* $\pi(\phi^a)$ *is a vertex or unbounded*, *then the same holds true for* $J(\xi)$ *in place of* $J^{1-\varrho_0}(\xi)$.

Let us again first assume that $\phi$ is analytic. To prove the first statement, we decompose the corresponding domain as in §8.2 into the domains $D_l$, $l=l_0,...,\lambda$, which become $\varkappa^l$-homogeneous in the coordinates $y$ defined by (8.1), and the transition domains $E_l$, $l=l_0-1,...,\lambda-1$. Accordingly, we decompose

$$J^{\varrho_1}(\xi)=\sum_{l=l_0}^{\lambda}J^{\varrho_l}(\xi)+\sum_{l=l_0-1}^{\lambda-1}J^{\tau_l}(\xi),$$

where $\varrho_l$ and $\tau_l$ are the cut-off functions defined in (8.10) and (8.11).

*Estimation of* $J^{\varrho_l}(\xi)$. In analogy with the proof of Lemma 8.3, after applying the change of coordinates (8.1) and performing a dyadic decomposition as before, only with the weight $\varkappa$ replaced by the weight $\varkappa^l$, we find that $J^{\varrho_l}(\xi)=\sum_{k=k_0}^{\infty}J_k(\xi)$, where

$$\begin{aligned}J_k(\xi)=2^{-k|\varkappa^l|}\int_{\mathbb{R}^2_+}&e^{i(2^{-k}\xi_3\phi^k(y)+2^{-k\varkappa_1^l}\xi_1y_1+2^{-k\varkappa_2^l}\xi_2y_2+2^{-k\varkappa_2^l}\xi_2\psi^k(y_1))}\\&\times\varrho_l^a(y)\eta^a(\delta^l_{2^{-k}}y)\chi(y)\,dy,\end{aligned}$$

with $\phi^k(y)$, $\psi^k(y_1)$, etc. defined as in §8.3.

By means of Corollary 3.2 (i) and (ii) we have seen in the proof of Lemma 8.3 that (8.13) holds true if $l\leqslant\lambda-1$, or if $l=\lambda$ and if the principal face $\pi(\phi^a)$ of the Newton polyhedron of $\phi^a$ is a vertex (case (b)) or unbounded as in case (c1) in §3. In these cases we can thus estimate $J_k(\xi)$ by means of Lemma 5.1 applied to the $y_2$-integration and a partition of unity argument and obtain

$$\begin{aligned}|J_k(\xi)|&\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-k|\varkappa^l|}(1+2^{-k}|\xi_3|)^{-1/d_h(\phi^a_{\varkappa^l})}\\&\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-k/h(\phi)}(1+2^{-k}|\xi|)^{-1/h}.\end{aligned}$$

The case where $l=\lambda$ and where $\pi(\phi^a)$ is a compact edge (case (a)) remains to be considered. Again, if $a=\varkappa_2^\lambda/\varkappa_1^\lambda\in\mathbb{N}$, then we may assume that (8.13) holds true in view of (8.6), so assume that $a\notin\mathbb{N}$. We are left with those points $y^0$ in the support of $\chi\varrho_\lambda^a$ for which $\partial_2^j\phi_{\rm pr}^a(y^0)=0$ for every $2\leqslant j\leqslant h$ but $\partial_2\phi_{\rm pr}^a(y^0)\neq 0$; all other points $y^0$ can be treated as before. However, in this case the estimate (11.3) is an immediate consequence of Corollary 6.4, if we apply a change of coordinates to a small $\varkappa^\lambda$-homogeneous neighborhood of $y^0$ as in the corresponding part of the proof of Lemma 8.3.

By summing over all $k$, we see that $J^{\varrho_l}(\xi)$ satisfies estimate (11.1).

*Estimation of* $J^{\tau_l}(\xi)$. Following §8.4, we decompose

$$J^{\tau_l}(\xi)=\sum_{j,k}J_{j,k}(\xi),$$

where summation takes place over all pairs $j,k$ satisfying (8.14), i.e.,

$$a_lj+M\leqslant k\leqslant a_{l+1}j-M, \tag{11.3}$$

with $J_{j,k}(\xi)$ given by

$$\begin{aligned}J_{j,k}(\xi)&:=\int_{\mathbb{R}^2}e^{i(\xi_3\phi^a(y)+\xi_1y_1+\xi_2y_2+\xi_2\psi(y_1))}\tau_l^a(y)\eta^a(y)\chi_{j,k}(y)\,dy\\&=2^{-j-k}\int_{\mathbb{R}^2}e^{i(\xi_3\tilde\phi^{j,k}(y)+2^{-j}\xi_1y_1+2^{-k}\xi_2y_2+\xi_2\psi(2^{-j}y_1))}\tilde\tau^{j,k}(y)\tilde\eta^{j,k}(y)\chi(y_1)\chi(y_2)\,dy.\end{aligned}$$

Here, we have kept the notation from §8.4. Assume first that $\phi$ is analytic. Then, by (8.16),

$$\tilde\phi^{j,k}(y)=2^{-(A_lj+B_lk)}(c_ly_1^{A_l}y_2^{B_l}+O(2^{-CM}))$$

for some constant $C>0$, where $A_l$ and $B_l$ are given by (3.2) and $M$ can still be chosen as large as we need, and where

$$\partial_2^2(y_1^{A_l}y_2^{B_l})\sim 1.$$

We can thus again apply Lemma 5.1, with $m=2$, to the $y_2$-integration in $J_{j,k}(\xi)$ and obtain

$$\begin{aligned}|J_{j,k}(\xi)|&\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}(1+2^{-(A_lj+B_lk)}|\xi_3|)^{-1/2}\\&\sim C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}(1+2^{-(A_lj+B_lk)}|\xi|)^{-1/2}.\end{aligned}$$

Then

$$|J^{\tau_l}(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}(J_0^{\tau_l}(\xi)+J_\infty^{\tau_l}(\xi)),$$

with

$$J_0^{\tau_l}(\xi) := \sum_{(j,k)\in I_0} 2^{-(1-A_l/2)j-(1-B_l/2)k}|\xi|^{-1/2},$$
$$J_\infty^{\tau_l}(\xi) := \sum_{(j,k)\in I_\infty} 2^{-j-k},$$

where $I_0$ and $I_\infty$ denote the index sets

$$I_0 := \{(j,k)\in\mathbb{N}^2 : A_l j + B_l k \leqslant \log|\xi| \text{ and } a_l j \leqslant k \leqslant a_{l+1} j\}$$

and

$$I_\infty := \{(j,k)\in\mathbb{N}^2 : A_l j + B_l k > \log|\xi|\}.$$

These estimates can easily be summed in $j$ and $k$ by means of the following auxiliary result.

LEMMA 11.3. *Let $0<a_1<a_2$ and $b_1, b_2\geqslant 0$ with $b_1+b_2>0$ be given. For $\gamma>0$, consider the triangle $A_\gamma := \{(t_1,t_2)\in(\mathbb{R}_+)^2 : a_1t_1\leqslant t_2\leqslant a_2t_1$ and $b_1t_1+b_2t_2\leqslant\gamma\}$, and denote by $(0,0)$, $\gamma X_1$ and $\gamma X_2$, with*

$$X_1 := \frac{1}{b_1+a_1b_2}(1,a_1) \quad \text{and} \quad X_2 := \frac{1}{b_1+a_2b_2}(1,a_2),$$

*the three vertices of $A_\gamma$. Assume that $\mu=(\mu_1,\mu_2)\in\mathbb{R}^2$ is such that*

$$\mu\cdot X_1 < \mu\cdot X_2. \tag{11.4}$$

(a) *If $\mu\cdot X_2>0$, then*
$$\int_{A_\gamma} e^{\mu\cdot t}\,dt \leqslant Ce^{\gamma\mu\cdot X_2};$$
(b) *If $\mu\cdot X_2=0$, then*
$$\int_{A_\gamma} e^{\mu\cdot t}\,dt \leqslant C\gamma;$$
(c) *If $\mu\cdot X_2<0$, then*
$$\int_{A_\gamma} e^{\mu\cdot t}\,dt \leqslant C,$$
*where the constant $C$ in these estimates depends only on the $a_j$, the $b_j$ and $\mu$.*

*Similarly, if we put $B_\gamma := \{(t_1,t_2)\in(\mathbb{R}_+)^2 : a_1t_1\leqslant t_2\leqslant a_2t_1$ and $b_1t_1+b_2t_2\geqslant\gamma\}$, then the following holds true:*

(d) *If $\mu\cdot X_2<0$, then*
$$\int_{B_\gamma} e^{\mu\cdot t}\,dt \leqslant Ce^{\gamma\mu\cdot X_1}.$$

*Proof.* Let us change to the coordinates $(x_1, x_2)$ given by

$$(t_1, t_2) = (x_1 + x_2, a_1 x_1 + a_2 x_2).$$

In these coordinates, $A_\gamma$, $X_1$ and $X_2$ correspond to

$$\tilde{A}_\gamma := \{(x_1, x_2) \in (\mathbb{R}_+)^2 : (b_1 + a_1 b_2) x_1 + (b_1 + a_2 b_2) x_2 \leqslant \gamma\},$$

$$\widetilde{X}_1 := \left(\frac{1}{b_1 + a_1 b_2}, 0\right) \quad \text{and} \quad \widetilde{X}_2 := \left(0, \frac{1}{b_1 + a_2 b_2}\right),$$

respectively. Moreover, $\mu \cdot t = \tilde{\mu} \cdot x$, where $\tilde{\mu} \cdot \widetilde{X}_1 < \tilde{\mu} \cdot \widetilde{X}_2$, i.e.,

$$\frac{\tilde{\mu}_1}{b_1 + a_1 b_2} < \frac{\tilde{\mu}_2}{b_1 + a_2 b_2}. \tag{11.5}$$

Now, in case (a) we have $\tilde{\mu}_2 > 0$, so that, because of (11.5),

$$\begin{aligned}
\int_{A_\gamma} e^{\mu \cdot t}\, dt = C \int_{\tilde{A}_\gamma} e^{\tilde{\mu} \cdot x}\, dx & \\
&= \frac{C}{\tilde{\mu}_2} e^{\tilde{\mu}_2 \gamma/(b_1 + a_2 b_2)} \int_0^{\gamma/(b_1 + a_1 b_2)} e^{(\tilde{\mu}_1 - \tilde{\mu}_2 (b_1 + a_1 b_2)/(b_1 + a_2 b_2)) x_1}\, dx_1 \\
&\leqslant \frac{C}{\tilde{\mu}_2} e^{\tilde{\mu}_2 \gamma/(b_1 + a_2 b_2)} \\
&= \frac{C}{\tilde{\mu}_2} e^{\gamma \mu \cdot X_2},
\end{aligned}$$

where $C$ depends only on $a_1$ and $a_2$.

In case (b), we have $\tilde{\mu}_2 = 0$ and $\tilde{\mu}_1 < 0$, so that a similar estimation as before leads to

$$\int_{A_\gamma} e^{\mu \cdot t}\, dt \leqslant C \gamma,$$

and case (c) is obvious, since here $\tilde{\mu}_1, \tilde{\mu}_2 < 0$.

The estimate in (d) is obtained in an analogous way as to the one in (a). $\square$

To estimate $J_0^{\tau_l}(\xi)$, we put $\mu := \left(\frac{1}{2} A_l - 1, \frac{1}{2} B_l - 1\right)$, $a_1 := a_l$, $a_2 := a_{l+1}$, $b_1 := A_l$, $b_l := B_l$ and $\gamma := \log |\xi|$ in Lemma 11.3. Then

$$X_1 = \frac{1}{A_l + a_l B_l} (1, a_l) \quad \text{and} \quad X_2 = \frac{1}{A_l + a_{l+1} B_l} (1, a_{l+1}),$$

and (cf. also the discussion in §3)

$$\mu \cdot X_1 = \frac{1}{2} - \frac{1 + a_l}{A_l + a_l B_l} = \frac{1}{2} - \frac{1}{d_h(\tilde{\phi}_{\varkappa_l})} \quad \text{and} \quad \mu \cdot X_2 = \frac{1}{2} - \frac{1 + a_{l+1}}{A_l + a_{l+1} B_l} = \frac{1}{2} - \frac{1}{d_h(\tilde{\phi}_{\varkappa_{l+1}})}.$$

Since $d_h(\tilde{\phi}_{\varkappa_l})<d_h(\tilde{\phi}_{\varkappa_{l+1}})$, we see that condition (11.4) is satisfied. Comparing the sum in $J_0^{\tau_l}(\xi)$ with the corresponding integral and applying Lemma 11.3, we thus find that

$$|J_0^{\tau_l}(\xi)|\leqslant C|\xi|^{-1/2}\log|\xi|\leqslant C|\xi|^{-1/h}\log(2+|\xi|),$$

if $\mu\cdot X_2\leqslant 0$, and

$$|J_0^{\tau_l}(\xi)|\leqslant C|\xi|^{-1/2}\exp\left(\log|\xi|\left(\frac{1}{2}-\frac{1}{d_h(\tilde{\phi}_{\varkappa_{l+1}})}\right)\right)\leqslant C|\xi|^{-1/d_h(\tilde{\phi}_{\varkappa_{l+1}})},$$

if $\mu\cdot X_2>0$. Since $d_h(\tilde{\phi}_{\varkappa_{l+1}})\leqslant h(\phi)$, this shows that $J_0^{\tau_l}(\xi)$ satisfies the estimate (11.1).

Similarly, in order to estimate $J_\infty^{\tau_l}(\xi)$, we put $\mu:=(-1,-1)$ in Lemma 11.3 (d). Then $\mu\cdot X_2=-1/d_h(\tilde{\phi}_{\varkappa_{l+1}})<0$ and $\mu\cdot X_1=-1/d_h(\tilde{\phi}_{\varkappa_l})<\mu\cdot X_2$, so that we obtain

$$|J_\infty^{\tau_l}(\xi)|\leqslant C\exp\left(\log|\xi|\left(-\frac{1}{d_h(\tilde{\phi}_{\varkappa_l})}\right)\right)\leqslant C|\xi|^{-1/h(\phi)}.$$

In combination, we have seen that all $J^{\tau_l}(\xi)$ satisfy the estimate (11.1).

In order to prove the second statement in Proposition 11.2, we assume that we are in case (b) or case (c1) of §3, so that the principal face of $\mathcal{N}(\phi^a)$ is either a vertex or unbounded. Recall from Corollary 3.2 (ii) that then $d_h(\phi^a_{\varkappa^\lambda})\leqslant h$ and $\partial_2^h\phi^a_{\varkappa^\lambda}(1,0)\neq 0$. Since, by the first part of Proposition 11.2, what remains to be controlled is the contribution to $J(\xi)$ given by a domain of the form $|x_2-\psi(x_1)|\leqslant\varepsilon_0x_1^a$, where $\varepsilon_0$ can be chosen as small as needed, we can thus argue as in the estimation of $J^{\varrho_\lambda}$ by means of Lemma 5.1.

We have thus completed the proof Proposition 11.2, at least when $\phi$ is analytic. However, the case of a general finite-type function $\phi$ can again be reduced to the analytic case along the lines of §8.5. Notice here that we have only made use of the van der Corput type Lemma 5.1 and Proposition 6.3, and the estimates in these results are stable under small perturbations of the phase function by error terms vanishing to sufficiently high order at the origin.

In view of Proposition 11.2, we may and shall from now on assume that the principal face of $\mathcal{N}(\phi^a)$ is a compact edge (case (a)). What remains to be estimated is the contribution of a small domain of the form (9.1) to $J(\xi)$, i.e., we are left with the oscillatory integral $J^{\varrho_0}(\xi)$ which, after a change of coordinates, is given by (6.2). With a slight abuse of notation, we shall therefore adapt the notation from §9 and write

$$J(\xi):=J^{\varrho_0}(\xi)=\int_{\mathbb{R}^2_+}e^{i(\xi_1x_1+\xi_2\psi(x_1)+\xi_2x_2+\xi_3\phi(x))}\varrho\left(\frac{x_2}{\varepsilon_0x_1^a}\right)\eta(x)\,dx,$$

where here $\phi$ and $\psi$ satisfy the Assumptions 9.2.

We may also in this context assume that condition (9.3) is satisfied, since otherwise we can obtain the desired estimate for $J(\xi)$ again by means of Lemma 5.1 applied to the $x_2$-integration in $J(\xi)$, provided that $j\geqslant 2$ in (9.3), and by means of Proposition 6.3, if $\partial_2\phi_{\mathrm{pr}}(1,0)\neq 0$, since $h\geqslant 2$. However, under these assumptions we have derived estimates for $J(\xi)$ in §9 and §10, and what remains to be shown is that these estimate are sufficient also in order to establish (11.1).

To this end, we apply the domain decomposition algorithm of §9.1 and are left with the estimation of the oscillatory integrals $J^{\tau_l}$ and $J^{\varrho_{l+1}}$ defined in that subsection.

We begin with $J^{\tau_l}(\xi)=\sum_{j,k} J_{j,k}(\xi)$, where $J_{j,k}$ is as defined in §10.1 and where summation takes place again over the set of indices $j$ and $k$ satisfying (11.3). Observe that, according to our discussion in §9.1, we have here that $\varkappa_1=1/n$, $\varkappa_2/\varkappa_1\geqslant 2$ and

$$\varkappa_1 A_1+\varkappa_2 B_1=1,$$

where $B_1=B\geqslant 3$ (cf. (9.5)). This implies that $\varkappa_2\leqslant\frac{1}{3}$, and hence

$$\varkappa_1\leqslant\tfrac{1}{6}\quad\text{and}\quad\varkappa_2\leqslant\tfrac{1}{3}.$$

From Proposition 10.3, we then conclude that

$$|J_{j,k}(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}(1+2^{-nj}|\xi|)^{-\varkappa_1}(1+\sigma_{j,k}|\xi|)^{-\varkappa_2},$$

and hence

$$|J_{j,k}(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}2^{-j-k}\frac{1}{1+2^{-j}2^{-(A_lj+B_lk)\varkappa_2}|\xi|^{\varkappa_1+\varkappa_2}}.$$

Then

$$|J^{\tau_l}(\xi)|\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)}(J_0^{\tau_l}(\xi)+J_\infty^{\tau_l}(\xi)),$$

where here

$$J_0^{\tau_l}(\xi):=\sum_{(j,k)\in I_0}2^{-k+(A_lj+B_lk)\varkappa_2}|\xi|^{-|\varkappa|}\quad\text{and}\quad J_\infty^{\tau_l}(\xi):=\sum_{(j,k)\in I_\infty}2^{-j-k},$$

with index sets

$$I_0:=\{(j,k)\in\mathbb{N}^2:j+(A_lj+B_lk)\varkappa_2\leqslant\log_2(|\xi|^{|\varkappa|})\text{ and }a_lj\leqslant k\leqslant a_{l+1}j\}$$

and

$$I_\infty:=\{(j,k)\in\mathbb{N}^2:j+(A_lj+B_lk)\varkappa_2>\log_2(|\xi|^{|\varkappa|})\}.$$

As $j\leqslant k/a_l$ and $k\leqslant c\log|\xi|$ in $I_0$, summing first in $j$ and then in $k$, we obtain that

$$|J_0^{\tau_l}(\xi)|\leqslant C\sum_{k\leqslant c\log|\xi|}2^{((A_l/a_l+B_l)\varkappa_2-1)k}|\xi|^{-|\varkappa|}=C\sum_{k\leqslant c\log_2|\xi|}2^{(\varkappa_2/\varkappa_2^l-1)k}|\xi|^{-|\varkappa|}.$$

But, by Lemma 10.2, $\varkappa_2/\varkappa_2^l \leqslant 1$, so that $|J_0^{\tau_l}(\xi)| \leqslant C|\xi|^{-|\varkappa|}\log|\xi|$.

Similarly, since $(A_l j + B_l k)\varkappa_2 \leqslant k$ (cf. Lemma 10.2), we have that $j+k > \log_2(|\xi|^{|\varkappa|})$. Putting $r:=j+k$, we thus see that

$$|J_\infty^{\tau_l}(\xi)| \leqslant C \sum_{r \geqslant \log_2(|\xi|^{|\varkappa|})} r 2^{-r} \leqslant C'|\xi|^{-|\varkappa|}\log|\xi|.$$

Since $|\varkappa| = 1/h(\phi)$, we thus see that $J^{\tau_l}(\xi)$ satisfies estimate (11.1).

What remains to be considered are the $J^{\varrho_{l+1}}(\xi)$, respectively the oscillatory integrals $J(\xi)$ given by (10.9), which we decompose according to (10.10) into $J(\xi) = \sum_{k=k_0}^\infty J_k(\xi)$. By Proposition 10.4, we have

$$|J_k(\xi)| \leqslant C\|\eta\|_{C^3(\mathbb{R}^2)} 2^{-|\varkappa'|k} \sigma_k^{-(l_m + c\varepsilon)} (2^{-\varkappa_1' nk}|\xi|)^{-1/2-\varepsilon/2}$$

for every sufficiently small $\varepsilon > 0$, where $l_m < \frac{1}{4}$, and by the definition of $J_k(\xi)$ in §10.2, we also have $|J_k(\xi)| \leqslant C\|\eta\|_{C^3(\mathbb{R}^2)} 2^{-|\varkappa'|k}$. Putting this in the definition of $\sigma_k$, we get that

$$\begin{aligned}
|J_k(\xi)| &\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)} 2^{-|\varkappa'|k} (1 + \sigma_k^{1/4} 2^{-\varkappa_1' nk/2} |\xi|^{1/2})^{-1} \\
&\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)} 2^{-|\varkappa'|k} (1 + 2^{-(1+\varkappa_1' n)k/2} |\xi|)^{-1/2} \\
&\leqslant C\|\eta\|_{C^3(\mathbb{R}^2)} 2^{-|\varkappa'|k} (1 + 2^{-(1+\varkappa_1' n)k/2} |\xi|)^{-|\varkappa|},
\end{aligned}$$

because $1/|\varkappa| = h(\phi) \geqslant 2$. Moreover, by (10.18), we have $\frac{1}{2}(1+\varkappa_1' n)|\varkappa| \leqslant |\varkappa'|$, so that

$$\sum_{k \lesssim \log|\xi|} 2^{-|\varkappa'|k} 2^{(1+\varkappa_1' n)|\varkappa|k/2} |\xi|^{-|\varkappa|} \leqslant C|\xi|^{-|\varkappa|}\log|\xi|,$$

and

$$\sum_{\substack{k \\ 2^{(1+\varkappa_1' n)k/2} > |\xi|}} 2^{-|\varkappa'|k} \leqslant C|\xi|^{-2|\varkappa'|/(1+\varkappa_1' n)} \leqslant |\xi|^{-|\varkappa|}.$$

This shows that also $J(\xi)$ given by (10.9) satisfies estimate (11.1), which completes the proofs of Theorems 11.1 and 1.9.

## 12. Proof of the remaining statements in the introduction and refined results

In this section, we shall prove the remaining results and claims that have been stated in the introduction.

### 12.1. Invariance of the notion of height $h(x^0, S)$ under affine transformations

We assume that $x^0=(0,0,1)=:e_3$ and $T_{x^0}=\{(x_1,x_2,x_3):x_3=0\}=:V$, and that our hypersurface $S$ is the graph

$$S=\{(x_1,x_2,1+\phi(x_1,x_2)):(x_1,x_2)\in\Omega\}$$

of a smooth function $1+\phi$ defined on an open neighborhood $\Omega$ of $0\in\mathbb{R}^2$ and satisfying the conditions

$$\phi(0,0)=0 \quad \text{and} \quad \nabla\phi(0,0)=0.$$

Consider an affine linear change of coordinates $F\colon u\mapsto w+Au$ of $\mathbb{R}^3$ which fixes the point $x^0$, i.e., $F(e_3)=e_3$, and so that the derivative $DF(x^0)$ leaves the tangent space $T_{x^0}S$ invariant, i.e., $A(V)=V$. Here, $A\in GL(3,\mathbb{R})$ and $w\in\mathbb{R}^3$ is a fixed translation vector. We then denote by $B:=A|_V$ the induced linear isomorphism of $V$. If we decompose $w=v+\mu e_3$, with $v\in V$ and $\mu\in\mathbb{R}$, and write elements of $\mathbb{R}^3$ as $(x,x_3)$, with $x\in\mathbb{R}^2$, then from $w+Ae_3=e_3$ one computes that

$$F(x,x_3)=(Bx+(1-x_3)v,\mu+(1-\mu)x_3).$$

Then

$$F(S)=\{(Bx-\phi(x)v,1+(1-\mu)\phi(x)):(x_1,x_2)\in\Omega\}.$$

Notice that $1-\mu\neq0$, since $F$ is assumed to be bijective. By our assumptions on $\phi$, the mapping $\varphi\colon x\mapsto y=Bx-\phi(x)v$ is a local diffeomorphism near the origin with $\varphi(0)=0$, and we can write $F(S)$ locally as the graph of the smooth function

$$1+\tilde\phi(y):=1+(1-\mu)\phi(\varphi^{-1}(y)).$$

Since $h(\phi)=h(\tilde\phi)$, we see that $h(x^0,S)=h(x^0,F(S))$, which proves the invariance of our notion of height $h(x^0,S)$ under affine linear changes of coordinates.

### 12.2. Proof of Proposition 1.7 and Remark 1.11 (a), and remarks on the critical exponent $p=h(x^0,S)$

We are first going to prove Proposition 1.7. As outlined in the introduction, we may assume without loss of generality that the hypersurface $S$ is given as the graph

$$S=\{(x_1,x_2,1+\phi(x_1,x_2)):(x_1,x_2)\in\Omega\}$$

of a smooth function $1+\phi$ defined on an open neighborhood $\Omega$ of $(0,0)\in\mathbb{R}^2$ and satisfying the conditions

$$\phi(0,0)=0 \quad \text{and} \quad \nabla\phi(0,0)=0,$$

and that $x^0=(0,0,1)$, so that the affine tangent plane $x^0+T_{x^0}S$ is $\{(x_1,x_2,x_3):x_3=1\}$. Then $d_{T,x^0}(x)=|\phi(x_1,x_2)|$, so that we have to show that for a neighborhood $\Omega$ of the origin, one has

$$\int_\Omega |\phi(x)|^{-1/p}(1+|\nabla\phi(x)|^2)^{1/2}\,dx=\infty,$$

or equivalently

$$\int_\Omega |\phi(x)|^{-1/p}\,dx=\infty \tag{12.1}$$

whenever $p<h(\phi)$. Moreover, if $\phi$ is analytic, then we need to show that (12.1) holds also for the critical exponent $p=h(\phi)$.

To this end, observe first that we may reduce ourselves to the case where the coordinates $x$ are adapted to $\phi$ by applying the change of coordinates (8.1) (cf. [41] and [19]) to the integral in (12.1). Recall that then one of the following three cases applies:

(a) $\pi(\phi)$ is a compact edge, and either $\varkappa_2/\varkappa_1\notin\mathbb{N}$, or $\varkappa_2/\varkappa_1\in\mathbb{N}$ and $m(\phi_{\mathrm{pr}})\leqslant d(\phi)$;

(b) $\pi(\phi)$ consists of a vertex;

(c) $\pi(\phi)$ is unbounded.

Moreover, in this case we have $h(\phi)=d(\phi_{\mathrm{pr}})$, where $\pi(\phi)$ denotes again the principal face of the Newton polyhedron $\mathcal{N}(\phi)$ and $\phi_{\mathrm{pr}}$ the principal part of $\phi$ (cf. (2.1)).

First, we consider the cases (a) and (b), where the principal face of the Newton polyhedron of $\phi$ is a compact set.

PROPOSITION 12.1. *If the principal face $\pi(\phi)$ of the Newton polyhedron of the function $\phi$, when expressed in adapted coordinates, is compact, then* (12.1) *holds for every $p\leqslant h(\phi)$.*

*Proof.* As in the proof of Corollary 7.2, we can choose a weight $\varkappa=(\varkappa_1,\varkappa_2)$ such that

$$h(\phi)=\frac{1}{|\varkappa|}=\frac{1}{(\varkappa_1+\varkappa_2)},$$

where $0<\varkappa_1\leqslant\varkappa_2$ without loss of generality. Then the $\varkappa$-principal part $\phi_\varkappa$ of the function $\phi$ is a weighted $\varkappa$-homogeneous polynomial of degree 1.

We may also assume that $\varkappa_1$ and $\varkappa_2$ are rational numbers. Then we can find even positive integers $q_1$ and $q_2$ and a positive integer $r$ such that $\varkappa_1=r/q_1$ and $\varkappa_2=r/q_2$.

The quasi-norm $N(x):=(x_1^{q_1}+x_2^{q_2})^{1/r}$ is then $\varkappa$-homogeneous of degree 1 and smooth away from the origin. Denote by $\Sigma:=\{(y_1,y_2):\varrho(y_1,y_2)=1\}$ the associated "unit circle", and let $(y_1(\theta),y_2(\theta))$, $0\leqslant\theta<1$, be a smooth parametrization of $\Sigma$. We can then introduce generalized polar coordinates $(\varrho,\theta)$ for $\mathbb{R}^2\setminus\{0\}$ by writing

$$x_1:=\varrho^{\varkappa_1}y_1(\theta)\quad\text{and}\quad x_2:=\varrho^{\varkappa_2}y_2(\theta),\quad \varrho>0.$$

It is well known and easy to see that the Lebesgue measure on $\mathbb{R}^2$ then decomposes as

$$dx_1\,dx_2=\varrho^{|\varkappa|-1}\,d\varrho\,d\gamma(\theta),$$

where $d\gamma(\theta)$ is a positive Radon measure such that $\int_\Sigma d\gamma(\theta)>0$. Let us also assume without loss of generality that $\Omega=\{(x_1,x_2):\varrho(x_1,x_2)<\varepsilon\}$, where $\varepsilon>0$.

If we now decompose $\phi=\phi_\varkappa+\phi_r$ as before into the $\varkappa$-principal part $\phi_\varkappa$ and the remainder term $\phi_r$, and express $\phi$ in polar coordinates $\tilde\phi(\varrho,\theta):=\phi(\varrho^{\varkappa_1}y_1(\theta),\varrho^{\varkappa_2}y_2(\theta))$, then

$$\tilde\phi(\varrho,\theta)=\varrho(\tilde\phi_\varkappa(1,\theta)+\tilde\phi_r(\varrho,\theta)),$$

where $\tilde\phi_r(\varrho,\theta)=O(\varrho^\delta)$ for some $\delta>0$ as $\varrho\to0$. In particular, also $\tilde\phi_r(\varrho,\theta)$ is bounded, which is all that we need. By passing to these polar coordinates, we obtain

$$\int_\Omega|\phi(x)|^{-1/h(\phi)}\,dx=\int_0^\varepsilon\frac{d\varrho}{\varrho}\int_\Sigma|\tilde\phi_\varkappa(1,\theta)+\tilde\phi_r(\varrho,\theta)|^{-1/h(\phi)}\,d\gamma(\theta)\geqslant c\int_0^\varepsilon\frac{d\varrho}{\varrho}.$$

In the last inequality $c$ is a positive constant and therefore the integral diverges. This proves the proposition. □

Case (c) where the principal face is unbounded remains to be considered.

PROPOSITION 12.2. *Assume that the principal face $\pi(\phi)$ of the Newton polyhedron of the function $\phi$, when expressed in adapted coordinates, is unbounded.*

(i) *Then* (12.1) *holds for every $p<h(\phi)$.*

(ii) *If $\phi$ is assumed to be analytic, then* (12.1) *holds also for $p=h(\phi)$.*

*Proof.* We first prove (i), so assume that $p<h(\phi)$. Here, we can apply a similar reasoning as in the proof of case (c) in Corollary 7.2. The principal face $\pi(\phi)$ is a horizontal half-line, with left endpoint $(\nu_1,N)$, where $\nu_1<N=h(\phi)$. Notice that $N\geqslant2$, since for $N=1$ we had $\nu_1=0$, which is not possible given our assumption $\nabla\phi(0,0)=0$. We can then choose $\varkappa$ with $0<\varkappa_1<\varkappa_2$ so that the line $\varkappa_1t_1+\varkappa_2t_2=1$ is a supporting line to the Newton polyhedron of $\phi$ and that the point $(\nu_1,N)$ is the only point of $\mathcal{N}(\phi)$ on this line. Moreover, we can choose $\varkappa_2/\varkappa_1$ as large as we wish, so that we may assume that

$$p<\frac{1}{|\varkappa|}<h(\phi).$$

Then the $\varkappa$-principal part $\phi_\varkappa$ of $\phi$ is of the form $\phi_\varkappa(x)=cx_1^{\nu_1}x_2^N$, with $c\neq0$, and it is $\varkappa$-homogeneous of degree 1.

By passing to generalized polar coordinates as in the proof of Proposition 12.1, we then see that

$$\int_\Omega|\phi(x)|^{-1/p}\,dx=\int_0^\varepsilon\frac{d\varrho}{\varrho^{1/p-|\varkappa|+1}}\int_\Sigma|\tilde\phi_\varkappa(1,\theta)+\tilde\phi_r(\varrho,\theta)|^{-1/p}\,d\gamma(\theta),$$

where again $\tilde{\phi}_r(\varrho,\theta)$ is bounded. Since $1/p-|\varkappa|>0$, we conclude that the last integral diverges.

In order to prove (ii), observe that if $\phi$ is analytic, then there exists a non-trivial analytic function $f$ near the origin so that $\phi(x_1,x_2)=x_2^N f(x_1,x_2)$, where again $N=h(\phi)$. Then, for sufficiently small $\varepsilon>0$, we have

$$\int_\Omega \frac{dx_1 dx_2}{|\phi(x_1,x_2)|^{1/h(\phi)}} \geqslant \int_{-\varepsilon}^{\varepsilon} \frac{dx_2}{|x_2|} \int_{-\varepsilon}^{\varepsilon} \frac{dx_1}{|f(x_1,x_2)|^{1/N}}.$$

Obviously the last integral diverges. □

*Remark* 12.3. If $\phi$ is a finite-type smooth function and the principal face is a non-compact set then the integral $\int_\Omega |\phi(x)|^{-1/h(\phi)}\,dx$ may be convergent.

An example is given by the function $\phi(x_1,x_2)=x_2^2+e^{-x_1^{-\alpha}}$ considered by A. Iosevich and E. Sawyer in [23]. Here we have $h(\phi)=2$, and the associated integral converges whenever $0<\alpha<1$. Correspondingly, it has been shown in [23] that the maximal operator associated with the hypersurface $x_3=1+x_2^2+e^{-x_1^{-\alpha}}$ is $L^2$ bounded whenever $0<\alpha<1$ and unbounded for $p<2$ (the latter statement follows of course also from Proposition 12.2). However, if $\alpha\geqslant 1$, then it is unbounded whenever $p\leqslant 2$.

We have thus obtained a confirmation of Iosevich–Sawyer's conjecture for analytic hypersurfaces [23], and for smooth finite-type hypersurfaces we have a partial confirmation of the conjecture. The conjecture remains open when $p=h(\phi)$ in the case where the principal face of $\phi$ is unbounded in an adapted coordinate system.

Let us finally indicate how to prove Remark 1.11 (a) (see also [27] for an analogous argument). We assume that $x^0=0$, and that $S$ is locally given near 0 as the graph of a function $\phi$, where $\phi(0,0)=\nabla\phi(0,0)=0$.

Consider again first cases (a) and (b). Similarly as in the proof of Proposition 12.1 we can choose a weight $\varkappa$ such that $\phi_{\mathrm{pr}}=\phi_\varkappa$ and $d(\phi)=1/|\varkappa|=1/(\varkappa_1+\varkappa_2)$. Then define for $r>0$ the function $f_r\in\mathcal{S}(\mathbb{R}^3)$ by

$$\hat{f}_r := \widehat{\chi}\Big(\frac{x_1}{r^{\varkappa_1}}\Big)\widehat{\chi}\Big(\frac{x_2}{r^{\varkappa_2}}\Big)\widehat{\chi}\Big(\frac{x_3}{r}\Big),$$

where $\chi$ is a Schwartz function such that $\widehat{\chi}\geqslant 0$ has compact support and $\widehat{\chi}(0)=1$. Since $\phi(r^{\varkappa_1}x_1,r^{\varkappa_2}x_2)=r(\phi_\varkappa(x_1,x_2)+O(r^\varepsilon))$ for some $\varepsilon>0$, where $\phi_\varkappa$ vanishes at the origin, one easily sees that, for $r$ sufficiently small,

$$\int_S |\hat{f}_r|^2\varrho\,d\sigma \gtrsim r^{|\varkappa|}\iint_{\mathbb{R}^2} |\hat{f}_r(r^{\varkappa_1}x_1, r^{\varkappa_2}x_2, \phi(r^{\varkappa_1}x_1, r^{\varkappa_2}x_2))|^2\,dx_1\,dx_2 \gtrsim r^{|\varkappa|}.$$

On the other hand, it is easy to check that $\|f_r\|_{L^p(\mathbb{R}^3)}\sim r^{(1+|\varkappa|)/p'}$, so that an estimate of the form $\|\hat{f}\|_{L^2(S)}\leqslant C\|f\|_{L^p(\mathbb{R}^3)}$ would imply that $r^{|\varkappa|/2}\lesssim r^{(1+|\varkappa|)/p'}$ for every sufficiently small $r$, and hence $p'\geqslant 2(1+1/|\varkappa|)=2(1+d(\phi))$.

In case (c), we choose $\varkappa$ similarly as in the proof of Proposition 12.2 such that $1/|\varkappa|<d(\phi)$ is arbitrarily close to $d(\phi)$. Arguing then as before, we find that necessarily $p'\geqslant 2(1+1/|\varkappa|)$ for every such $\varkappa$, and hence again $p'\geqslant 2(1+d(\phi))$.

### 12.3. Proof of Theorem 1.12

By means of a smooth partition of unity consisting of non-negative functions, we may reduce ourselves to the situation where $\varrho$ is supported in a sufficiently small neighborhood of some given point $z\in S$. Without loss of generality we may then assume that $z=0$, and that our hypersurface $S$ is the graph

$$S=\{(x,\phi(x)):x\in\Omega\}$$

of a smooth function $\phi$ defined on an open neighborhood $\Omega$ of $0\in\mathbb{R}^{n-1}$ and satisfying the conditions

$$\phi(0)=0 \quad \text{and} \quad \nabla\phi(0)=0.$$

Then the Fourier transform $\widehat{\varrho\,d\sigma}(0,...,0,\lambda)$ of the surface carried measure $\varrho\,d\sigma$ in direction of the unit normal to $S$ at $z=0$ is an oscillatory integral of the form

$$J(\lambda)=\int_{\mathbb{R}^{n-1}} e^{-i\lambda\phi(x)}\eta(x)\,dx,$$

where $0\leqslant\eta\in C_0^\infty(\Omega)$. By (1.9), we have in particular that

$$|J(\lambda)|\leqslant C_\beta(1+|\lambda|)^{-\beta} \quad \text{for every } \lambda\in\mathbb{R}, \tag{12.2}$$

where $\beta>0$. Crucial for us is the next result, which can be found in a discussion on p. 539 in [32]. We remark that related connections between the oscillation index and the function of $\gamma$ defined by the integral in (12.3) played a crucial role already in the classical work by Varchenko [41] (see also [3]).

LEMMA 12.4. (Phong–Stein–Sturm) *If* (12.2) *holds true*, *then*

$$\int_{\mathbb{R}^{n-1}} |\phi(x)|^{-\gamma}\eta(x)\,dx<\infty \tag{12.3}$$

*for every* $\gamma<1$ *such that* $\gamma<\beta$.

Theorem 1.12 is now an easy consequence of Lemma 12.4. Indeed, by Remark 1.6 it suffices to prove the estimate (1.10) only for affine tangent planes $H=z+T_zS$ to $S$, where $z\in S$ is sufficiently close to the support of $\varrho$. For these, the previous reasoning applies, and since then $d_H(x)=|\phi(x)|$, we see that (1.10) is an immediate consequence of Lemma 12.4.

*Remark* 12.5. By the same reasoning, Lemma 12.4 also shows that if $z\in S$, and if $0<\beta\in\mathfrak{B}(z,S)$ and $\gamma<\min\{1,\beta\}$, then $\gamma\in\mathfrak{C}(z,S)$.

### 12.4. Proof of Corollary 1.13

Note first that always $h(x^0,S)\geqslant 1$. We first assume that $h(x^0,S)>1$. If we had

$$\beta > \frac{1}{h(x^0,S)},$$

then we could choose some $p>1$ in this case such that

$$\beta > \frac{1}{p} > \frac{1}{h(x^0,S)}.$$

Then Theorem 1.12 in combination with Proposition 1.7 would imply that $p\leqslant 1/\beta$, a contradiction.

The case where $h(x^0,S)=1$ remains to be considered. We may again assume that $S$ is given as the graph of a smooth function $\phi$, with $\phi$ satisfying (1.3) and $x^0=(0,0,0)$. Assuming without loss of generality that the coordinates are adapted to $\phi$, it is then easy to see that the Hessian matrix $D^2\phi(0,0)$ is non-degenerate. The asymptotic form of the method of stationary phase then shows that $\gamma\leqslant 1=1/h(\phi)=1/h(x^0,S)$.

### 12.5. Proof of Theorem 1.14

Let $S$ be a smooth, finite-type hypersurface in $\mathbb{R}^3$, and let $x^0\in S$ be given. Notice first that Theorem 1.9 implies that

$$\beta_u(x^0,S) \geqslant \frac{1}{h(x^0,S)}.$$

Moreover, by Corollary 1.13 we have $\beta_u(x^0,S)\leqslant 1/h(x^0,S)$. Indeed, since its proof was based on Proposition 1.7, which made only use of the affine tangent hyperplane at the point $x^0$, with the same arguments restricted to these tangent hyperplane we even obtain that

$$\beta(x^0,S) \leqslant \frac{1}{h(x^0,S)}.$$

In combination with (1.12) these estimates imply that

$$\beta_u(x^0,S) = \beta(x^0,S) = \frac{1}{h(x^0,S)} \leqslant 1. \tag{12.4}$$

Observe next that if $\beta\in\mathfrak{B}_u(x^0,S)$, then by Theorem 1.12 and (12.4) we have $\beta\leqslant 1$, and then $\beta-\varepsilon\in\mathfrak{C}_u(x^0,S)$ for every sufficiently small $\varepsilon>0$. This implies that

$$\beta_u(x^0,S) \leqslant \gamma_u(x^0,S),$$

and hence, by (12.4) and (1.14),

$$\frac{1}{h(x^0,S)} \leqslant \gamma_u(x^0,S) \leqslant \gamma(x^0,S). \tag{12.5}$$

Finally, if $\gamma \in \mathfrak{C}(x^0,S)$, then putting $p:=1/\gamma$ in Proposition 1.7 yields $1/\gamma \geqslant h(x^0,S)$, and hence $\gamma \leqslant 1/h(x^0,S)$. This implies that $\gamma(x^0,S) \leqslant 1/h(x^0,S)$, and in combination with (12.5), we also get that

$$\gamma(x^0,S) = \gamma_u(x^0,S) = \frac{1}{h(x^0,S)}.$$

This concludes the proof of Theorem 1.14.

*Acknowledgement.* We wish to express our gratitude to the referee whose comments and suggestions were most helpful.

## References


[1] Arkhipov, G. I., Karatsuba, A. A. & Chubarikov, V. N., Trigonometric integrals. *Izv. Akad. Nauk SSSR Ser. Mat.*, 43 (1979), 971–1003, 1197 (Russian); English translation in *Math. USSR–Izv.*, 15 (1980), 211–239.

[2] Arnold, V. I., Remarks on the method of stationary phase and on the Coxeter numbers. *Uspekhi Mat. Nauk*, 28 (1973), 17–44 (Russian); English translation in *Russian Math. Surveys*, 28 (1973), 19–48.

[3] Arnold, V. I., Guseĭn-Zade, S. M. & Varchenko, A. N., *Singularities of Differentiable Maps.* Vol. II. Monographs in Mathematics, 83. Birkhäuser, Boston, MA, 1988.

[4] Bourgain, J., Averages in the plane over convex curves and maximal operators. *J. Anal. Math.*, 47 (1986), 69–85.

[5] Bruna, J., Nagel, A. & Wainger, S., Convex hypersurfaces and Fourier transforms. *Ann. of Math.*, 127 (1988), 333–365.

[6] Carbery, A., Wainger, S. & Wright, J., Singular integrals and the Newton diagram. *Collect. Math.*, Vol. Extra (2006), 171–194.

[7] Chester, C., Friedman, B. & Ursell, F., An extension of the method of steepest descents. *Proc. Cambridge Philos. Soc.*, 53 (1957), 599–611.

[8] Cowling, M. & Mauceri, G., Inequalities for some maximal functions. II. *Trans. Amer. Math. Soc.*, 296 (1986), 341–365.

[9] — Oscillatory integrals and Fourier transforms of surface carried measures. *Trans. Amer. Math. Soc.*, 304 (1987), 53–68.

[10] Domar, Y., On the Banach algebra $A(G)$ for smooth sets $\Gamma \subset R^n$. *Comment. Math. Helv.*, 52 (1977), 357–371.

[11] Duistermaat, J. J., Oscillatory integrals, Lagrange immersions and unfolding of singularities. *Comm. Pure Appl. Math.*, 27 (1974), 207–281.

[12] Greenblatt, M., Newton polygons and local integrability of negative powers of smooth functions in the plane. *Trans. Amer. Math. Soc.*, 358 (2006), 657–670.

[13] — The asymptotic behavior of degenerate oscillatory integrals in two dimensions. *J. Funct. Anal.*, 257 (2009), 1759–1798.

[14] — Oscillatory integral decay, sublevel set growth, and the Newton polyhedron. *Math. Ann.*, 346 (2010), 857–895.
[15] — Resolution of singularities, asymptotic expansions of oscillatory integrals, and related phenomena. Preprint, 2007. `arXiv:0709.2496v2 [math.CA]`.
[16] Greenleaf, A., Principal curvature and harmonic analysis. *Indiana Univ. Math. J.*, 30 (1981), 519–537.
[17] Hörmander, L., *The Analysis of Linear Partial Differential Operators.* I. Grundlehren der Mathematischen Wissenschaften, 256. Springer, Berlin–Heidelberg, 1990.
[18] Ikromov, I. A., Kempe, M. & Müller, D., Damped oscillatory integrals and boundedness of maximal operators associated to mixed homogeneous hypersurfaces. *Duke Math. J.*, 126 (2005), 471–490.
[19] Ikromov, I. A. & Müller, D., On adapted coordinate systems. To appear in *Trans. Amer. Math. Soc.*
[20] — Uniform estimates for the Fourier transform of surface carried measures and a sharp $L^p$-$L^2$ Fourier restriction theorem for hypersurfaces in $\mathbb{R}^3$. In preparation.
[21] Iosevich, A., Maximal operators associated to families of flat curves in the plane. *Duke Math. J.*, 76 (1994), 633–644.
[22] Iosevich, A. & Sawyer, E., Oscillatory integrals and maximal averages over homogeneous surfaces. *Duke Math. J.*, 82 (1996), 103–141.
[23] — Maximal averages over surfaces. *Adv. Math.*, 132 (1997), 46–119.
[24] Iosevich, A., Sawyer, E. & Seeger, A., On averaging operators associated with convex hypersurfaces of finite type. *J. Anal. Math.*, 79 (1999), 159–187.
[25] Karpushkin, V. N., A theorem on uniform estimates for oscillatory integrals with a phase depending on two variables. *Trudy Sem. Petrovsk.*, 10 (1984), 150–169, 238 (Russian); English translation in *J. Soviet Math.*, 35 (1986), 2809–2826.
[26] Levinson, N., Transformation of an analytic function of several variables to a canonical form. *Duke Math. J.*, 28 (1961), 345–353.
[27] Magyar, Á., On Fourier restriction and the Newton polygon. *Proc. Amer. Math. Soc.*, 137 (2009), 615–625.
[28] Mockenhaupt, G., Seeger, A. & Sogge, C. D., Wave front sets, local smoothing and Bourgain's circular maximal theorem. *Ann. of Math.*, 136 (1992), 207–218.
[29] — Local smoothing of Fourier integral operators and Carleson–Sjölin estimates. *J. Amer. Math. Soc.*, 6 (1993), 65–130.
[30] Nagel, A., Seeger, A. & Wainger, S., Averages over convex hypersurfaces. *Amer. J. Math.*, 115 (1993), 903–927.
[31] Phong, D. H. & Stein, E. M., The Newton polyhedron and oscillatory integral operators. *Acta Math.*, 179 (1997), 105–152.
[32] Phong, D. H., Stein, E. M. & Sturm, J. A., On the growth and stability of real-analytic functions. *Amer. J. Math.*, 121 (1999), 519–554.
[33] Randol, B., On the asymptotic behavior of the Fourier transform of the indicator function of a convex set. *Trans. Amer. Math. Soc.*, 139 (1969), 279–285.
[34] Schulz, H., Convex hypersurfaces of finite type and the asymptotics of their Fourier transforms. *Indiana Univ. Math. J.*, 40 (1991), 1267–1275.
[35] Sogge, C. D., *Fourier Integrals in Classical Analysis.* Cambridge Tracts in Mathematics, 105. Cambridge University Press, Cambridge, 1993.
[36] — Maximal operators associated to hypersurfaces with one nonvanishing principal curvature, in *Fourier Analysis and Partial Differential Equations* (Miraflores de la Sierra, 1992), Stud. Adv. Math., pp. 317–323. CRC, Boca Raton, FL, 1995.

[37] Sogge, C. D. & Stein, E. M., Averages of functions over hypersurfaces in $\mathbf{R}^n$. *Invent. Math.*, 82 (1985), 543–556.

[38] Stein, E. M., Maximal functions. I. Spherical means. *Proc. Nat. Acad. Sci. U.S.A.*, 73 (1976), 2174–2175.

[39] — *Harmonic Analysis*: *Real-Variable Methods*, *Orthogonality*, *and Oscillatory Integrals*. Princeton Mathematical Series, 43. Princeton University Press, Princeton, NJ, 1993.

[40] Svensson, I., Estimates for the Fourier transform of the characteristic function of a convex set. *Ark. Mat.*, 9 (1971), 11–22.

[41] Varchenko, A. N., Newton polyhedra and estimates of oscillatory integrals. *Funktsional. Anal. i Prilozhen.*, 10 (1976), 13–38 (Russian); English translation in *Functional Anal. Appl.*, 18 (1976), 175–196.

[42] Vasil′ev, B. A., The asymptotic behavior of exponential integrals, the Newton diagram and the classification of minima. *Funktsional. Anal. i Prilozhen.*, 11 (1977), 1–11, 96 (Russian); English translation in *Functional Anal. Appl.*, 11 (1977), 163–172.

Isroil A. Ikromov
Department of Mathematics
Samarkand State University
University Boulevard 15
14014, Samarkand
Uzbekistan
ikromov1@rambler.ru

Michael Kempe
Mathematisches Seminar,
C.A.-Universität Kiel
Ludewig-Meyn-Strasse 4
DE-24098 Kiel
Germany
kempe@math.uni-kiel.de

Detlef Müller
Mathematisches Seminar
C.A.-Universität Kiel
Ludewig-Meyn-Strasse 4
DE-24098 Kiel
Germany
mueller@math.uni-kiel.de

*Received February 18, 2008*
*Received in revised form September 7, 2009*