



Resolution Analysis of Inverting the Generalized N -Dimensional Radon Transform in $\mathbb{R}^n$ from Discrete Data

Alexander Katsevich¹

Received: 7 February 2022 / Revised: 16 September 2022 / Accepted: 30 September 2022 /
Published online: 28 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Let $\mathcal{R}$ denote the generalized Radon transform, which integrates over a family of N -dimensional smooth submanifolds $\mathcal{S}_{\tilde{y}} \subset \mathcal{U}$, $1 \leq N \leq n-1$, where an open set $\mathcal{U} \subset \mathbb{R}^n$ is the image domain. The submanifolds are parametrized by points $\tilde{y} \in \tilde{\mathcal{V}}$, where an open set $\tilde{\mathcal{V}} \subset \mathbb{R}^n$ is the data domain. We assume that the canonical relation $\tilde{C}$ from $T^*\mathcal{U}$ to $T^*\tilde{\mathcal{V}}$ of $\mathcal{R}$ is a local canonical graph (when $\mathcal{R}$ is viewed as a Fourier Integral Operator). The continuous data are denoted by g , and the reconstruction is $\check{f} = \mathcal{R}^* \mathcal{B} g$. Here $\mathcal{R}^*$ is a weighted adjoint of $\mathcal{R}$, $\mathcal{B}$ is a pseudo-differential operator, and g is a conormal distribution. Discrete data consists of the values of g on a regular lattice with step size $O(\epsilon)$. Let $\mathcal{S}$ denote the singular support of $\check{f}$, and $\check{f}_\epsilon = \mathcal{R}^* \mathcal{B} g_\epsilon$ be the reconstruction from interpolated discrete data $g_\epsilon(\tilde{y})$. Pick a point $x_0 \in \mathcal{S}$, i.e. the singularity of $\check{f}$ at x_0 is visible from the data. The main result of the paper is the computation of the limit

$$\text{DTB}(\check{x}) := \lim_{\epsilon \rightarrow 0} \epsilon^\kappa \check{f}_\epsilon(x_0 + \epsilon \check{x}).$$

Here $\kappa \geq 0$ is selected based on the strength of the reconstructed singularity, and $\check{x}$ is confined to a bounded set. The limiting function $\text{DTB}(\check{x})$, which we call the discrete transition behavior, contains full information about the resolution of reconstruction.

Keywords Generalized Radon transform · Discrete data · Resolution analysis · Fourier integral operators · Singularities

Communicated by Todd Quinto.

This work was supported in part by NSF Grant DMS-1906361.

✉ Alexander Katsevich
Alexander.Katsevich@ucf.edu

¹ Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA

Mathematics Subject Classification 44A12 · 65R10 · 92C55

1 Introduction

Analysis of resolution of tomographic reconstruction from discrete Radon transform data is a practically important problem. In many applications one needs to know how accurately and with what resolution singularities of the object f (e.g., a jump discontinuity across a smooth surface $\mathcal{S} = \text{singsupp}(f)$) are reconstructed. Let $\check{f}$ denote the reconstruction from continuous data, and $\check{f}_\epsilon$ denote the reconstruction from discrete data, where ϵ represents the data sampling rate. In the latter case, interpolated discrete data are substituted into the “continuous” inversion formula. In [17–20] the author initiated an analysis of reconstruction, which is focused on the behavior of $\check{f}_\epsilon$ near $\mathcal{S}$. One of the main results of these papers is the computation of the limit

$$\text{DTB}(\check{x}) := \lim_{\epsilon \rightarrow 0} \epsilon^\kappa \check{f}_\epsilon(x_0 + \epsilon \check{x}). \quad (1.1)$$

Here $x_0 \in \mathcal{S}$ (x_0 is selected subject to some constraints, see Definition 3.4 below), $\kappa \geq 0$ is a unique number selected based on the strength of the singularity of $\check{f}$ at x_0 (see 4.19), and $\check{x}$ is confined to a bounded set. Let f_0 be the leading singularity of $\check{f}$ in a neighborhood of x_0 (see Definition 5.2). For example, if $\check{f}$ is a conormal distribution with a homogeneous top symbol, then f_0 is the distribution determined by the top symbol. If f_0 is a homogeneous distribution of degree $-\kappa$, i.e., $f_0(t\check{x}) = t^{-\kappa} f_0(\check{x})$, then this value of κ is used in (1.1).

It is important to emphasize that both the size of the neighborhood around x_0 and the data sampling rate go to zero simultaneously in (1.1). The limiting function $\text{DTB}(\check{x})$, which we call the discrete transition behavior (or DTB for short), contains complete information about the resolution of reconstruction. The limit in (1.1) is computed for a fixed x_0 , so the dependence of the DTB, f_0 , and κ on x_0 is omitted for simplicity.

The DTB in (1.1) is a complete description of the reconstruction from discrete data in a neighborhood of a singularity. To put it simply, DTB is an accurate estimate of the reconstruction itself, which is the most one can ever obtain in resolution analysis. Conventional measures of resolution such as Full Width at Half Maximum (FWHM), line pairs per unit length, characteristic scale, etc., are a single number each. Once the full DTB function is computed, getting any desired resolution measurement from it (i.e., converting the DTB into a single number) is trivial. See Remark 4.10 for an example.

The results obtained to date can be summarized as follows. Even though we study reconstruction from discrete data, the classification of the cases is based on their continuous analogues. In [17] we find $\text{DTB}(\check{x})$ for the Radon transform in $\mathbb{R}^2$ in two cases: f is static and f changes during the scan (dynamic tomography). In the static case the reconstruction formula is exact (i.e., $\check{f} = f$), and in the dynamic case the reconstruction formula is quasi-exact (i.e., $\check{f} - f$ is smoother than f). In [18] we find $f_0(\check{x})$ for the classical Radon transform (CRT) in $\mathbb{R}^3$ assuming the reconstruction is exact and f has jumps. In [20] we consider a similar setting as in [18], i.e., f has

Table 1 Summary of the cases considered prior to this paper

	RT type	Dimension	Type of inversion	Singularity of f
[17]	CRT & GRT	2	Exact/quasi-exact	Jumps
[18]	CRT	3	Exact	Jumps
[20]	GRT	3	Quasi-exact	Jumps
[19]	CRT	n	More general	More general

jumps and reconstruction is quasi-exact, but consider a wide family of generalized Radon transforms (GRT) in $\mathbb{R}^3$. Finally, in [19], the data still comes from the classical Radon transform, but the dimension is increased to $\mathbb{R}^n$, the reconstruction operators are more general, and f may have singularities other than jumps. See Table 1 for a summary of the cases.

Let $\mathcal{R}$ denote the GRT, which integrates over a family of N -dimensional smooth submanifolds $\mathcal{S}_{\tilde{y}} \subset \mathcal{U} \subset \mathbb{R}^n$, $1 \leq N \leq n - 1$. When integration is performed over affine subspaces and $N < n - 1$, the GRT is known as the N -plane transform. If $N = 1$, the GRT is called the ray (or, X-ray) transform. The open set $\mathcal{U}$ represents the image domain. The submanifolds $\mathcal{S}_{\tilde{y}}$ are parametrized by points $\tilde{y} \in \tilde{\mathcal{V}}$, where an open set $\tilde{\mathcal{V}} \subset \mathbb{R}^n$ is the data domain. Our only other condition on $\mathcal{R}$ (besides that $\mathcal{S}_{\tilde{y}}$ be embedded manifolds, see Assumption 3.1(G1)) is that the canonical relation $\tilde{C}$ from $T^*\mathcal{U}$ to $T^*\tilde{\mathcal{V}}$ of $\mathcal{R}$ be a local canonical graph (see Assumption 3.1(G2) and Sect. 5.1). Here we view $\mathcal{R}$ as a Fourier Integral Operator (FIO). Assumption 3.1(G2) implies also that all the singularities of f microlocally near $(x_0, \xi_0) \in T^*\mathcal{U}$, where ξ_0 is conormal to $\mathcal{S}$ at x_0 , are visible in the GRT data $\mathcal{R}f(\tilde{y})$, $\tilde{y} \in \tilde{\mathcal{V}}$ (see Remark 3.2).

Reconstruction from continuous data $g = \mathcal{R}f$ is achieved by $\check{f} = \mathcal{R}^*\mathcal{B}g$. Here $\mathcal{R}^*$ is a weighted adjoint of $\mathcal{R}$, which integrates over submanifolds $\tilde{\mathcal{T}}_x := \{\tilde{y} \in \tilde{\mathcal{V}} : x \in \mathcal{S}_{\tilde{y}}\}$, and $\mathcal{B}$ is a fairly general pseudo-differential operator (Ψ DO). In fact, g does not even have to be the GRT of some f . All we need is that g be a sufficiently regular conormal distribution associated with a smooth hypersurface $\Gamma \subset \tilde{\mathcal{V}}$ [13, Sect. 18.2]. The data g are sampled on a regular lattice $\tilde{y}_j = \epsilon Dj$, $j \in \mathbb{Z}^n$, covering $\tilde{\mathcal{V}}$, where D is a sampling matrix.

To illustrate the effect of $\mathcal{B}$ on reconstruction, suppose $g = \mathcal{R}f$, where f is a sufficiently regular conormal distribution associated with a smooth hypersurface $\mathcal{S} \subset \mathcal{U}$. The choice of $\mathcal{B}$ determines whether the reconstruction is quasi-exact (i.e., $\check{f} - f$ is smoother than f), preserves the order of singularities of f ($\check{f}$ and f are in the same Sobolev space), or is singularity-enhancing ($\check{f}$ is more singular than f). A common example of the latter is Lambda (also known as local) tomography [8, 27].

The setting considered in this paper includes all the cases considered previously [17–20], and is substantially more general than before. In particular, in the previous work we always had $N = n - 1$. Now, N can be any integer $1 \leq N \leq n - 1$. This includes the practically most important case of cone beam CT: $n = 3$ and $N = 1$, on which the overwhelming majority of all medical, industrial, and security CT scans are based (see e.g., [15, 24] and references therein).

The main result of this paper is the derivation of the DTB (1.1) under these general conditions (see Theorem 4.7). Our result shows that even though g is sampled on a regular lattice, due to the geometric properties of the GRT, the resolution is both location- and direction-dependent (see Remark 4.8). We also show that the DTB equals to the convolution of the continuous transition behavior (CTB) with the suitably scaled classical Radon transform of the interpolation kernel (see Theorem 5.4). Loosely speaking, the CTB is the continuous analogue of the DTB (see Definition 5.3):

$$\text{CTB}(\check{x}) := \lim_{\epsilon \rightarrow 0} \epsilon^\kappa \check{f}(x_0 + \epsilon \check{x}). \quad (1.2)$$

To put it differently, $\text{CTB}(\check{x})$ is the leading singularity of the reconstruction $\check{f}$ at x_0 (the same as f_0 mentioned above). Since the reconstruction is not always intended to compute f or its singularities exactly (e.g., for singularity-enhancing reconstructions), it is important to know what the CTB look like in these more general situations.

The operator $\mathcal{R}^*$ (and, of course, $\mathcal{R}$ as well) can be viewed as an FIO, which is associated to a phase function linear in the frequency variables (see [12, Sect. 2.4] and [10, Sect. 1.3]):

$$(\mathcal{R}^*g)(x) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \int_{\tilde{\mathcal{V}}} e^{i\mu \cdot \Psi(x, \tilde{y})} w(x, \tilde{y}) g(\tilde{y}) d\tilde{y} d\mu. \quad (1.3)$$

Here $w \in C_0^\infty(\mathcal{U} \times \tilde{\mathcal{V}})$, and $\Psi \in C^\infty(\mathcal{U} \times \tilde{\mathcal{V}})$ is any $\mathbb{R}^{n-N}$ valued function that satisfies some nondegeneracy conditions (so that $\tilde{\mathcal{C}}$ is a local canonical graph). Any such Ψ determines a pair $\mathcal{R}, \mathcal{R}^*$ by setting $\mathcal{S}_{\tilde{\mathcal{Y}}} = \{x \in \mathcal{U} : \Psi(x, \tilde{y}) = 0\}$, $\tilde{\mathcal{T}}_x = \{\tilde{y} \in \tilde{\mathcal{V}} : \Psi(x, \tilde{y}) = 0\}$, and selecting integration weights to ensure that $\mathcal{R}$ and $\mathcal{R}^*$ are properly supported FIOs. As is easily seen, any properly supported FIO $F : \mathcal{E}'(\tilde{\mathcal{V}}) \rightarrow \mathcal{D}'(\mathcal{U})$ with the same phase can be represented in the form $F = \mathcal{R}^* \mathcal{B}$ modulo a regularizing operator for some $\mathcal{B}$ (at least, microlocally where $\mathcal{R}^*$ is elliptic). Indeed, we can just take $\mathcal{B} = (\mathcal{R}^*)^{-1} F$, where $(\mathcal{R}^*)^{-1}$ is a (left and right) parametrix for $\mathcal{R}^*$ (see [6, Proposition 5.1.2]). We assume here that an appropriate cut-off is introduced in $(\mathcal{R}^*)^{-1}$, so the composition is well-defined. Also, we do not worry about global conditions on $\mathcal{R}$, because $\mathcal{U}$ and $\tilde{\mathcal{V}}$ are sufficiently small. Then $\mathcal{B}$ is a Ψ DO (its canonical relation $\tilde{\mathcal{C}} \circ \tilde{\mathcal{C}}^t$ is the diagonal from $T^*\tilde{\mathcal{V}}$ to $T^*\tilde{\mathcal{V}}$), and $F - \mathcal{R}^* \mathcal{B}$ is regularizing. Thus, the reconstruction algorithm is the application of an FIO F with a phase function, which is linear in the frequency variables, to discrete data $g(\tilde{y}_j)$.

We emphasize $\mathcal{R}^*$ when discussing parallels between $\mathcal{R}^*$ and FIOs, because in this paper we investigate the resolution of computing $\mathcal{R}^* \mathcal{B}$ from discrete data (and not of $\mathcal{R}$).

Various methods for applying FIOs to discrete data have been proposed, see e.g., [3–5, 35] and references therein. This appears to be the *first analysis of resolution* of the reconstructed image Fg for fairly general classes of FIOs F and (conormal) distributions g . Some results along this direction are in [20]. Here $g = \mathcal{I}f$, where $\mathcal{I}$ is an imaging operator (frequently an FIO), and f is the unknown original object. Analyses of such sort are especially important, because they apply not only when an exact inversion formula for $\mathcal{I}$ is known (e.g., when $\mathcal{I}$ is the classical Radon transform),

but even when no such formula exists (e.g., when $\mathcal{I}$ is a weighted GRT integrating over nonplanar submanifolds). In the latter cases a common approach is to use a parametrix for $\mathcal{I}$ as the reconstruction operator F , so Fg accurately recovers only the singularities of f . Reconstruction of the smooth part of f with this approach is usually not accurate even if the data are ideal (i.e., known exactly everywhere). See, for example, Remark 1 in [29]. Our approach, which we call local resolution analysis, is well suited to the analysis of such linear reconstruction algorithms because the analysis is localized to an immediate neighborhood of the singularities of f .

Let g be a conormal distribution associated with a smooth hypersurface $\tilde{\Gamma}$, i.e., $WF(g) \subset N^*\tilde{\Gamma}$, the latter is the conormal bundle of $\tilde{\Gamma}$. Even if g is not in the range of $\mathcal{R}$, our assumptions ensure that there is a smooth hypersurface $\mathcal{S} \subset \mathcal{U}$ such that (1) $N^*\tilde{\Gamma} = \tilde{C} \circ N^*\mathcal{S}$, and (2) $WF(\check{f}) \subset N^*\tilde{\mathcal{S}}$. Let $\tilde{\mathcal{T}}_{\mathcal{S}}$ be the set of all $\tilde{y} \in \tilde{\mathcal{V}}$ such that $\mathcal{S}_{\tilde{y}}$ is tangent to $\mathcal{S}$. As is well-known, $\tilde{\Gamma} = \tilde{\mathcal{T}}_{\mathcal{S}}$.

A common thread through our work is that the well-behaved DTB (i.e., the limit in (1.1)) is guaranteed to exist only if a pair $(x_0, \tilde{y}_0) \in \mathcal{U} \times \tilde{\mathcal{V}}$ is generic. Here $x_0 \in \mathcal{S}$, and $\tilde{y}_0 \in \tilde{\Gamma}$ is the data point from which the singularity of $\check{f}$ at x_0 is visible, i.e. $\mathcal{S}_{\tilde{y}_0}$ is tangent to $\mathcal{S}$ at x_0 . Roughly, the pair is generic if in a small neighborhood of $\tilde{y}_0$ the sampling lattice $\tilde{y}^j$ is in general position relative to a local patch of $\tilde{\Gamma}$ containing $\tilde{y}_0$ (see Definition 3.4 for a precise statement). The property of a pair to be generic is closely related with the uniform distribution theory [21].

If $(x_0, \tilde{y}_0)$ is not generic, the DTB may be different from the generic one predicted by our theory, and certain non-local artifacts that depend on the shape of $\mathcal{S}$ can appear as well (see e.g. [19]) even if $\mathcal{R}^*\mathcal{R}$ is a Ψ DO. This shows also that the case of discrete data is more complicated than when the data are continuous, because in the latter case $WF(\check{f}) \subset WF(f)$ whenever $\mathcal{R}^*\mathcal{R}$ is a Ψ DO.

Alternative approaches to study resolution are in the framework of the sampling theory. The key assumption in these approaches is that f be essentially bandlimited in the classical sense [7, 23, 25] or in the semiclassical sense [22, 31, 32]. However, the methodologies of these approaches are quite different from ours, and the results obtained are different as well. The latter include sampling rate required to resolve details of a given size, and analysis of aliasing artifact if the sampling requirements are violated.

The paper is organized as follows. In Sect. 2 we introduce the GRT $\mathcal{R}$ and its adjoint $\mathcal{R}^*$, the sampling matrix D , the sampling lattice $\tilde{y}^j = \epsilon Dj$, $j \in \mathbb{Z}^n$, and fix a pair $(x_0, \tilde{y}_0) \in \mathcal{U} \times \mathcal{V}$ such that $\mathcal{S}_{\tilde{y}_0}$ is tangent to $\mathcal{S}$ at x_0 . In Sect. 3 we select convenient coordinates both in the data and image domains, state the main geometric assumptions about $\mathcal{R}$ and the shape of $\mathcal{S}$, and define a generic pair $(x_0, \tilde{y}_0)$. In Sect. 4 we formulate the main assumptions about the operator $\mathcal{B}$, interpolation kernel φ , and data function g . Essentially, $\mathcal{B}$ is a Ψ DO with a homogeneous top symbol. Likewise, g is a conormal distribution with a homogeneous top symbol associated with a smooth hypersurface $\tilde{\Gamma}$. The top symbol decays sufficiently fast, so g is a continuous function. We do not require that g be in the range of $\mathcal{R}$. We also give a formula for κ in terms of N and the orders of $\mathcal{B}$ and g (see 4.19). Then we state our main result as Theorem 4.7, where explicit formulas for the DTB are provided.

In Sect. 5 we look at $\mathcal{R}$ as an FIO, and discuss some of our assumptions from the FIO perspective. We also state a theorem about the relationship between the DTB and CTB (Theorem 5.4), and provide some intuition behind our results.

The proof of Theorem 4.7 is spread over Sects. 6–11. Preliminary results are in Sects. 6 and 7. In Sect. 6 we show that $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{T}_{x_0}$ are tangent at y_0 and investigate the contact. All calculations are done in the new y -coordinates, so we drop the tildas in $\tilde{y}$, $\tilde{\Gamma}$, $\tilde{\mathcal{V}}$, etc. Let g_ϵ be the interpolated data (see 4.6). In Sect. 7 we obtain various bounds on g , g_ϵ , $g - g_\epsilon$, and their derivatives. The core of the proof is in Sects. 8–11. To help the reader, Sect. 8.1 describes the main ideas of the proof and outlines what is done in each of the Sects. 8–11.

Theorem 5.4 is proven in Sect. 12. In Appendix B we show that our assumptions about g are reasonable. For example, they are satisfied when f is a conormal distribution associated with a smooth hypersurface $\mathcal{S}$. The fact that $g = \mathcal{R}f$ is conormal follows from the calculus of FIOs, see e.g. [33, Sect. VIII.5]. We present the necessary calculations here, because they are short, make the paper self-contained, and these calculations are used elsewhere in the paper. Finally, proofs of various auxiliary lemmas are collected in the other appendices.

2 Preliminaries

Let $\tilde{\Phi}(t, \tilde{y}) \in C^\infty(\mathbb{R}^N \times \tilde{\mathcal{V}})$ be a defining function for the GRT $\mathcal{R}$. Here $t \in \mathbb{R}^N$ is an auxiliary variable that parametrizes smooth manifolds $\mathcal{S}_{\tilde{y}} := \{x \in \mathcal{U} : x = \tilde{\Phi}(t, \tilde{y}), t \in \mathbb{R}^N\}$ over which $\mathcal{R}$ integrates, an open set $\mathcal{U} \subset \mathbb{R}^n$ is the image domain, $\tilde{y} \in \tilde{\mathcal{V}}$ is the data domain variable, and an open set $\tilde{\mathcal{V}} \subset \mathbb{R}^n$ is the data domain. Both $\mathcal{U}$ and $\tilde{\mathcal{V}}$ are endowed by the usual Euclidean metric. The corresponding GRT is given by

$$\mathcal{R}f(\tilde{y}) = \int_{\mathcal{S}_{\tilde{y}}} f(x)b(x, \tilde{y})dx, \quad \tilde{y} \in \tilde{\mathcal{V}}, \quad (2.1)$$

where $b \in C_0^\infty(\mathcal{U} \times \tilde{\mathcal{V}})$, $dx = (\det G^{\mathcal{S}}(t, \tilde{y}))^{1/2}dt$ is the volume form on $\mathcal{S}_{\tilde{y}}$ induced by the embedding $\mathcal{S}_{\tilde{y}} \hookrightarrow \mathcal{U}$, and $G^{\mathcal{S}}$ is the Gram matrix

$$G_{jk}^{\mathcal{S}}(t, \tilde{y}) = \frac{\partial \tilde{\Phi}(t, \tilde{y})}{\partial t_j} \cdot \frac{\partial \tilde{\Phi}(t, \tilde{y})}{\partial t_k}, \quad 1 \leq j, k \leq N. \quad (2.2)$$

Therefore, more explicitly,

$$\mathcal{R}f(\tilde{y}) = \int_{\mathbb{R}^N} f(x)b(x, \tilde{y})(\det G^{\mathcal{S}}(t, \tilde{y}))^{1/2}dt, \quad x = \tilde{\Phi}(t, \tilde{y}). \quad (2.3)$$

We assume that f is compactly supported, $\text{supp}(f) \subset \mathcal{U}$, and f is sufficiently smooth, so that $\mathcal{R}f(\tilde{y})$ is a continuous function. Exact (i.e., continuous) reconstruction is computed by

$$\check{f}(x) = (\mathcal{R}^* \mathcal{B}g)(x) = \int_{\tilde{\mathcal{T}}_x} (\mathcal{B}g)(\tilde{y}) w(x, \tilde{y}) d\tilde{y}, \quad x \in \mathcal{U}, \quad g = \mathcal{R}f, \quad (2.4)$$

where $w \in C_0^\infty(\mathcal{U} \times \tilde{\mathcal{V}})$, $d\tilde{y}$ is the volume form on $\tilde{\mathcal{T}}_x := \{\tilde{y} \in \tilde{\mathcal{V}} : x \in \mathcal{S}_{\tilde{y}}\}$ (which is induced by the embedding $\tilde{\mathcal{T}}_x \hookrightarrow \tilde{\mathcal{V}}$, see the paragraph following (4.1) below), $\mathcal{R}^*$ is a weighted adjoint of $\mathcal{R}$, and $\mathcal{B}$ is a fairly arbitrary pseudo-differential operator (Ψ DO). Lemma 3.7 below asserts that $\tilde{\mathcal{T}}_x \subset \tilde{\mathcal{V}}$ is a smooth, embedded submanifold. The reconstruction formula in (2.4) is of the Filtered-Backprojection type. Application of $\mathcal{B}$ is the filtering step, and integration with respect to $\tilde{y}$ (i.e., the application of $\mathcal{R}^*$) is the backprojection step. By reconstruction here we mean any function (or, distribution) $\check{f}$ that is reconstructed from the data using (2.4). The reconstruction is intended to recover the visible wave-front set of f , but the strength of the singularities of $\check{f}$ and f need not match.

Let D be a data sampling matrix, $\det D = 1$. Discrete data $g(\tilde{y}^j)$ are given on the lattice

$$\tilde{y}^j = \epsilon D j, \quad j \in \mathbb{Z}^n. \quad (2.5)$$

Reconstruction from discrete data is given by the same formula (2.4), where we replace g with its interpolated version g_ϵ (see 4.6).

We assume that $\mathcal{S} := \text{singsupp}(f)$ is a smooth hypersurface. Pick some $x_0 \in \mathcal{S}$. This point is fixed throughout the paper. Our goal is to study the function reconstructed from discrete data in a neighborhood of x_0 . All our results are local, so we assume that $\mathcal{U}$ is a sufficiently small neighborhood of x_0 . Let $\tilde{y}_0 \in \tilde{\mathcal{V}}$ be such that $\mathcal{S}_{\tilde{y}_0}$ is tangent to $\mathcal{S}$ at x_0 . Only a small neighborhood of $\tilde{y}_0$ is relevant for the recovery of the singularity of f at x_0 . Hence we assume that $\tilde{\mathcal{V}}$ is a sufficiently small neighborhood of $\tilde{y}_0$.

3 Selecting Coordinates, Geometric Assumptions

Let $\Psi(x) = 0$ be an equation of $\mathcal{S}$, and $d\Psi(x) \neq 0$, $x \in \mathcal{U}$. Introduce the matrix

$$\tilde{M} := \begin{pmatrix} \tilde{\Phi}_t & \tilde{\Phi}_y \\ (\xi_0 \cdot \tilde{\Phi})_{tt} & (\xi_0 \cdot \tilde{\Phi})_{ty} \end{pmatrix}, \quad \xi_0 := d\Psi \in T_{x_0}^* \mathcal{U}, \quad (3.1)$$

which is the Jacobian matrix for the equations

$$\tilde{\Phi}(t, \tilde{y}) = x_0, \quad \xi_0 \cdot \tilde{\Phi}_t(t, \tilde{y}) = 0, \quad (3.2)$$

where $(t, \tilde{y}) \in \mathbb{R}^N \times \tilde{\mathcal{V}}$ are the unknowns. See Fig. 1 for an illustration of $\mathcal{S}$ and ξ_0 . For convenience, in (3.1) and in the rest of the paper we frequently drop the arguments of Ψ , $\tilde{\Phi}$, and similar functions whenever they are x_0 and $(t_0, \tilde{y}_0)$, as appropriate. Here t_0 is the unique point such that $x_0 = \tilde{\Phi}(t_0, \tilde{y}_0)$. Our convention is that a variable in the subscript of a function denotes the partial derivative of the function with respect to the variable, e.g.,

$$\tilde{\Phi}_{\tilde{y}_j} = \begin{pmatrix} \partial_{\tilde{y}_j} \tilde{\Phi}_1 \\ \vdots \\ \partial_{\tilde{y}_j} \tilde{\Phi}_n \end{pmatrix}, \quad \tilde{\Phi}_{\tilde{y}} = \begin{pmatrix} \partial_{\tilde{y}_1} \tilde{\Phi}_1 & \dots & \partial_{\tilde{y}_n} \tilde{\Phi}_1 \\ \vdots & \dots & \vdots \\ \partial_{\tilde{y}_1} \tilde{\Phi}_n & \dots & \partial_{\tilde{y}_n} \tilde{\Phi}_n \end{pmatrix}. \quad (3.3)$$

Assumption 3.1 (*Geometry of the GRT*)

- G1. $\text{rank } \tilde{\Phi}_t = N$;
 G2. $\det \tilde{M} \neq 0$.

Remark 3.2 Assumption G1 implies that $\mathcal{S}_{\tilde{y}} \subset \mathcal{U}$ is a smooth N -dimensional embedded submanifold for any $\tilde{y} \in \tilde{\mathcal{V}}$ provided that $\mathcal{U} \ni x_0$ and $\tilde{\mathcal{V}} \ni \tilde{y}_0$ are sufficiently small neighborhoods. Assumption G2 guarantees that any singularity of f microlocally near (x_0, ξ_0) is visible from the GRT data $\mathcal{R}f(\tilde{y})$, $\tilde{y} \in \tilde{\mathcal{V}}$. In addition, Assumption G2 ensures that $\tilde{\mathcal{T}}_x = \{\tilde{y} \in \tilde{\mathcal{V}} : x \in \mathcal{S}_{\tilde{y}}\}$ is a codimension $n - N$ embedded manifold for any $x \in \mathcal{U}$ (see (3.14) and Lemma 3.7).

Example 3.3 Excluding some exceptional cases, Assumptions 3.1 are satisfied by the X-ray transform in $\mathbb{R}^3$, where $\mathcal{S}_{\tilde{y}}$ are lines intersecting a smooth curve $\mathcal{C}$. Since $\mathcal{S}_{\tilde{y}}$ are lines, it is trivial that one can find $\tilde{\Phi}$ so that G1 holds. It is well-known that G2 holds for some $\tilde{\Phi}$ if the plane $\Pi_0 := \{x \in \mathcal{U} : \xi_0 \cdot (x - x_0) = 0\}$ intersects $\mathcal{C}$ transversely. By (3.1) and (3.2), this is, essentially, the Tuy condition [34]. Assumption 3.1(G2) fails to hold if Π_0 is either tangent to $\mathcal{C}$ or does not intersect it. In the latter case the singularity at (x_0, ξ_0) is invisible.

Definition 3.4 The pair $(x_0, \tilde{y}_0)$, such that $\mathcal{S}_{\tilde{y}_0}$ is tangent to $\mathcal{S}$ at x_0 , is generic for the sampling matrix D if

- (1) There is no vector $m \in \mathbb{Z}^n$, $m \neq 0$, such that the 1-form

$$\tilde{\omega} = (D^{-T}m)_1 d\tilde{y}_1 + \dots + (D^{-T}m)_n d\tilde{y}_n \in T_{\tilde{y}_0}^* \tilde{\mathcal{T}}_{x_0} \quad (3.4)$$

vanishes identically on $T_{\tilde{y}_0} \tilde{\mathcal{T}}_{x_0}$, i.e. $\tilde{\omega} \notin N_{\tilde{y}_0}^* \tilde{\mathcal{T}}_{x_0}$, and

- (2) The matrix $(\Psi \circ \tilde{\Phi})_{tt}$ is either positive definite or negative definite.

In simple terms, condition (1) says that there is no nonzero vector $m \in \mathbb{Z}^n$ such that $D^{-T}m$ is orthogonal to $\tilde{\mathcal{T}}_{x_0}$ at $\tilde{y}_0$. For more information about this condition see Remark 5.1.

In the rest of the paper, we assume that the pair $(x_0, \tilde{y}_0)$ is generic in the sense of Definition 3.4, and $(\Psi \circ \tilde{\Phi})_{tt}$ is *negative* definite. The latter assumption is not restrictive, because the positive and negative definite cases can be converted into each other by a change of the x coordinates. To illustrate the notation convention described above, $(\Psi \circ \tilde{\Phi})_{tt}$ stands for the matrix of the second derivatives of the function $(\Psi \circ \tilde{\Phi})(t, y)$ with respect to t evaluated at $(t_0, \tilde{y}_0)$.

Using Assumption 3.1(G1), select x coordinates so that

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}, \quad x^{(1)} \in \mathbb{R}^{n-N}, \quad x^{(2)} \in \mathbb{R}^N, \quad x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \tilde{\Phi}(t_0, \tilde{y}_0), \quad (3.5)$$

$$d\Psi = (|d\Psi|, 0, \dots, 0), \quad \tilde{\Phi}_t^{(1)} = 0, \quad \det \tilde{\Phi}_t^{(2)} \neq 0.$$

We also denote $x^\perp := (x_2, \dots, x_n)^T$. See Fig. 1 for an illustration of the coordinates x_1 and $x^\perp$ (the plane $\{x : x_1 = 0, x^\perp \in \mathbb{R}^{n-1}\}$ is shown as a shaded oval). The notation $\tilde{\Phi}_*^{(j)}$, $j = 1, 2$, stands for the derivative of the j -th group of coordinates of $x = \tilde{\Phi}(t, \tilde{y})$ (either along $x^{(1)}$ or along $x^{(2)}$) with respect to $*$ = t or $\tilde{y}$. By the last inequality in (3.5), we can select $x^{(2)}$ as the t variables. With this choice we have (with some other defining function $\tilde{\Phi}$):

$$x^{(1)} = \tilde{\Phi}^{(1)}(x^{(2)}, \tilde{y}), \quad x^{(2)} \equiv \tilde{\Phi}^{(2)}(x^{(2)}, \tilde{y}), \quad \tilde{\Phi}_{x^{(2)}}^{(1)} = 0, \quad \tilde{\Phi}_{x^{(2)}}^{(2)}(x^{(2)}, \tilde{y}) \equiv I_N, \quad (3.6)$$

where I_N is the $N \times N$ identity matrix. This definition of $\tilde{\Phi}$ is assumed in what follows.

We also need a convenient y coordinate system. Since the data points in (2.5) are given in the original coordinates, we have to keep track of both the original and new coordinates. Points in the original and new y coordinates are denoted $\tilde{y}$ and y , respectively. Data domains in the original and new y coordinates are denoted $\tilde{\mathcal{V}}$ and $\mathcal{V}$, respectively. Suppose that $\tilde{y} = Uy + \tilde{y}_0$ and $\tilde{\mathcal{V}} = U\mathcal{V} + \tilde{y}_0$, where U is some orthogonal matrix $U : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

In what follows we will be using mostly the new y coordinates, so we further modify the defining function:

$$\Phi(x^{(2)}, y) := \tilde{\Phi}(x^{(2)}, \tilde{y}(y)) = \tilde{\Phi}(x^{(2)}, Uy + \tilde{y}_0). \quad (3.7)$$

Since the x variable remains the same, the function $\Phi(x^{(2)}, y)$ satisfies (3.6) (with the derivative computed at $(x_0^{(2)}, y_0) = (0, 0)$):

$$x^{(1)} = \Phi^{(1)}(x^{(2)}, y), \quad x^{(2)} \equiv \Phi^{(2)}(x^{(2)}, y), \quad \Phi_{x^{(2)}}^{(1)} = 0, \quad \Phi_{x^{(2)}}^{(2)}(x^{(2)}, y) \equiv I_N. \quad (3.8)$$

For the same reason, condition (2) in Definition 3.4 implies that the matrix $(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}$ is either positive definite or negative definite. The following lemma is proven in Appendix A.

Lemma 3.5 Suppose $x_0 \in \mathcal{S}$, $\mathcal{S}_{\tilde{y}_0}$ is tangent to $\mathcal{S}$ at x_0 , $\det(\Psi \circ \tilde{\Phi})_{tt} \neq 0$, and Assumptions 3.1 hold. The orthogonal matrix U and the function Ψ , which satisfies (3.5), can be selected so that the new y coordinates and the new function Φ satisfy

$$y = \begin{pmatrix} y_1 \\ y^\perp \end{pmatrix}, \quad y_1 \in \mathbb{R}, \quad y^\perp \in \mathbb{R}^{n-1}, \quad y_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (\Psi \circ \Phi)_{y_1} = 1, \quad (\Psi \circ \Phi)_{y^\perp} = 0; \quad (3.9)$$

and

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}, \quad y^{(1)} \in \mathbb{R}^{n-N}, \quad y^{(2)} \in \mathbb{R}^N, \quad \det \Phi_{y^{(1)}}^{(1)} \neq 0, \quad \Phi_{y^{(2)}}^{(1)} = 0. \quad (3.10)$$

Remark 3.6 The representations $y = (y_1, y^\perp)^T$ and $y = (y^{(1)}, y^{(2)})^T$ are two different ways to split the y coordinates, which are convenient in different contexts. See Fig. 1

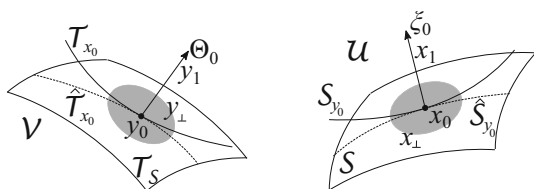


Fig. 1 Main geometric objects used in the paper

for an illustration of the coordinates y_1 and $y^\perp$ (the plane $\{y : y_1 = 0, y^\perp \in \mathbb{R}^{n-1}\}$ is shown as a shaded oval).

Similarly to (3.1), we compute the matrix M by replacing $\tilde{y}$ and $\tilde{\Phi}$ with y and Φ , respectively. In block form

$$\begin{aligned} M &:= \begin{pmatrix} \Phi_{x^{(2)}} & \Phi_y \\ \xi_0 \cdot \Phi_{x^{(2)}x^{(2)}} & \xi_0 \cdot \Phi_{x^{(2)}y} \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \\ M_{11} &= \Phi_{(x^{(2)}, y^{(1)})} \in \mathbb{R}^{n \times n}, \quad M_{12} = \Phi_{y^{(2)}} \in \mathbb{R}^{n \times N}, \\ M_{21} &\in \mathbb{R}^{N \times n}, \quad M_{22} = \xi_0 \cdot \Phi_{x^{(2)}y^{(2)}} \in \mathbb{R}^{N \times N}. \end{aligned} \quad (3.11)$$

In the selected x - and y -coordinates (see (3.5), (3.10)), the matrix M becomes

$$M = \begin{pmatrix} 0 & \Phi_{y^{(1)}}^{(1)} & 0 \\ I_N & 0 & 0 \\ \xi_0 \cdot \Phi_{x^{(2)}x^{(2)}} & \xi_0 \cdot \Phi_{x^{(2)}y^{(1)}} & \xi_0 \cdot \Phi_{x^{(2)}y^{(2)}} \end{pmatrix}. \quad (3.12)$$

By (3.12), Assumption 3.1(G2) is equivalent to

$$\det \Phi_{y^{(1)}}^{(1)} \neq 0, \quad \det M_{22} = \det (\xi_0 \cdot \Phi_{x^{(2)}y^{(2)}}) \neq 0. \quad (3.13)$$

Let us introduce two important sets:

$$\mathcal{T}_S := \{y \in \mathcal{V} : \mathcal{S}_y \text{ is tangent to } \mathcal{S}\}, \quad \mathcal{T}_x := \{y \in \mathcal{V} : x \in \mathcal{S}_y\}. \quad (3.14)$$

Here and in what follows, with some mild abuse of notation, $\mathcal{S}_y$ denotes $\mathcal{S}_{\tilde{y}(y)}$. See Fig. 1 for an illustration of $\mathcal{S}_{y_0}$, $\mathcal{T}_S$, and $\mathcal{T}_{x_0}$.

Lemma 3.7 Suppose $x_0 \in \mathcal{S}$, $\mathcal{S}_{y_0}$ is tangent to $\mathcal{S}$ at x_0 , $\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}} \neq 0$, and Assumptions 3.1 hold. One can find sufficiently small neighborhoods $\mathcal{U} \ni x_0$ and $\mathcal{V} \ni y_0$ so that

- (1) $\mathcal{T}_S \subset \mathcal{V}$ is a smooth, codimension one embedded manifold; and
- (2) $\mathcal{T}_x \subset \mathcal{V}$ is a smooth, codimension $n - N$ embedded manifold for any $x \in \mathcal{U}$.

Proof To find $\mathcal{T}_S$, we solve the equations

$$\Psi(\Phi(x^{(2)}, y)) = 0, \quad (\Psi \circ \Phi)_{x^{(2)}}(x^{(2)}, y) = 0 \quad (3.15)$$

for $x^{(2)}$ and y_1 in terms of $y^\perp$. The Jacobian matrix is

$$\begin{pmatrix} (\Psi \circ \Phi)_{x^{(2)}} & (\Psi \circ \Phi)_{y_1} \\ (\Psi \circ \Phi)_{x^{(2)}x^{(2)}} & (\Psi \circ \Phi)_{x^{(2)}y_1} \end{pmatrix}. \quad (3.16)$$

By condition (2) in Definition 3.4 (with $t = x^{(2)}$), $\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}} \neq 0$. Moreover, $(\Psi \circ \Phi)_{x^{(2)}} = 0$ and $(\Psi \circ \Phi)_{y_1} \neq 0$, and the Jacobian is non-degenerate. Therefore, solving (3.15) determines $x^{(2)}(y^\perp)$ and $y_1(y^\perp)$ as smooth functions of $y^\perp$ in a small neighborhood of $y^\perp = 0$. In particular, $y_1(y^\perp)$ is a local equation of the smooth, codimension 1 embedded submanifold $\mathcal{T}_S \subset \mathcal{V}$. The point of tangency $x_* = \Phi(x^{(2)}(y^\perp), (y_1(y^\perp), y^\perp))$ also depends smoothly on $y^\perp$.

To prove assertion (2), solve $x^{(1)} = \Phi^{(1)}(x^{(2)}, y)$ for $y^{(1)}$. This gives an equation for $\mathcal{T}_x$ in the form $y = Y(y^{(2)}, x)$ (where $y^{(2)} \equiv Y^{(2)}(y^{(2)}, x)$). The property $\det \Phi_{y^{(1)}}^{(1)} \neq 0$ (cf. (3.10)) implies that $\mathcal{T}_x \subset \mathcal{V}$ is a smooth, codimension $n - N$ embedded submanifold for any $x \in \mathcal{U}$ provided that both $\mathcal{U}$ and $\mathcal{V}$ are sufficiently small. Since $\Phi_{y^{(2)}}^{(1)} = 0$, we also get

$$\partial Y^{(1)}(y^{(2)}, x_0) / \partial y^{(2)} \Big|_{y^{(2)}=0} = 0. \quad (3.17)$$

□

In what follows, we use

$$\Theta_0 := d_y(\Psi \circ \Phi) = \Phi^* \xi_0; \quad Y_0(y^{(2)}) := Y(y^{(2)}, x_0), \quad y^{(2)} \in \mathcal{V}. \quad (3.18)$$

Thus $\Theta_0 \in T_{y_0}^* \mathcal{V}$ is the pull-back of $\xi_0 \in T_{x_0}^* \mathcal{U}$ by $\Phi(0, \cdot)$. By (3.9), $\Theta_0 = dy_1$ (see Fig. 1).

4 Main Assumptions and Main Result

To simplify notations, in the rest of the paper we set $b(x, y) := b(x, \tilde{y}(y))$, $w(x, y) := w(x, \tilde{y}(y))$, and $g(y) := g(\tilde{y}(y))$. The original versions of these functions are used only in Sect. 2. Thus the reconstruction is computed by

$$(\mathcal{R}^* \mathcal{B}g)(x) = \int_{\mathcal{T}_x} (\mathcal{B}g)(y) w(x, y) dy, \quad (4.1)$$

where $dy = (\det G^T(y^{(2)}, x))^{1/2} dy^{(2)}$ is the volume form on $\mathcal{T}_x$, G^T is the Gram matrix:

$$G_{ij}^T(y^{(2)}, x) = \frac{\partial Y(y^{(2)}, x)}{\partial y_i^{(2)}} \cdot \frac{\partial Y(y^{(2)}, x)}{\partial y_j^{(2)}}, \quad 1 \leq i, j \leq N, \quad (4.2)$$

$\mathcal{B}$ is a Ψ DO

$$(\mathcal{B}g)(y) := \frac{1}{(2\pi)^n} \int \tilde{B}(y, \eta) \tilde{g}(\eta) e^{-iy \cdot \eta} d\eta, \quad \tilde{g} = \mathcal{F}g, \quad (4.3)$$

and $\mathcal{F}$ is the Fourier transform in $\mathbb{R}^n$. To clarify the use of indices in (4.2), if $y^{(2)}$ is viewed as part of y , then $y_j^{(2)} = y_{n-N+j}$, $1 \leq j \leq N$. Using (2.5) and that $\tilde{y} = Uy + \tilde{y}_0$, the discrete data $g(\hat{y}^j)$ are known at the points

$$\hat{y}^j = U^T(\epsilon Dj - \tilde{y}_0), \quad j \in \mathbb{Z}^n. \quad (4.4)$$

Reconstruction from discrete data is given by

$$\check{f}_\epsilon(x) = (\mathcal{R}^* \mathcal{B}g_\epsilon)(x) = \int_{T_x} (\mathcal{B}g_\epsilon)(y) w(x, y) dy, \quad (4.5)$$

where $g_\epsilon(y)$ is the interpolated data:

$$g_\epsilon(y) := \sum_{|j| \leq \vartheta/\epsilon} \varphi\left(\frac{y - \hat{y}^j}{\epsilon}\right) g(\hat{y}^j), \quad (4.6)$$

φ is an interpolation kernel, $\vartheta = \sigma_{\min}^{-1} \sup_{\tilde{y} \in \tilde{\mathcal{V}}} |\tilde{y}|$, and $\sigma_{\min}$ is the smallest singular value of the sampling matrix D . The value of ϑ is selected in such a way that $\hat{y}^j \in \text{supp}(g)$ implies $|j| \leq \vartheta/\epsilon$. In what follows we call (4.1) *reconstruction from continuous data* (as opposed to reconstruction from discrete data (4.5)). Denote $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$.

Definition 4.1 Given an open set $V \subset \mathbb{R}^n$, $r \in \mathbb{R}$, and $N \in \mathbb{N}$, $S^r(V \times \mathbb{R}^N)$ denotes the space of $C^\infty(V \times (\mathbb{R}^N \setminus \{0\}))$ functions, $\tilde{B}(y, \eta)$, having the following properties

$$\begin{aligned} |\partial_y^m \tilde{B}(y, \eta)| &\leq c_m |\eta|^{a-N} \quad \forall m \in \mathbb{N}_0^n, y \in V, 0 < |\eta| \leq 1; \\ |\partial_y^{m_1} \partial_\eta^{m_2} \tilde{B}(y, \eta)| &\leq c_{m_1, m_2} |\eta|^{r-|m_2|}, \quad \forall m_1, m_2 \in \mathbb{N}_0^n, y \in V, |\eta| \geq 1; \end{aligned} \quad (4.7)$$

for some constants $c_m, c_{m_1, m_2} > 0$ and $a > 0$.

Now we state all the assumptions about $\mathcal{B}$, φ , and g .

Assumption 4.2 (*Properties of $\mathcal{B}$*)

- B1. $\tilde{B}(y, \eta) \equiv 0$ outside a small conic neighborhood of (y_0, Θ_0) ; and
 B2. The amplitude of $\mathcal{B}$ satisfies

$$\begin{aligned} \tilde{B} &\in S^{\beta_0}(\mathcal{V} \times \mathbb{R}^n); \quad \tilde{B} - \tilde{B}_0 \in S^{\beta_1}(\mathcal{V} \times \mathbb{R}^n); \\ \tilde{B}_0(y, \lambda \eta) &= \lambda^{\beta_0} \tilde{B}_0(y, \eta); \quad \beta_0 > \beta_1; \end{aligned} \quad (4.8)$$

for some $\tilde{B}_0$, β_0 , and β_1 , and for all $y \in \mathcal{V}$, $\lambda > 0$.

Recall that $\lfloor r \rfloor$, $r \in \mathbb{R}$, denotes the largest integer not exceeding r . Similarly, $\lceil r \rceil$ denotes the smallest integer greater than or equal to r . We also introduce:

$$\lfloor r^- \rfloor = \lim_{\epsilon \rightarrow +0} \lfloor r - \epsilon \rfloor = \begin{cases} \lfloor r \rfloor, & r \notin \mathbb{Z}, \\ r - 1, & r \in \mathbb{Z}, \end{cases} \quad \lceil r^+ \rceil = \lim_{\epsilon \rightarrow +0} \lceil r + \epsilon \rceil = \begin{cases} \lceil r \rceil, & r \notin \mathbb{Z}, \\ r + 1, & r \in \mathbb{Z}. \end{cases} \quad (4.9)$$

Assumption 4.3 (*Properties of the interpolation kernel φ*)

IK1. $\varphi \in C_0^{\lceil \beta_0^+ \rceil}(\mathbb{R}^n)$, i.e., φ is compactly supported, and all of its derivatives up to order $\lceil \beta_0^+ \rceil$ are L^∞ ;

IK2. φ is exact up to order $\lceil \beta_0 \rceil$ for the sampling lattice determined by $D_1 := U^T D$, i.e.,

$$\sum_{j \in \mathbb{Z}^n} (D_1 j)^m \varphi(u - D_1 j) \equiv u^m, \quad |m| \leq \lceil \beta_0 \rceil, \quad m \in \mathbb{N}_0^n, \quad u \in \mathbb{R}^n, \quad (4.10)$$

for all indicated m and u .

Assumption IK2 with $m = 0$ implies that φ is normalized. By assumption, $|\det D_1| = 1$. Then

$$\begin{aligned} 1 &= \int_{D_1[0,1]^n} \sum_{j \in \mathbb{Z}^n} \varphi(u - D_1 j) du = \int_{[0,1]^n} \sum_{j \in \mathbb{Z}^n} \varphi(D_1(v - j)) dv \\ &= \int_{\mathbb{R}^n} \varphi(D_1 v) dv = \int_{\mathbb{R}^n} \varphi(v) dv. \end{aligned} \quad (4.11)$$

Assume that g is given by

$$g(y) = \frac{1}{2\pi} \int_{\mathbb{R}} \tilde{v}(y, \lambda) e^{-i\lambda P(y)} d\lambda, \quad y \in \mathcal{V}, \quad (4.12)$$

for some $P \in C^\infty(\mathcal{V})$ with $dP(y) \neq 0$ on $\mathcal{V}$. We need the smooth hypersurface determined by P :

$$\Gamma := \{y \in \mathcal{V} : P(y) = 0\}. \quad (4.13)$$

Even if g is not in the range of $\mathcal{R}$, Assumptions 3.1 and Definition 3.4 imply that there exists a smooth surface $\mathcal{S} \subset \mathcal{U}$ such that $\Gamma = \mathcal{T}_{\mathcal{S}}$ (see Sect. 5.1 for additional information). Let $\hat{\mathcal{S}}_{y_0}$ be the projection of $\mathcal{S}_{y_0}$ onto $\mathcal{S}$ along the first coordinate x_1 (see Fig. 1).

Definition 4.4 Set

$$\Delta \Pi_{\mathcal{S}} := \Pi_{\mathcal{S}_{y_0}} - \Pi_{\hat{\mathcal{S}}_{y_0}}, \quad (4.14)$$

where $\Pi_{\mathcal{S}_{y_0}}$ is the matrix of the second fundamental form of $\mathcal{S}_{y_0}$ at x_0 written in the coordinates $(x_1, x^{(2)})^T$, and similarly for $\hat{\mathcal{S}}_{y_0}$.

Assumption 4.5 (*Properties of the data function g*)

- g1. $\tilde{v} \in S^{-(s_0+1)}(\mathcal{V} \times \mathbb{R})$, and there exists a compact $K \subset \mathcal{V}$ such that $\tilde{v}(y, \lambda) \equiv 0$ if $y \in \mathcal{V} \setminus K$;
 g2. $\tilde{v}$ satisfies

$$\begin{aligned} \tilde{v}(y, \lambda) &= \tilde{v}^+(y) \lambda_+^{-(s_0+1)} + \tilde{v}^-(y) \lambda_-^{-(s_0+1)} + \tilde{R}(y, \lambda), \quad \forall y \in \mathcal{V}, |\lambda| \geq 1, \\ \tilde{v}^\pm &\in C_0^\infty(\mathcal{V}), \quad \tilde{R} \in S^{-(s_1+1)}(\mathcal{V} \times \mathbb{R}), \quad 0 < s_0 < s_1, \quad s_1 \notin \mathbb{N}; \end{aligned} \quad (4.15)$$

for some $\tilde{v}^\pm$, $\tilde{R}$, s_0 , and s_1 ;

- g3. $P \in C^\infty(\mathcal{V})$ is given by

$$P(y) = y_1 - \psi(y^\perp), \quad (4.16)$$

where ψ is smooth, $\psi(0) = 0$, and $d\psi(0) = 0$;

- g4. If $s_0 \in \mathbb{N}$, one has

$$\tilde{v}^+(y) = (-1)^{s_0+1} \tilde{v}^-(y) \quad \text{for any } y \in \Gamma. \quad (4.17)$$

- g5. The matrix $\Delta \Pi_S$ is negative definite.

We use superscripts $'\pm'$ to distinguish between two different functions as opposed to the positive and negative parts of a number. The latter are denoted by the subscripts $'\pm'$: $\lambda_\pm := \max(\pm\lambda, 0)$.

Assumptions g1, g2 imply that g is sufficiently regular. The assumption $s_1 \notin \mathbb{N}$ is not restrictive. It is made to simplify some of the proofs. Assumption g3 is not restrictive either. An equivalent assumption is $dP_1(y_0) = \Theta_0$ for some other smooth P_1 (see also (B.5) and (B.9) below). Indeed, by shrinking $\mathcal{V}$, if necessary, we can find $\psi(y^\perp)$ with the required properties such that the function $u(y) := P_1(y)/P(y)$ (where P is as in (4.16)) satisfies $u \in C^\infty(\mathcal{V})$ and $c_1 \leq |u(y)|, |du(y)| \leq c_2$ for some $c_{1,2} > 0$ and all $y \in \mathcal{V}$. Substituting $P_1(y) = u(y)P(y)$ into (4.12) and changing variables $\lambda_1 = \lambda u(y)$, we see that the new amplitude $\tilde{v}_1(y, \lambda_1) := \tilde{v}(y, \lambda_1/u(y))/u(y)$ satisfies g1, g2 (with the same s_0 and s_1), and g4. See Remark B.1 about the meaning of assumption g4. Define

$$e(a) := \exp\left(i \frac{\pi}{2} a\right). \quad (4.18)$$

Assumption 4.6 (*Joint properties of $\mathcal{B}$ and g*)

- C1. The constants β_0 and s_0 , defined in (4.8) and (4.15), respectively, satisfy

$$\kappa := \beta_0 - s_0 - (N/2) \geq 0; \quad (4.19)$$

- C2. The functions $\tilde{B}_0$ and $\tilde{v}^\pm$, defined in (4.8) and (4.15), respectively, satisfy

$$\tilde{B}_0(y, d_y P(y)) \tilde{v}^+(y) = -e(2(\beta_0 - s_0)) \tilde{B}_0(y, -d_y P(y)) \tilde{v}^-(y) \quad \text{if } \kappa = 0 \quad \forall y \in \Gamma. \quad (4.20)$$

The role of conditions (4.17) and (4.20) is that they prevent the appearance of logarithmic terms in g and $\check{f} = \mathcal{R}^* \mathcal{B}g$ in a neighborhood of Γ and $\mathcal{S}$, respectively, see (C.6) and (12.7).

Set $x_\epsilon := x_0 + \epsilon \check{x}$. Adopt the convention that the interior side of $\mathcal{S}$ is the one where the x_1 axis points and define

$$x_0^{\text{int}} := \lim_{x_1 \rightarrow 0^+} (x_1, x^\perp = 0), \quad x_0^{\text{ext}} := \lim_{x_1 \rightarrow 0^-} (x_1, x^\perp = 0). \quad (4.21)$$

Let $\hat{\varphi}$ denote the classical Radon transform of φ (see (8.8)). Recall that $D_1 = U^T D$. Now we can state our main result.

Theorem 4.7 *Suppose $\mathcal{U}$ and $\mathcal{V}$ are sufficiently small neighborhoods of x_0 and y_0 , respectively, and*

- (1) $\mathcal{S}_{y_0}$ is tangent to $\mathcal{S}$ at x_0 ;
- (2) Assumptions 3.1, 4.2, 4.3, 4.5, and 4.6 are satisfied; and
- (3) The pair (x_0, y_0) is generic for the sampling matrix D_1 .

Then one has

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \epsilon^\kappa \check{f}_\epsilon(x_\epsilon) &= C_1 \int \hat{\varphi}(\Theta_0, (\check{x}_1/\partial_{y_1} \Phi_1) - p) \left(c_1^+(p - i0)^{-\kappa} + c_1^-(p + i0)^{-\kappa} \right) dp, \\ C_1 &= (2\pi)^{N/2} w(x_0, y_0) |\det \Delta H_S|^{1/2} (\partial_{y_1} \Phi_1)^{-N/2} \left| \det \frac{\partial^2 \Phi_1}{\partial x^{(2)} \partial y^{(2)}} \right|^{-1}, \quad (4.22) \\ c_1^\pm &= \frac{\Gamma(\kappa)}{2\pi} \tilde{B}_0(y_0, \pm \Theta_0) \tilde{v}^\pm(y_0) e(\mp(\beta_0 - s_0)) \quad \text{if } \kappa > 0, \end{aligned}$$

and

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \check{f}_\epsilon(x_\epsilon) &= \check{f}(x_0^{\text{int}}) + C_1 c_1 \int_{-\infty}^0 \hat{\varphi}(\Theta_0, (\check{x}_1/\partial_{y_1} \Phi_1) - p) dp, \quad (4.23) \\ c_1 &:= i \tilde{B}_0(y_0, \Theta_0) \tilde{v}^+(y_0) e(-(\beta_0 - s_0)) \quad \text{if } \kappa = 0. \end{aligned}$$

See [9, Chapter I, Sect. 3.6] for the definition and properties of the distributions $(p \pm i0)^a$.

Remark 4.8 Since $\check{x}_1$ is a rescaled coordinate, (4.22) and (4.23) imply that in the original coordinates x the resolution of reconstruction is $\sim \epsilon r(\partial_{y_1} \Phi_1)(x_0^{(2)}, y_0)$, where $r > 0$ is (loosely speaking) a measure of the spread of $\hat{\varphi}(\Theta_0, p)$ as a function of p . This shows that the resolution is not only location-dependent (via the $x^{(2)}$ dependence of Φ), but also direction-dependent (via the y dependence of Φ).

Even though the formulas (4.22), (4.23) do not contain the sampling matrix, the dependence on D_1 is still there. It is implicit, and manifests itself via the kernel φ , which is required to be exact on the lattice determined by D_1 (see Assumption 4.3(IK2)).

Remark 4.9 The limits in (4.22) and (4.23) are functions of the scalar argument $h = \check{x}_1/\partial_{y_1} \Phi_1$. Hence, it is more appropriate to view the DTBs as functions on $\mathbb{R}$ rather

than on $\mathbb{R}^n$. With this convention, the expressions in (4.22) and (4.23) can be written as $\text{DTB}(\check{x}_1/\partial_{y_1}\Phi_1)$. The same convention applies to the CTB as well. This convention will be used in the rest of the paper.

Remark 4.10 Set $x = x_0 + (h/|\xi_0|)\xi_0$, i.e. h is physical (not rescaled) signed distance from x to x_0 . Equation (4.23) implies that the derivative of the *edge response function* of the reconstruction (in $\mathbb{R}^2$, this derivative is known as the line spread function) is $E'(h) := \hat{\varphi}(\Theta_0, h/(\epsilon\partial_{y_1}\Phi_1))$. Thus, to compute, for example, FWHM, we perform the following steps: (1) Find the maximum $M := \max_h E'(h)$. Frequently, $M = \hat{\varphi}(\Theta_0, 0)$; and (2) Find the length $\text{FWHM} = |\{h \in \mathbb{R} : E'(h) \geq M/2\}|$.

The proof of the theorem is broken into several sections. The contact between $\mathcal{T}_S$ and $\mathcal{T}_{x_0}$ in a neighborhood of y_0 is investigated in Sect. 6. Properties of the continuous data function g and its interpolated version are investigated in Sect. 7. These two sections prepare the groundwork for the remainder of the proof in Sects. 8–11. A high-level overview of the remainder of the proof is at the end of Sect. 8.1.

5 Additional Results: Discussion

5.1 FIO Point of View

Introduce the function

$$\phi(x, y, \lambda) = \lambda \cdot (x^{(1)} - \Phi^{(1)}(x^{(2)}, y)), \quad \lambda \in \mathbb{R}^{n-N}. \quad (5.1)$$

Clearly, ϕ is the phase function of the FIO $\mathcal{R}$ (cf. (2.3)):

$$\mathcal{R}f(y) = \frac{1}{(2\pi)^{n-N}} \int_{\mathbb{R}^{n-N}} \int_{\mathbb{R}^n} f(x) e^{i\phi(x, y, \lambda)} b(x, y) (\det G^S(x^{(2)}, y))^{1/2} dx d\lambda \quad (5.2)$$

with the canonical relation from $T^*\mathcal{U}$ to $T^*\mathcal{V}$:

$$C := \{(x, d_x\phi(x, y, \lambda), y, -d_y\phi(x, y, \lambda) : x^{(1)} = \Phi^{(1)}(x^{(2)}, y)\} \subset T^*\mathcal{U} \times T^*\mathcal{V}. \quad (5.3)$$

In (5.2), G^S is computed similarly to (2.2), but with $\tilde{y}$ and $\tilde{\Phi}$ replaced by y and Φ , respectively.

Set $\lambda_0 := (1, 0, \dots, 0)^T$. Clearly, $\xi_0 = |d\Psi|d_x\phi \neq 0$ and $\Theta_0 = -|d\Psi|d_y\phi \neq 0$. This follows easily from (3.5), (3.18), and (5.1). This also implies that the differentials $d_{x,\lambda}\phi$ and $d_{y,\lambda}\phi$ do not vanish anywhere in a conic neighborhood of (x_0, y_0, λ_0) (see [33, Definition 2.1, Sect. VI.2]). As a reminder about our convention, the differentials are evaluated at (x_0, y_0, λ_0) . Clearly $\partial^2\phi/\partial x^{(1)}\partial\lambda = I_{n-N}$, so the differentials $d_{x,y,\lambda}(\partial\phi/\partial\lambda_j)$, $1 \leq j \leq n - N$, are linearly independent, and ϕ is nondegenerate [33, Definition 1.1, Sect. VIII.1].

Setting $\lambda = \lambda_0$ gives

$$\det \begin{pmatrix} \phi_{x,y} & \phi_{x,\lambda} \\ \phi_{\lambda,y} & \phi_{\lambda,\lambda} \end{pmatrix} = \det \begin{pmatrix} 0 & 0 & I_{n-N} \\ -\frac{\partial^2 \Phi_1}{\partial x^{(2)} \partial y^{(1)}} & -\frac{\partial^2 \Phi_1}{\partial x^{(2)} \partial y^{(2)}} & 0 \\ -\Phi_{y^{(1)}}^{(1)} & 0 & 0 \end{pmatrix}. \quad (5.4)$$

Here we have used that $\Phi_{x^{(2)}}^{(1)} = 0$ and $\Phi_{y^{(2)}}^{(1)} = 0$ (see (3.5) and (3.10)). In fact, the determinant in (5.4) is non-zero if and only if $\det M \neq 0$ (see (3.11)–(3.13)). Hence Assumption 3.1(G2) implies that C is a local canonical graph (see the discussion following Eq. (4.23) in [33, Sect. VI] and Definition 6.1 in [33, Sect. VIII]). In particular, there is a unique, smooth hypersurface $S \subset \mathcal{U}$ such that $(N^*\Gamma \setminus \mathbf{0}) = C \circ (N^*S \setminus \mathbf{0})$. Here $N^*\Gamma$ and N^*S are the conormal bundles of Γ and S , respectively, and $\mathbf{0}$ is the zero section.

From (2.2) and (3.6), $\det G^S = 1$. Clearly, $\mathcal{R}$ is an elliptic FIO in a neighborhood of $((x_0, \xi_0), (y_0, \Theta_0)) \in (T^*\mathcal{U} \setminus \mathbf{0}) \times (T^*\mathcal{V} \setminus \mathbf{0})$ if $b(x_0, y_0) \neq 0$ and $\mathcal{U}, \mathcal{V}$ are sufficiently small.

The fact that the GRT and its adjoint can be viewed as FIOs has been known for a long time (see e.g., [11, 26]). The material in this section is well-known, and is presented for the convenience of the reader to make the paper self-contained.

Remark 5.1 We can now discuss condition (1) in Definition 3.4 in more detail. Here it is convenient to argue in the original $\tilde{y}$ coordinates. The preceding discussion shows that for every $x \in S$ there is $\tilde{y}(x) \in \tilde{\Gamma}$, which depends smoothly on x , such that $S_{\tilde{y}(x)}$ is tangent to S and x . Consider the N -dimensional tangent space to $\tilde{T}_x$ at $\tilde{y}(x)$. By (3.6) it is determined by (1) solving $x^{(1)} = \tilde{\Phi}^{(1)}(x^{(2)}, (\tilde{y}^{(1)}, \tilde{y}^{(2)}))$ for $\tilde{y}^{(1)}$ in terms of x and $\tilde{y}^{(2)}$, and (2) computing the partial derivatives $\partial \tilde{y}^{(1)} / \partial \tilde{y}_j^{(2)}$, $1 \leq j \leq N$, at $(x, \tilde{y}^{(2)}(x))$. Consider the $n \times N$ matrix:

$$\Xi(x) := \begin{pmatrix} \partial \tilde{y}^{(1)} / \partial \tilde{y}^{(2)} \\ I_N \end{pmatrix}. \quad (5.5)$$

Condition (1) in Definition 3.4 is violated for an exceptional $x \in S$ if $(D^{-T}m)\Xi(x) = 0$ for some $m \in \mathbb{Z}^n$.

5.2 CTB and Its Relationship with DTB

When describing the leading singularity of a distribution at a point, the following definition (which is a slight modification of the one in [16]) is convenient.

Definition 5.2 [16] Given a distribution $f \in \mathcal{D}'(\mathbb{R}^n)$ and a point $x_0 \in \mathbb{R}^n$, suppose there exists a distribution $f_0 \in \mathcal{D}'(\mathbb{R}^n)$ so that for some $a \in \mathbb{R}$ the following equality

holds

$$\lim_{\epsilon \rightarrow 0} \epsilon^{-a} \int f(x_0 + \epsilon \check{x}) \partial_{\check{x}}^m \omega(\check{x}) d\check{x} = \int f_0(\check{x}) \partial_{\check{x}}^m \omega(\check{x}) d\check{x}, \quad (5.6)$$

$$\forall m \in \mathbb{N}_0^n, |m| = \max(0, \lceil a^+ \rceil),$$

for any $\omega \in C_0^\infty(\mathbb{R}^n)$. Then we call f_0 the leading order singularity of f at x_0 .

Definition 5.3 CTB is defined as the leading order singularity of the reconstruction from continuous data $\mathcal{R}^* \mathcal{B}g$ at x_0 .

Similarly to the DTB, the following theorem shows that the CTB can be viewed as a function of a scalar argument.

Theorem 5.4 Under the assumptions of Theorem 4.7, one has

$$\lim_{\epsilon \rightarrow 0} \epsilon^\kappa \mathcal{R}^* \mathcal{B}g(x_\epsilon) = CTB(\check{x}_1 / \partial_{y_1} \Phi_1), \quad (5.7)$$

where

$$CTB(h) = C_1 (c_1^+(h - i0)^{-\kappa} + c_1^-(h + i0)^{-\kappa}), \quad \kappa > 0, \quad (5.8)$$

$$CTB(h) = -C_1 c_1 (1/2) \operatorname{sgn}(h) + \text{const}, \quad \kappa = 0,$$

and C_1 , c_1 , and $c_1^\pm$ are the same as in Theorem 4.7. Thus, the DTB is the convolution of the CTB with the scaled classical Radon transform of the interpolating kernel:

$$DTB(h) = \int \hat{\varphi}(\Theta_0, h - p) CTB(p) dp, \quad \kappa > 0, \quad (5.9)$$

$$DTB(h) = \int \hat{\varphi}(\Theta_0, h - p) CTB(p) dp + \text{const}, \quad \kappa = 0.$$

5.3 Discussion

Recall that $\check{f}(x) := \mathcal{R}^* \mathcal{B}g(x)$ denotes the reconstruction from continuous data. Suppose $\kappa = 0$, i.e. $\check{f}$ has a jump across $\mathcal{S}$. The second term on the right in (4.23) equals zero for all $\check{x}_1 > c$. Because φ is normalized: $\int \hat{\varphi}(\Theta_0, p) dp = 1$, the second term equals $C_1 c_1$ for all $\check{x}_1 < -c$. Here $c > 0$ is sufficiently large, and we used that $\partial_{y_1} \Phi_1 > 0$ (cf. (3.9)). By (12.7), the product $C_1 c_1$ is precisely the jump of $\check{f}(x)$ across $\mathcal{S}$ at x_0 : $-C_1 c_1 = \check{f}(x_0^{\text{int}}) - \check{f}(x_0^{\text{ext}})$, see (10.2) and (4.21). Thus, the right-hand side of (4.23) equals to $\check{f}(x_0^{\text{int}})$ if $\check{x}_1 > c$, and to $\check{f}(x_0^{\text{ext}}) -$ if $\check{x}_1 < -c$. This shows that (4.23) describes a smooth transition of the discrete reconstruction $\check{f}_\epsilon(x_\epsilon)$ from the value $\check{f}(x_0^{\text{int}})$ on the interior side of $\mathcal{S}$ to the value $\check{f}(x_0^{\text{ext}})$ on the exterior side of $\mathcal{S}$. Loosely speaking, the transition happens over a region of size $O(\epsilon)$:

$$DTB(\check{x}_1 / \partial_{y_1} \Phi_1) = \lim_{\epsilon \rightarrow 0} \check{f}_\epsilon(x_0 + \epsilon \check{x}) = \begin{cases} \check{f}(x_0^{\text{int}}), & \check{x}_1 > c, \\ \check{f}(x_0^{\text{ext}}), & \check{x}_1 < -c. \end{cases} \quad (5.10)$$

Then the DTB is a “stretched” version of the abrupt jump of $\check{f}$ across $\mathcal{S}$ in the continuous case. This is most apparent from the last equations in (5.8), (5.9). See also Section 6 of [20] for a similar discussion in the setting of quasi-exact inversion of the GRT in $\mathbb{R}^3$.

6 Beginning of the Proof of Theorem 4.7. Tangency of $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{T}_{x_0}$

In this section we show that $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{T}_{x_0}$ are tangent at y_0 , and investigate their properties near the point of tangency.

Lemma 6.1 *Suppose the assumptions of Lemma 3.7 are satisfied. The submanifolds $\mathcal{T}_{\mathcal{S}}$ and $\mathcal{T}_{x_0}$ are tangent at $y_0 = 0$, and $\Theta_0 = d_y(\Psi \circ \Phi)$ is conormal to both of them at $y_0 = 0$.*

Proof Begin with $\mathcal{T}_{\mathcal{S}}$. Following the proof of the first assertion of Lemma 3.7, all we need to do is compute $\partial_{y^\perp} y_1$. Viewing $x^{(2)}$ and y_1 as functions of $y^\perp$ and differentiating the equation $(\Psi \circ \Phi)(x^{(2)}, y) = 0$ (cf. (3.15)) gives:

$$d\Psi \cdot (\Phi_{x^{(2)}} \partial_{y^\perp} x^{(2)} + \Phi_{y_1} \partial_{y^\perp} y_1 + \Phi_{y^\perp}) = (\Psi \circ \Phi)_{y_1} \partial_{y^\perp} y_1 = 0, \quad (6.1)$$

which implies $\partial_{y^\perp} y_1 = 0$. Here we have used that (cf. (3.5), (3.9))

$$(\Psi \circ \Phi)_{x^{(2)}} = 0, \quad (\Psi \circ \Phi)_{y_1} \neq 0, \quad (\Psi \circ \Phi)_{y^\perp} = 0. \quad (6.2)$$

Therefore, in the selected y coordinates, the equation of the tangent space $T_{y_0} \mathcal{T}_{\mathcal{S}}$ (viewed as a subspace of $\mathbb{R}^n$) is $y_1 = 0$. See the shaded ellipse on the left in Fig. 1. Since only the first component of Θ_0 is not zero (see the line following (3.18)), it follows that Θ_0 is conormal to $\mathcal{T}_{\mathcal{S}}$ at y_0 .

Consider next $\mathcal{T}_{x_0}$. Differentiating $x_0^{(1)} = \Phi^{(1)}(x_0^{(2)}, y)$, where $y^{(1)} = Y_0^{(1)}(y^{(2)})$ (see (3.18)), and using (3.10) gives $\partial y^{(1)} / \partial y^{(2)} = 0$. Hence the equation of the tangent space $T_{y_0} \mathcal{T}_{x_0}$ is $y^{(1)} = 0$, i.e. $T_{y_0} \mathcal{T}_{x_0}$ is a subspace of $T_{y_0} \mathcal{T}_{\mathcal{S}}$. $\square$

Next we look more closely at the contact between $\mathcal{T}_{x_0}$ and $\mathcal{T}_{\mathcal{S}}$.

Lemma 6.2 *Suppose the assumptions of Lemma 3.7 are satisfied. Let $y = Y_0(y^{(2)})$ be the equation of $\mathcal{T}_{x_0}$ defined in (3.18). Let $y = Z(y^{(2)})$ be the equation of the projection of $\mathcal{T}_{x_0}$ onto $\mathcal{T}_{\mathcal{S}}$ along the first coordinate, i.e., $Z(y^{(2)}) \in \mathcal{T}_{\mathcal{S}}$ and $Y_0(y^{(2)}) - Z(y^{(2)}) = (h(y^{(2)}), 0, \dots, 0)^T$ for a scalar function $h(y^{(2)})$. Then, with M_{22} as in (3.11), and Θ_0 as in (3.18), we have*

$$\begin{aligned} \Theta_0 \cdot (Y_0(y^{(2)}) - Z(y^{(2)})) &= -\frac{1}{2} Q y^{(2)} \cdot y^{(2)} + O(|y^{(2)}|^3), \\ Q &= M_{22}^T (\Psi \circ \Phi)_{x^{(2)} x^{(2)}}^{-1} M_{22}, \quad \det Q \neq 0. \end{aligned} \quad (6.3)$$

Proof We solve separately two sets of equations (recall that $x_0 = 0$):

$$\begin{aligned}\mathcal{T}_{x_0} : \Phi^{(1)}(0, y) &= 0; \\ \mathcal{T}_S : (\Psi \circ \Phi)(x^{(2)}, y) &= 0, (\Psi \circ \Phi)_{x^{(2)}}(x^{(2)}, y) = 0.\end{aligned}\quad (6.4)$$

Since $(x^{(2)}, y) = (0, 0)$ solves (6.4), we have to first order in $y^{(2)}$

$$\mathcal{T}_{x_0} : \Phi_y^{(1)} y = O(|y^{(2)}|^2); \quad (6.5)$$

$$\mathcal{T}_S : (\Psi \circ \Phi)_{x^{(2)}\check{x}^{(2)}} + (\Psi \circ \Phi)_y \check{y} = O(|y^{(2)}|^2), \quad (6.6)$$

$$(\Psi \circ \Phi)_{x^{(2)}x^{(2)}\check{x}^{(2)}} + (\Psi \circ \Phi)_{x^{(2)}y} \check{y} = O(|y^{(2)}|^2). \quad (6.7)$$

The solution to the second system (i.e., related to $\mathcal{T}_S$) is denoted with a check. Because $\mathcal{T}_{x_0}$ and $\mathcal{T}_S$ are tangent at y_0 , $y = \check{y} + O(|y^{(2)}|^2)$. Recall that we search not for the general solution $y \in \mathcal{T}_S$, but for the points $y = Z(y^{(2)})$ obtained by projecting $\mathcal{T}_{x_0}$ onto $\mathcal{T}_S$ along y_1 . Therefore, $\check{y}^{(2)} \equiv y^{(2)}$. By (3.10) and (6.5), $y^{(1)} = O(|y^{(2)}|^2)$. By (6.7) and the assumption $\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}} \neq 0$, $x^{(2)} = O(|y^{(2)}|)$.

Since $\Phi_y^{(2)} \equiv 0$, (6.5) implies $\Phi_y y, \Phi_y \check{y} = O(|y^{(2)}|^2)$. Let $\Delta x^{(2)}$ and $\Delta y = \begin{pmatrix} \Delta y^{(1)} \\ 0 \end{pmatrix}$ denote second order perturbations, i.e. $y^{(1)} = \Delta y^{(1)} + O(|y^{(2)}|^3)$ (and analogously for $x^{(2)}$ and $\check{y}^{(1)}$). Since $y^{(2)}$ is an independent variable, its perturbation is not considered. Then

$$\mathcal{T}_{x_0} : \Phi_{y^{(1)}}^{(1)} \Delta y^{(1)} + (1/2) \Phi_{y^{(2)}y^{(2)}}^{(1)} y^{(2)} \cdot y^{(2)} = O(|y^{(2)}|^3); \quad (6.8)$$

$$\begin{aligned}\mathcal{T}_S : (\Psi \circ \Phi)_{x^{(2)}} \Delta \check{x}^{(2)} + (\Psi \circ \Phi)_y \Delta \check{y} \\ + (1/2) \left((\Psi \circ \Phi)_{x^{(2)}x^{(2)}\check{x}^{(2)}} \cdot \check{x}^{(2)} + 2(\Psi \circ \Phi)_{x^{(2)}y\check{x}^{(2)}} \cdot \check{y} + (\Psi \circ \Phi)_{yy\check{y}} \cdot \check{y} \right) \\ = O(|y^{(2)}|^3).\end{aligned}\quad (6.9)$$

Only (6.6) was used to derive (6.9). Using (6.7) and that $y^{(1)} = O(|y^{(2)}|^2)$, $x^{(2)} = O(|y^{(2)}|)$, $(\Psi \circ \Phi)_y = (1, 0, \dots, 0)$, $(\Psi \circ \Phi)_{x^{(2)}} = 0$, and $\Phi_y \check{y} = O(|y^{(2)}|^2)$ yields

$$\mathcal{T}_{x_0} : (d\Psi \Phi_{y_1}) \Delta y_1 + (1/2) d\Psi (\Phi_{y^{(2)}y^{(2)}} y^{(2)} \cdot y^{(2)}) = O(|y^{(2)}|^3); \quad (6.10)$$

$$\mathcal{T}_S : (d\Psi \Phi_{y_1}) \Delta \check{y}_1 + (1/2) d\Psi (\Phi_{x^{(2)}y^{(2)}\check{x}^{(2)}} \cdot y^{(2)} + \Phi_{y^{(2)}y^{(2)}} y^{(2)} \cdot y^{(2)}) = O(|y^{(2)}|^3).$$

Subtracting the two equations gives (recall that $\xi_0 = d\Psi$):

$$\Delta y_1 - \Delta \check{y}_1 = \Theta_0 \cdot (\Delta y - \Delta \check{y}) = (1/2) \xi_0 \cdot \Phi_{x^{(2)}y^{(2)}\check{x}^{(2)}} y^{(2)} + O(|y^{(2)}|^3). \quad (6.11)$$

Solving (6.7) for $\check{x}^{(2)}$ and substituting into (6.11) we get the formula for Q in (6.3). That Q is non-degenerate follows from (3.13). $\square$

Remark 6.3 In this remark we discuss the meaning of condition (2) in Definition 3.4 from two perspectives. First, consider the image domain perspective. An equation for $\hat{\mathcal{S}}_{y_0}$ is $\Psi(x_1, \Phi^\perp(x^{(2)}, y_0)) = 0$, where x_1 is viewed as a function of $x^{(2)}$. Hence

$$\Psi_{x_1} \partial_{x^{(2)}} x_1 + \Psi_{x^\perp} \partial_{x^{(2)}} \Phi^\perp = 0, \quad \Psi_{x_1} \partial_{x^{(2)}}^2 x_1 + \Psi_{x^\perp x^\perp} \partial_{x^{(2)}} \Phi^\perp = 0. \quad (6.12)$$

Here we have used that $\Psi_{x^\perp} = 0$ implies $\partial_{x^{(2)}} x_1 = 0$. By (3.5) and (3.8),

$$|d\Psi| \partial_{x^{(2)}}^2 x_1 + \Psi_{x^{(2)} x^{(2)}} = 0. \quad (6.13)$$

Using (3.5) and (3.8) again and then (4.14) gives

$$(\Psi \circ \Phi)_{x^{(2)} x^{(2)}} = \Psi''_{x^{(2)} x^{(2)}} + |d\Psi| \partial_{x^{(2)}}^2 \Phi_1 = |d\Psi| \partial_{x^{(2)}}^2 (\Phi_1 - x_1) = |d\Psi| \Delta \Pi_{\mathcal{S}}. \quad (6.14)$$

This shows that if Assumptions 3.1 hold, condition (2) is equivalent to the requirement that $\Delta \Pi_{\mathcal{S}}$ be either positive definite or negative definite.

To understand condition (2) from the data domain perspective, look at the surfaces $\mathcal{T}_{x_0}$ and $\mathcal{T}_{\mathcal{S}}$. Define similarly to (4.14):

$$\Delta \Pi_{\mathcal{T}} := \Pi_{\mathcal{T}_{x_0}} - \Pi_{\hat{\mathcal{T}}_{x_0}}, \quad (6.15)$$

where $\hat{\mathcal{T}}_{x_0}$ is the projection of $\mathcal{T}_{x_0}$ onto $\mathcal{T}_{\mathcal{S}}$ along the first coordinate y_1 , and $\Pi_{\mathcal{T}_{x_0}}, \Pi_{\hat{\mathcal{T}}_{x_0}}$ are the matrices of the second fundamental form of $\mathcal{T}_{x_0}, \hat{\mathcal{T}}_{x_0}$, respectively, at y_0 written in the coordinates $(y_1, y^{(2)})^T$. By construction, $y = Z(y^{(2)})$ is the equation of $\hat{\mathcal{T}}_{x_0}$. By Lemma 6.2, $\Delta \Pi_{\mathcal{T}} = -Q$. Hence, if Assumptions 3.1 hold, condition (2) in Definition 3.4 is equivalent to the requirement that $\Delta \Pi_{\mathcal{T}}$ be either positive definite or negative definite.

Let $y = Y(y^{(2)}, x_\epsilon)$ be the equation for $\mathcal{T}_{x_\epsilon}$ (see the proof of Lemma 3.7). This equation is obtained by solving $\epsilon \check{x}^{(1)} = \Phi^{(1)}(\epsilon \check{x}^{(2)}, y)$ for $y^{(1)}$ and setting $y^{(2)} \equiv Y^{(2)}(y^{(2)}, x_\epsilon)$. Suppose $|\check{x}| = O(1)$ and $|y^{(2)}| = O(\epsilon^{1/2})$. The term $\epsilon \check{x}$ is of a lower order than $y^{(2)}$, so the equation for $\mathcal{T}_{x_0}$ in (6.5) is accurate on $\mathcal{T}_{x_\epsilon}$ to the order $\epsilon^{1/2}$. Due to $\Phi_{x^{(2)}}^{(1)} = 0$, the updated version of (6.8) becomes

$$\Phi_{y^{(1)} y^{(1)}}^{(1)} + (1/2) \Phi_{y^{(2)} y^{(2)}}^{(1)} y^{(2)} y^{(2)} = \epsilon \check{x}^{(1)} + O(\epsilon^{3/2}). \quad (6.16)$$

The terms $(1/2) \Phi_{y^{(2)} y^{(2)}}^{(1)} y^{(2)} y^{(2)}$ in (6.8) and in (6.16) are the same. Also, $\Delta y^{(1)}$ in (6.8) is the analogue of $y^{(1)}$ in (6.16). Therefore, to order ϵ , introduction of the term $\epsilon \check{x}$ requires only a linear correction compared with $Y_0(y^{(2)}) = Y(y^{(2)}, x_0)$, and we have

$$Y^{(1)}(y^{(2)}, x_\epsilon) = Y_0^{(1)}(y^{(2)}) + \epsilon (\Phi_{y^{(1)}}^{(1)})^{-1} \check{x}^{(1)} + O(\epsilon^{3/2}) \text{ if } |y^{(2)}| = O(\epsilon^{1/2}). \quad (6.17)$$

Hence

$$Y_x^{(1)} = ((\Phi_{y^{(1)}}^{(1)})^{-1}, 0), \quad Y_x^{(2)} = 0. \quad (6.18)$$

7 On Some Properties of the Continuous Data g and Its Interpolated Version g_ϵ

By [14, Proposition 25.1.3], g is a conormal distribution with respect to $\Gamma (= \mathcal{T}_S)$. The wave front set of g is contained in the conormal bundle of Γ :

$$WF(g) \subset N^*\Gamma \setminus \mathbf{0} = \{(y, \eta) \in T^*\mathcal{V} : P(y) = 0, \eta = \lambda dP(y), \lambda \neq 0\}. \quad (7.1)$$

See also Sect. 18.2 and Definition 18.2.6 in [13] for a formal definition and in-depth discussion of conormal distributions. A discussion of closely related Lagrangian distributions is in Sect. 25.1 of [14].

In this paper we use two types of spaces of continuous functions. First, $C_b^k(\mathbb{R}^n)$, $k \in \mathbb{N}_0$, is the Banach space of functions with bounded derivatives up to order k . The norm in $C_b^k(\mathbb{R}^n)$ is given by

$$\|h\|_{C_b^k} := \max_{|m| \leq k} |h^{(m)}|_{L^\infty}. \quad (7.2)$$

The subscript ‘0’ in C_0^k means that we consider the subspace of compactly supported functions, $C_0^k(\mathbb{R}^n) \subset C_b^k(\mathbb{R}^n)$.

The second type is the Hölder-Zygmund spaces $C_*^r(\mathbb{R}^n)$, $r > 0$. Pick any $\mu_0 \in C_0^\infty(\mathbb{R}^n)$ such that $\mu_0(\eta) = 1$ for $|\eta| \leq 1$, $\mu_0(\eta) = 0$ for $|\eta| \geq 2$, and define $\mu_j(\eta) := \mu_0(2^{-j}\eta) - \mu_0(2^{-j+1}\eta)$, $j \in \mathbb{N}$ [1, Sect. 5.4]. Then

$$\begin{aligned} C_*^r(\mathbb{R}^n) &:= \{h \in C_b^0(\mathbb{R}^n) : \|h\|_{C_*^r} < \infty\}, \\ \|h\|_{C_*^r} &:= \sup_{j \in \mathbb{N}_0} 2^{jr} \|\mathcal{F}^{-1}(\mu_j(\eta) \tilde{h}(\eta))\|_{L^\infty}, \end{aligned} \quad (7.3)$$

where $\tilde{h} = \mathcal{F}h$. If $r \notin \mathbb{Z}$, i.e. $r = k + \gamma$, $k \in \mathbb{N}_0$, $0 < \gamma < 1$, then $C_*^r(\mathbb{R}^n)$ consists of $C_b^k(\mathbb{R}^n)$ functions, which have Hölder continuous k -th order derivatives (see [30, Definition 2.4 and Example 2.3]):

$$\max_{|m|=k} \sup_{x \in \mathbb{R}^n, |h|>0} \frac{|f^{(m)}(x+h) - f^{(m)}(x)|}{|h|^\gamma} < \infty. \quad (7.4)$$

As is easily seen, $C_b^k \subset C_*^k$ if $k \in \mathbb{N}$. The Hölder-Zygmund spaces are a particular case of the Besov spaces: $C_*^r(\mathbb{R}^n) = B_{p,q}^r(\mathbb{R}^n)$, where $p, q = \infty$ [1, item 2 in Remark 6.4].

The following two lemmas are proven in Appendix C.

Lemma 7.1 Suppose g satisfies Assumption 4.5. There exist $c_m > 0$ such that

$$|\partial_y^m g(y)| \leq c_m \begin{cases} |P(y)|^{s_0-|m|}, & |m| > s_0, \\ 1, & |m| \leq s_0, \end{cases} \quad \forall m \in \mathbb{N}_0^n, y \in \mathcal{V} \setminus \Gamma. \quad (7.5)$$

Additionally,

$$g \in C_*^{s_0}(\mathcal{V}) \quad \forall s_0 > 0 \text{ and } g \in C_0^{s_0}(\mathcal{V}) \text{ if } s_0 \in \mathbb{N}. \quad (7.6)$$

If the leading term in $\tilde{v}$ is missing, i.e., $\tilde{v}^\pm \equiv 0$, then $g \in C_*^{s_1}(\mathcal{V})$, and (7.5) holds with s_0 replaced by s_1 .

Lemma 7.2 Suppose $\mathcal{B}$ and g satisfy Assumptions 4.2, 4.5, and 4.6. There exists $c_\beta > 0$ such that

$$|(\mathcal{B}g)(y)| \leq c_\beta |P(y)|^{s_0-\beta_0}, \quad \forall y \in \mathcal{V} \setminus \Gamma. \quad (7.7)$$

If $\kappa = 0$, we additionally have with some $c > 0$

$$|(\mathcal{B}g)(y)| \leq c_\beta P(y)^{s_0-\beta_0+c}, \text{ for any } y \in \mathcal{V} \text{ if } P(y) > 0. \quad (7.8)$$

Define

$$g_\epsilon^{(l)}(y) := \sum_{|j| \leq \vartheta/\epsilon} \partial_{y_1}^l \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) g(\hat{y}^j), \quad \Delta g_\epsilon^{(l)}(y) := \partial_{y_1}^l (g_\epsilon(y) - g(y)), \quad 0 \leq l \leq \lceil \beta_0^+ \rceil. \quad (7.9)$$

The following two lemmas are proven in Appendix C.

Lemma 7.3 Suppose φ and g satisfy Assumptions 4.3 and 4.5, respectively. There exists $\varkappa_1 > 0$ such that

$$|g_\epsilon^{(l)}(y)| \leq c \begin{cases} |P(y)|^{s_0-l}, & |P(y)| \geq \varkappa_1 \epsilon, \quad s_0 < l \leq \lceil \beta_0^+ \rceil, \\ \epsilon^{s_0-l}, & |P(y)| \leq \varkappa_1 \epsilon, \quad s_0 < l \leq \lceil \beta_0^+ \rceil, \\ 1, & 0 \leq l \leq s_0, \end{cases} \quad \forall y \in \mathcal{V}, \quad (7.10)$$

for some $c > 0$.

If the top order term in $\tilde{v}$ is missing, i.e., $\tilde{v}^\pm \equiv 0$, then (7.10) holds with s_0 replaced by s_1 as long as $l \leq \lceil \beta_0^+ \rceil$.

Lemma 7.4 Suppose φ and g satisfy Assumptions 4.3 and 4.5, respectively. Let $\varkappa_1$ be the same as in Lemma 7.3. One has

$$|\Delta g_\epsilon^{(l)}(y)| \leq c \epsilon |P(y)|^{s_0-1-l}, \quad y \in \mathcal{V}, \quad |P(y)| \geq \varkappa_1 \epsilon, \quad \lfloor s_0^- \rfloor \leq l \leq \lceil \beta_0^+ \rceil, \quad (7.11)$$

$$|\Delta g_\epsilon^{(l)}(y)| \leq c \epsilon^{s_0-l}, \quad y \in \mathcal{V}, \quad 0 \leq l \leq \lfloor s_0^- \rfloor, \quad (7.12)$$

for some $c > 0$.

If the top order term in $\tilde{v}$ is missing, i.e. $\tilde{v}^\pm \equiv 0$, then (7.11), (7.12) hold with s_0 replaced by s_1 as long as $l \leq \lceil \beta_0^+ \rceil$.

8 Computing the First Part of the Leading Term

8.1 Splitting the Reconstruction into Two Parts: $f_\epsilon = f_\epsilon^{(1)} + f_\epsilon^{(2)}$

By (4.5),

$$\check{f}_\epsilon(x_\epsilon) = \int_{y \in \mathcal{T}_{x_\epsilon}} (\mathcal{B}g_\epsilon)(y) w(x_\epsilon, y) dy. \quad (8.1)$$

Pick some large $A \gg 1$ and introduce two sets

$$\begin{aligned} \Omega_1 &:= \left\{ y^{(2)} \in \mathbb{R}^N : |y^{(2)}| \leq A\epsilon^{1/2} \right\}, \\ \Omega_2 &:= \left\{ y^{(2)} \in \mathbb{R}^N : |y^{(2)}| \geq A\epsilon^{1/2}, \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix} \in \mathcal{V} \right\}. \end{aligned} \quad (8.2)$$

Let $f_\epsilon^{(l)}(x_\epsilon)$ denote the reconstruction obtained using (8.1), where the y integration is restricted to the part of $\mathcal{T}_{x_\epsilon}$ corresponding to Ω_l , $l = 1, 2$, respectively. The main ideas behind the split are that (1) The contribution of $f_\epsilon^{(2)}(x_\epsilon)$ to the DTB goes to zero when $A \rightarrow \infty$; (2) For each fixed $A > 0$, the fact that $|y^{(2)}| = O(\epsilon^{1/2})$ greatly simplifies estimation of $f_\epsilon^{(1)}(x_\epsilon)$; and (3) The double limit $\lim_{A \rightarrow \infty} \lim_{\epsilon \rightarrow 0} f_\epsilon^{(1)}(x_\epsilon)$ exists and gives the DTB.

The remainder of the proof consists of four parts: (1) Show that the leading singular part of $f_\epsilon^{(1)}(x_\epsilon)$ (i.e., when only the top order terms are retained in $\mathcal{B}$ and g) gives the main contribution to the DTB. This is done in the rest of this section; (2) Show that the remaining, less singular part of $f_\epsilon^{(1)}(x_\epsilon)$ does not contribute to the DTB (Sect. 9); (3) Show that the contribution of $f_\epsilon^{(2)}(x_\epsilon)$ to the DTB can be made as small as one likes by selecting $A > 0$ sufficiently large (Sect. 10); and (4) Compute the DTB (Sect. 11).

8.2 Estimation of the Leading Term of $f_\epsilon^{(1)}(x_\epsilon)$

Throughout this section we assume that $\mathcal{B}$ in (4.3) satisfies $\tilde{B}(y, \eta) \equiv \tilde{B}_0(y, \eta)$, i.e., we assume that the symbol of $\mathcal{B}$ contains only the top order term. Let $\mathcal{B}_0$ denote the Ψ DO of the form (4.3), where $\tilde{B}(y, \eta) \equiv \tilde{B}_0(y_0, \eta)$. Likewise, we assume that the symbol of g coincides with its top order term (i.e., $\tilde{R} \equiv 0$ in (4.15)). It then follows from (C.1) to (C.7) (see also (11.3)) that g is given by

$$g(y) = a^+(y)P_+^{s_0}(y) + a^-(y)P_-^{s_0}(y), \quad a^\pm \in C_0^\infty(\mathcal{V}), \quad (8.3)$$

where $a^\pm(y)$ are linear combinations of $\tilde{v}^\pm(y)$. Substitute (8.3) into (8.1)

$$f_\epsilon^{(1)}(x_\epsilon) = \int_{\substack{y \in \mathcal{T}_{x_\epsilon} \\ |y^{(2)}| \leq A\epsilon^{1/2}}} (\mathcal{B}g_\epsilon)(y)w(x_\epsilon, y)dy. \quad (8.4)$$

Introduce the operator

$$\mathcal{B}_{1d}g := \mathcal{F}_{1d}^{-1}(\tilde{b}(\lambda)\tilde{g}(\lambda)), \quad \tilde{b}(\lambda) := \tilde{B}_0(y_0, \Theta_0)\lambda_+^{\beta_0} + \tilde{B}_0(y_0, -\Theta_0)\lambda_-^{\beta_0}, \quad (8.5)$$

where g is sufficiently smooth and decays sufficiently fast, and $\mathcal{F}_{1d}$ denotes the 1D Fourier transform. Introduce an auxiliary function:

$$\Upsilon(p) := \int \mathcal{B}_{1d}\hat{\varphi}(\Theta_0, p-q)\mathcal{A}(q)dq = \mathcal{F}_{1d}^{-1}(\tilde{\varphi}(\lambda\Theta_0)\tilde{b}(\lambda)\tilde{\mathcal{A}}(\lambda)), \quad p \in \mathbb{R}, \quad (8.6)$$

where $\tilde{\varphi} = \mathcal{F}\varphi$, $\tilde{\mathcal{A}} = \mathcal{F}_{1d}\mathcal{A}$,

$$\mathcal{A}(p) := a^+(y_0)p_+^{s_0} + a^-(y_0)p_-^{s_0}, \quad p \in \mathbb{R}, \quad (8.7)$$

$\mathcal{B}_{1d}$ acts with respect to the affine variable, and the hat denotes the classical Radon transform that integrates over hyperplanes:

$$\hat{\varphi}(\Theta_0, p) := \int \varphi(x)\delta(\Theta_0 \cdot x - p)dx. \quad (8.8)$$

Both $\tilde{b}(\lambda)$ and $\tilde{\mathcal{A}}(\lambda)$ are not smooth at $\lambda = 0$, so the product $\tilde{b}(\lambda)\tilde{\mathcal{A}}(\lambda)$ needs to be computed carefully, see the discussion between (11.4) and (11.6).

The main result in this section is the following lemma.

Lemma 8.1 *Suppose the symbols of $\mathcal{B}$ and g contain only the top order terms as described above. Under the assumptions of Theorem 4.7 one has*

$$\lim_{\epsilon \rightarrow 0} \epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon) = w(x_0, y_0) \int_{|\check{y}^{(2)}| \leq A} \Upsilon\left(\frac{\partial Y_1}{\partial x_1}\check{x}_1 - \frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2}\right) d\check{y}^{(2)}. \quad (8.9)$$

8.3 Proof of Lemma 8.1

We begin by investigating the sum in (8.4). The key result is the following lemma (see Appendix D for the proof).

Lemma 8.2 *Suppose $y, z \in \mathcal{V}$ satisfy*

$$|y - y_0| \leq c\epsilon^{1/2}, \quad |y - z| \leq c\epsilon, \quad z \in \Gamma, \quad (8.10)$$

for some $c > 0$, and g_ϵ is obtained by interpolating g in (8.3) (cf. (4.6)). One has

$$\epsilon^{\beta_0-s_0}(\mathcal{B}g_\epsilon)(y) = \sum_{j \in \mathbb{Z}^n} \mathcal{B}_0 \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) \mathcal{A} \left(\Theta_0 \cdot \frac{\hat{y}^j - z}{\epsilon} \right) + O(\epsilon^{\min(s_0, 1)/2}), \quad \epsilon \rightarrow 0, \quad (8.11)$$

where the series on the right converges absolutely, and the big- O term is uniform with respect to z , y satisfying (8.10). Moreover, the left-hand side of (8.11) remains bounded as $\epsilon \rightarrow 0$ uniformly with respect to z , y satisfying (8.10).

The next step is to use (8.11) in (8.4):

$$\begin{aligned} \epsilon^{\beta_0-s_0} f_\epsilon^{(1)}(x_\epsilon) &= O(\epsilon^{(N+\min(s_0, 1))/2}) \\ &+ w(x_0, y_0) \int_{\Omega_1} \sum_{j \in \mathbb{Z}^n} \mathcal{B}_0 \varphi \left(\frac{Y(y^{(2)}, x_\epsilon) - \hat{y}^j}{\epsilon} \right) \mathcal{A} \left(\Theta_0 \cdot \frac{\hat{y}^j - Z(y^{(2)})}{\epsilon} \right) dy^{(2)}. \end{aligned} \quad (8.12)$$

Here $Z(y^{(2)})$ is obtained by projecting $Y_0(y^{(2)})$ onto Γ along y_1 , see Lemma 6.2. From (3.17) and $Y_0(y^{(2)}, x) \equiv y^{(2)}$ we have $G^T = I_N$ and $\det G^T = 1$, where G^T is the Gram matrix (4.2) evaluated at $y^{(2)} = 0, x = x_0$. This property was used to obtain (8.12). By (6.17) and Lemma 6.2,

$$\begin{aligned} |Y(y^{(2)}, x_\epsilon) - Z(y^{(2)})| &= |Y_0(y^{(2)}) + O(\epsilon) - Z(y^{(2)})| \\ &= |Y_0(y^{(2)}) - Z(y^{(2)})| + O(\epsilon) = O(|y^{(2)}|^2) + O(\epsilon) = O(\epsilon), \end{aligned} \quad (8.13)$$

so the conditions in (8.10) hold, and (8.11) applies.

In view of (8.12), define

$$\phi(v, p) := w(x_0, y_0) \sum_{j \in \mathbb{Z}^n} \mathcal{B}_0 \varphi(v - D_1 j) \mathcal{A}(\Theta_0 \cdot D_1 j - p), \quad v \in \mathbb{R}^n, p \in \mathbb{R}. \quad (8.14)$$

Recall that $D_1 = U^T D$ (cf. Assumption 4.3(IK2)). As is easily checked,

$$\begin{aligned} \phi(v + D_1 m, p + \Theta_0 \cdot D_1 m) &= \phi(v, p), \quad \forall m \in \mathbb{Z}^n, v \in \mathbb{R}^n, p \in \mathbb{R}; \quad (8.15) \\ \int_{[0, 1]^n} \phi(v + D_1 u, p + \Theta_0 \cdot D_1 u) du &= w(x_0, y_0) \int_{\mathbb{R}^n} \mathcal{B}_0 \varphi(v - u) \mathcal{A}(\Theta_0 \cdot u - p) du \\ &= w(x_0, y_0) \Upsilon(\Theta_0 \cdot v - p), \quad \forall v \in \mathbb{R}^n, p \in \mathbb{R}. \end{aligned} \quad (8.16)$$

Use (8.14) to rewrite (8.12):

$$\epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon) = I(\epsilon) + O(\epsilon^{\min(s_0, 1)/2}), \quad (8.17)$$

where

$$\begin{aligned}
 I(\epsilon) &:= \epsilon^{-N/2} \int_{\Omega_1} \phi \left(\frac{Y(y^{(2)}, x_\epsilon) + U^T \tilde{y}_0}{\epsilon}, \Theta_0 \cdot \frac{Z(y^{(2)}) + U^T \tilde{y}_0}{\epsilon} \right) dy^{(2)} \\
 &= \epsilon^{-N/2} \int_{\Omega_1} \phi \left(\frac{Y(y^{(2)}, x_\epsilon) - Y_0(y^{(2)})}{\epsilon} + D_1 u_\epsilon, \right. \\
 &\quad \left. \Theta_0 \cdot \left(-\frac{Y_0(y^{(2)}) - Z(y^{(2)})}{\epsilon} + D_1 u_\epsilon \right) \right) dy^{(2)}, \\
 u_\epsilon &:= \frac{D_1^{-1}(U^T \tilde{y}_0 + Y_0(y^{(2)}))}{\epsilon}.
 \end{aligned} \tag{8.18}$$

Using that $\check{x} = O(1)$ and $|y^{(2)}| = O(\epsilon^{1/2})$, (6.17) and (6.18) imply

$$\frac{Y(y^{(2)}, x_\epsilon) - Y_0(y^{(2)})}{\epsilon} = Y_x(y^{(2)}, 0)\check{x} + O(\epsilon) = v_0 + O(\epsilon^{1/2}), \quad v_0 := \begin{pmatrix} Y_x^{(1)} \check{x} \\ 0 \end{pmatrix}. \tag{8.19}$$

By Lemma 6.2,

$$\Theta_0 \cdot (Y_0(y^{(2)}) - Z(y^{(2)})) = -\frac{Qy^{(2)} \cdot y^{(2)}}{2} + O(|y^{(2)}|^3). \tag{8.20}$$

Therefore,

$$I(\epsilon) = \int_{|\check{y}^{(2)}| \leq A} \phi \left(v_0 + D_1 u_\epsilon + O(\epsilon^{1/2}), \frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2} + \Theta_0 \cdot D_1 u_\epsilon + O(\epsilon^{1/2}) \right) d\check{y}^{(2)}, \tag{8.21}$$

where $y^{(2)} = \epsilon^{1/2} \check{y}^{(2)}$.

The following result is proven in Appendix E.

Lemma 8.3 *Pick any c , $0 < c < \infty$. One has*

$$\begin{aligned}
 \phi(v + \Delta v, p) - \phi(v, p) &= O(|\Delta v|^{1-\{\beta_0\}}), \quad \Delta v \rightarrow 0, \quad |v|, |p| \leq c; \\
 \phi(v, p + \Delta p) - \phi(v, p) &= O(|\Delta p|^{\min(s_0, 1)}), \quad \Delta p \rightarrow 0, \quad |v|, |p| \leq c;
 \end{aligned} \tag{8.22}$$

and the two big- O terms are uniform in v and p confined to the indicated sets.

Introduce an auxiliary function

$$\phi_1(q; u) := \phi(v_0 + D_1 u, q + \Theta_0 \cdot D_1 u), \quad u \in \mathbb{R}^n, \quad q \in \mathbb{R}. \tag{8.23}$$

By Lemma 8.3, the integrand in (8.21) can be written in the form:

$$\phi_1 \left(\frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2}; u_\epsilon \right) + O(\epsilon^{a/2}), \quad (8.24)$$

where $a = \min(1 - \{\beta_0\}, s_0, 1) > 0$.

Consider now u_ϵ . The assumption $|y^{(2)}| = O(\epsilon^{1/2})$ implies

$$Y_0(y^{(2)}) = Y'_0(0)y^{(2)} + \frac{Y''_0(0)y^{(2)} \cdot y^{(2)}}{2} + O(\epsilon^{3/2}), \quad (8.25)$$

where we have used that $Y_0(0) = 0$. Here $Y'_0(0) = \partial_{y^{(2)}} Y_0(y^{(2)})|_{y^{(2)}=0}$ and $Y''_0(0) = \partial_{y^{(2)}}^2 Y_0(y^{(2)})|_{y^{(2)}=0}$. For completeness, note that $Y'_0(0) = (0, I_N)^T$ (see (3.17)), but this is not used in what follows.

From (8.15), $\phi_1(q; u + m) = \phi_1(q; u)$ for any $m \in \mathbb{Z}^n$. Using Lemma 8.3 again and (8.25), (8.24) becomes

$$\phi_1 \left(\frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2}; \frac{D^{-1}\tilde{y}_0}{\epsilon} + D_1^{-1} \left(\frac{Y'_0(0)\check{y}^{(2)}}{\epsilon^{1/2}} + \frac{Y''_0(0)\check{y}^{(2)} \cdot \check{y}^{(2)}}{2} \right) \right) + O(\epsilon^{a/2}). \quad (8.26)$$

Thus, we need to compute the limit of the following integral as $\epsilon \rightarrow 0$:

$$\begin{aligned} J(\epsilon) &:= \int_{|\check{y}^{(2)}| \leq A} \phi_1(q; u) d\check{y}^{(2)}, \quad q = q(\check{y}^{(2)}) = \frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2}, \\ u &= u(\check{y}^{(2)}, \epsilon) = \frac{D_1^{-1}Y'_0(0)\check{y}^{(2)}}{\epsilon^{1/2}} + \left[\frac{D^{-1}\tilde{y}_0}{\epsilon} + D_1^{-1} \frac{Y''_0(0)\check{y}^{(2)} \cdot \check{y}^{(2)}}{2} \right]. \end{aligned} \quad (8.27)$$

Represent ϕ_1 in terms of its Fourier series:

$$\phi_1(q; u) = \sum_{m \in \mathbb{Z}^n} \tilde{\phi}_{1,m}(q) e^{2\pi i m \cdot u}. \quad (8.28)$$

The N columns of $Y'_0(0)$ form a basis for the tangent space to $\mathcal{T}_{x_0}$ at $y_0 = 0$ written in the new y coordinates. The columns of $UY'_0(0)$ span the tangent space to $\tilde{\mathcal{T}}_{x_0}$ at $\tilde{y}_0$ written in the original $\tilde{y}$ coordinates. By assumption, $\tilde{\mathcal{T}}_{x_0}$ is generic at $\tilde{y}_0$ with respect to D (cf. Definition 3.4), so there is no $m \in \mathbb{Z}^n$ such that $m \neq 0$ and $mD_1^{-1}Y'_0(0) = 0$. The same argument as in (5.8)–(5.14) in [20] implies

$$\lim_{\epsilon \rightarrow 0} J(\epsilon) = \int_{|\check{y}^{(2)}| \leq A} \int_{[0,1]^n} \phi_1 \left(\frac{Q\check{y}^{(2)} \cdot \check{y}^{(2)}}{2}, u \right) du d\check{y}^{(2)}. \quad (8.29)$$

Here is an outline of the argument. Break up the integral with respect to $\check{y}^{(2)}$ in (8.27) into a sum of integrals over a finite, pairwise disjoint covering of the domain of integration by subdomains B_k with diameter $0 < \delta \ll 1$. Then approximate each of these integrals by assuming that $\check{y}^{(2)}$ is constant everywhere except in the first term of u outside of brackets in (8.27). This is done by choosing $\check{y}_k^{(2)} \in B_k$ in an arbitrary fashion:

$$J(\epsilon) = \sum_k [J_k(\epsilon) + O(\delta^a) \text{Vol}(B_k)], \quad J_k(\epsilon) := \int_{B_k} \phi_1(q(\check{y}_k^{(2)}); u_k(\check{y}^{(2)}, \epsilon)) d\check{y}^{(2)},$$

$$u_k(\check{y}^{(2)}, \epsilon) = \frac{D_1^{-1} Y_0'(0) \check{y}^{(2)}}{\epsilon^{1/2}} + \left[\frac{D^{-1} \tilde{y}_0}{\epsilon} + D_1^{-1} \frac{Y_0''(0) \check{y}_k^{(2)} \cdot \check{y}_k^{(2)}}{2} \right]. \quad (8.30)$$

Thus, the variable of integration $\check{y}^{(2)}$ is present only in the rapidly changing term with $\epsilon^{1/2}$ in the denominator. The magnitude of the error term $O(\delta^a)$ follows from Lemma 8.3. Represent each ϕ_1 in (8.30) in terms of its Fourier series (8.28). Using the fact that there is no $m \in \mathbb{Z}^n$ such that $m \neq 0$ and $m D_1^{-1} Y_0'(0) = 0$ implies

$$\lim_{\epsilon \rightarrow 0} \int_{B_k} \exp(2\pi i m \cdot u(\check{y}^{(2)}, \epsilon)) d\check{y}^{(2)} = 0 \text{ if } m \neq 0. \quad (8.31)$$

The first term in u_k (cf. (8.30)) is the only one that contains $\check{y}^{(2)}$ and changes rapidly as $\epsilon \rightarrow 0$. In turn, (8.31) implies

$$\lim_{\epsilon \rightarrow 0} J_k(\epsilon) = \tilde{\phi}_{1,m=0}(q(\check{y}_k^{(2)})) = \int_{[0,1]^n} \phi_1(q(\check{y}_k^{(2)}); u) du. \quad (8.32)$$

Using that $\delta > 0$ can be as small as we like finishes the proof of (8.29).

Combining (8.17), (8.21), (8.23), (8.26), (8.27), (8.29) gives

$$\lim_{\epsilon \rightarrow 0} \epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon) = \int_{|\check{y}^{(2)}| \leq A} \int_{[0,1]^n} \phi \left(v_0 + D_1 u, \frac{Q \check{y}^{(2)} \cdot \check{y}^{(2)}}{2} + \Theta_0 \cdot D_1 u \right) du d\check{y}^{(2)}. \quad (8.33)$$

By (8.16) and (8.33),

$$\lim_{\epsilon \rightarrow 0} \epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon) = w(x_0, y_0) \int_{|\check{y}^{(2)}| \leq A} \Upsilon \left(\Theta_0 \cdot v_0 - \frac{Q \check{y}^{(2)} \cdot \check{y}^{(2)}}{2} \right) d\check{y}^{(2)}. \quad (8.34)$$

Since $\partial Y_1 / \partial x^\perp = 0$ (cf. (11.8)) and $\Theta_0 = dy_1$, from the definition of v_0 in (8.19)

$$\Theta_0 \cdot \begin{pmatrix} Y_x^{(1)} \check{x} \\ 0 \end{pmatrix} = \frac{\partial Y_1}{\partial x_1} \check{x}_1. \quad (8.35)$$

Combining (8.34) and (8.35) proves Lemma 8.1.

9 Estimation of the Remaining Parts of $f_{\epsilon}^{(1)}(x_{\epsilon})$

In this section we prove that the lower order terms of $f_{\epsilon}^{(1)}(x_{\epsilon})$ do not contribute to the DTB. In this section, by c we denote various positive constants that may have different values in different places. From (4.16), (6.3), (6.17), and (8.2) it follows that there exists $c_1 > 0$ such that

$$y^{(2)} \in \Omega_1, y \in \mathcal{T}_{x_{\epsilon}} \text{ implies } |P(y)| \leq c_1 \epsilon. \quad (9.1)$$

Let β and s denote the remaining highest order exponents in (4.8) and (4.15), respectively. By construction, $\beta_0 - s_0 > \beta - s$. This means that either $s = s_0$ if the first term in $\mathcal{B}$ is missing (i.e., $\beta = \beta_1 < \beta_0$), or $\beta = \beta_0$ if the first term in $\tilde{v}$ is missing (i.e., $s = s_1 > s_0$).

Suppose initially that $\beta > \lfloor s^- \rfloor$. Set $k := \lceil \beta^+ \rceil$, $\nu := k - \beta$. Thus, $0 < \nu \leq 1$, $\nu = 1$ if $\beta \in \mathbb{N}_0$, and $s_0 \leq s < k \leq \lceil \beta_0^+ \rceil$. Clearly,

$$\mathcal{B} = \mathcal{W}_1 \partial_{y_1}^k + \mathcal{W}_2, \quad (9.2)$$

for some $\mathcal{W}_1 \in S^{-\nu}(\mathcal{V} \times \mathbb{R}^n)$ and $\mathcal{W}_2 \in S^{-\infty}(\mathcal{V} \times \mathbb{R}^n)$. Here we use a cut-off near $\eta = 0$ and the fact that the amplitude of $\mathcal{B}$ is supported in a small conic neighborhood of $(y_0, \Theta_0 = dy_1)$. Then

$$(\mathcal{B}g_{\epsilon})(y) = \int K(y, y-w) g_{\epsilon}^{(k)}(w) dw + O(1), \quad (9.3)$$

where $K(y, w)$ is the Schwartz kernel of $\mathcal{W}_1$, and $O(1)$ represents $\mathcal{W}_2 g_{\epsilon}(y)$. The latter statement follows, because $g_{\epsilon}(y)$ is uniformly bounded as $\epsilon \rightarrow 0$ for all $y \in \mathcal{V}$ (cf. (4.6) and (7.6)) and compactly supported. By the estimate [1, Eq. (5.13)],

$$|\partial_{w_1}^l K(y, y-w)| \leq c(l) |y-w|^{-(n-\nu+l)}, \quad l \geq 0, \quad y, w \in \mathcal{V}. \quad (9.4)$$

Combining (9.2), (7.9), the two top cases in (7.10) with $l = k$ and s_0 replaced by s , (9.3), and (9.4) with $l = 0$, gives

$$|(\mathcal{B}g_{\epsilon})(y)| \leq c(J_1 + J_2) + O(1),$$

$$J_1 := \int_{\varkappa_1 \epsilon \leq |P(w)| \leq O(1)} \frac{|w_1 - \psi(w^{\perp})|^{s-k}}{|y-w|^{n-\nu}} dw, \quad J_2 := \int_{|P(w)| \leq \varkappa_1 \epsilon} \frac{\epsilon^{s-k}}{|y-w|^{n-\nu}} dw, \quad (9.5)$$

where $\varkappa_1$ is the same as in Lemma 7.3. Consider J_1 :

$$J_1 = \int \int_{\varkappa_1 \epsilon \leq |p| \leq O(1)} \frac{|p|^{s-k}}{|([P-p] + \psi(y^{\perp}) - \psi(w^{\perp}), y^{\perp} - w^{\perp})|^{n-\nu}} dp dw^{\perp}, \quad (9.6)$$

where we denoted $P := P(y)$ and changed variables $w_1 \rightarrow p = w_1 - \psi(w^\perp)$. There exists $0 < c' < 1$ so that

$$|a + \psi(y^\perp) - \psi(w^\perp)| + |y^\perp - w^\perp| \geq c'(|a| + |y^\perp - w^\perp|) \quad \forall a \in \mathbb{R}, y, w \in \mathcal{V}. \quad (9.7)$$

By construction, $\partial_{y^\perp} \psi(0) = 0$. Assume $\mathcal{V}$ is sufficiently small, so that $|\psi(y^\perp) - \psi(w^\perp)| \leq c''|y^\perp - w^\perp|$, $y, w \in \mathcal{V}$, for some $0 < c'' < 1$. Then any c' such that $0 < c' < 1 - c''$ works. This implies

$$\begin{aligned} J_1 &\leq c \int_{|x_1| \leq |p| \leq O(1)} \int \frac{|p|^{s-k}}{(|P - p| + |w^\perp|)^{n-v}} dw^\perp dp \\ &\leq c \int_{|x_1| \leq |p| \leq O(1)} \frac{|p|^{s-k}}{|P - p|^{1-v}} dp = \begin{cases} O(\epsilon^{s-\beta}), & \beta > s, \\ O(\ln(1/\epsilon)), & \beta = s, \\ O(1), & \beta < s. \end{cases} \end{aligned} \quad (9.8)$$

Here we have used that $P = O(\epsilon)$.

The term J_2 can be estimated analogously, and we get an estimate similar to (9.8), where the bound is $O(\epsilon^{s-\beta})$ in all three cases.

Suppose $\beta > s$. By (9.5), $(\mathcal{B}g_\epsilon)(y) = O(\epsilon^{s-\beta})$. Estimate the integral in (8.4):

$$\begin{aligned} |\epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon)| &\leq \epsilon^\kappa \int_{\Omega_1} O(\epsilon^{s-\beta}) dy^{(2)} = O(\epsilon^{\kappa+s-\beta}) \int_0^{A\epsilon^{1/2}} r^{N-1} dr \\ &= O(\epsilon^{(\beta_0-s_0)-(\beta-s)}) \text{ if } \beta > s. \end{aligned} \quad (9.9)$$

In a similar fashion,

$$|\epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon)| = \begin{cases} O(\epsilon^{\beta_0-s_0} \ln(1/\epsilon)), & \beta = s, \\ O(\epsilon^{\beta_0-s_0}), & \beta < s. \end{cases} \quad (9.10)$$

Since $\beta_0 - s_0 \geq N/2 \geq 1/2$, $\epsilon^\kappa f_\epsilon^{(1)}(x_\epsilon) \rightarrow 0$ in all three cases.

Suppose now $0 < \beta \leq \lfloor s^- \rfloor$. Similarly to (9.2), $\mathcal{B} = \mathcal{W}_1 \partial_{y_1}^k + \mathcal{W}_2$, where $k = \lceil \beta \rceil \geq 1$, $v = k - \beta \geq 0$, $s > k$, $\mathcal{W}_1 \in S^{-v}(\mathcal{V} \times \mathbb{R}^n)$, and $\mathcal{W}_2 \in S^{-\infty}(\mathcal{V} \times \mathbb{R}^n)$. The kernel of $\mathcal{W}_1$ is an L^1 function (see e.g. Theorem 5.15 in [1]) and $\sup_{y \in \mathcal{V}} |\Delta g_\epsilon^{(k)}(y)| = O(\epsilon^{s-k})$ (cf. (7.12) with s_0 replaced by s). This implies that $\sup_{y \in \mathcal{V}} |(\mathcal{B}g_\epsilon)(y) - \mathcal{B}g(y)| = O(\epsilon^{s-k})$. From Lemma 7.1, $\mathcal{B}g \in C_*^{s-\beta}(\mathcal{V})$, and $s > \beta$. Thus, $(\mathcal{B}g_\epsilon)(y) = O(1)$, and the desired result follows similarly to the case $\beta < s$ in (9.10). The case $\beta \leq 0$ is proven using the same argument with $l = 0$ in (7.12) and without splitting $\mathcal{B}$ into two parts.

10 Estimation of $f_{\epsilon}^{(2)}(x_{\epsilon})$

10.1 Statement of Results

In this section we prove the following two lemmas.

Lemma 10.1 *Under the assumptions of Theorem 4.7 one has*

$$\lim_{\epsilon \rightarrow 0} \begin{cases} \epsilon^{\kappa} \check{f}_{\epsilon}(x_{\epsilon}), & \kappa > 0 \\ \check{f}_{\epsilon}(x_{\epsilon}) - f^{(2)}(x_{\epsilon}), & \kappa = 0 \end{cases} = w(x_0, y_0) \int_{\mathbb{R}^N} \Upsilon \left(\frac{\partial Y_1}{\partial x_1} \check{x}_1 - \frac{Q \check{y}^{(2)} \cdot \check{y}^{(2)}}{2} \right) d\check{y}^{(2)}. \quad (10.1)$$

Lemma 10.2 *Suppose $\kappa = 0$. Under the assumptions of Theorem 4.7 one has*

$$\lim_{\epsilon \rightarrow 0} f^{(2)}(x_{\epsilon}) = (\mathcal{R}^* \mathcal{B}g)(x_0^{\text{int}}). \quad (10.2)$$

Recall that x_0^{int} is defined in (4.21). Lemma 10.1 is proven by considering the last remaining term $f_{\epsilon}^{(2)}(x_{\epsilon})$:

$$f_{\epsilon}^{(2)}(x_{\epsilon}) = \int_{\substack{y \in \mathcal{T}_{x_{\epsilon}} \\ |y^{(2)}| > A\epsilon^{1/2}}} (\mathcal{B}g_{\epsilon})(y) w(x_{\epsilon}, y) dy, \quad (10.3)$$

where both $\mathcal{B}$ and g are given by their full expressions. The continuous counterpart of (10.3) is

$$f^{(2)}(x_{\epsilon}) = \int_{\substack{y \in \mathcal{T}_{x_{\epsilon}} \\ |y^{(2)}| > A\epsilon^{1/2}}} (\mathcal{B}g)(y) w(x_{\epsilon}, y) dy. \quad (10.4)$$

10.2 Proof of Lemma 10.1

The following lemma is proven in Appendix F.

Lemma 10.3 *Suppose $\mathcal{B}$, g , and φ satisfy Assumptions 4.2, 4.3, 4.5, and 4.6. There exist $c, \kappa_2 > 0$ such that for all $\epsilon > 0$ sufficiently small one has*

$$|(\mathcal{B}g_{\epsilon})(y) - (\mathcal{B}g)(y)| \leq c\epsilon |P(y)|^{s_0-1-\beta_0} \begin{cases} 1, & \beta_0 \notin \mathbb{N}, \\ |\ln(P(y)/\epsilon)|, & \beta_0 \in \mathbb{N}, \end{cases} \quad y \in \mathcal{V}, \quad (10.5)$$

whenever $|P(y)| > \kappa_2\epsilon$.

Return now to (10.3). Pick any $y \in \mathcal{T}_{x_{\epsilon}}$. Recall that $y = Y(y^{(2)}, x)$ is obtained by solving $x^{(1)} = \Phi^{(1)}(x^{(2)}, y)$ for $y^{(1)}$, and that $y^{(2)} \equiv Y^{(2)}(y^{(2)}, x)$ (see the proof of

Lemma 3.7). Hence $|Y(y^{(2)}, x_\epsilon) - Y_0(y^{(2)})| = O(\epsilon)$. Strictly speaking, we cannot invoke (6.17) here, because in (6.17) the assumption is $|y^{(2)}| = O(\epsilon^{1/2})$. Therefore,

$$P(Y(y^{(2)}, x_\epsilon)) = P(Y_0(y^{(2)})) + O(\epsilon) = \Theta_0 \cdot (Y_0(y^{(2)}) - Z(y^{(2)})) + O(\epsilon). \quad (10.6)$$

Recall that $Z(y^{(2)})$ is the projection of $Y_0(y^{(2)})$ onto Γ , cf. Lemma 6.2. Using (6.3), (8.2), and that Q is negative definite, by shrinking $\mathcal{V}$, if necessary, and taking $A \gg 1$ large enough, we can make sure that (a) $P(y) \geq c|y^{(2)}|^2$ for some $c > 0$ and (b) inequality (10.5) applies (i.e. $P(y) > \kappa_2\epsilon$) if $y \in \mathcal{T}_{x_\epsilon}$ and $y^{(2)} \in \Omega_2$ for all $\epsilon > 0$ small enough.

Suppose first that $\kappa > 0$. Using (10.3), (10.5), and (7.7) gives an estimate

$$\begin{aligned} |\epsilon^\kappa f_\epsilon^{(2)}(x_\epsilon)| &\leq O(\epsilon^\kappa) \int_{\Omega_2} \left(\epsilon P(y)^{s_0-\beta_0-1} \ln(P(y)/\epsilon) + P(y)^{s_0-\beta_0} \right) dy^{(2)} \\ &\leq O(\epsilon^\kappa) \int_{\Omega_2} \left(\epsilon |y^{(2)}|^{2(s_0-\beta_0-1)} \ln(P(y)/\epsilon) + |y^{(2)}|^{2(s_0-\beta_0)} \right) dy^{(2)} \\ &= O(A^{-2\kappa}). \end{aligned} \quad (10.7)$$

If $\kappa = 0$, we get from (10.4), (10.5), and (7.8)

$$\begin{aligned} |f_\epsilon^{(2)}(x_\epsilon) - f^{(2)}(x_\epsilon)| &\leq \int_{\Omega_2} \epsilon P(y)^{s_0-\beta_0-1} \ln(P(y)/\epsilon) dy^{(2)} \\ &\leq \epsilon \int_{\Omega_2} |y^{(2)}|^{2(s_0-\beta_0-1)} \ln(P(y)/\epsilon) dy^{(2)} = O(A^{-2} \ln A). \end{aligned} \quad (10.8)$$

As $A \gg 1$ can be arbitrarily large, combining (10.7) and (10.8) with (8.9), (8.35) proves (10.1). Here we use that the integral on the right in (10.1) is absolutely convergent (see Sect. 11).

10.3 Proof of Lemma 10.2

Recall that $\kappa = 0$. By (6.3) and (10.6), for any $c_1 > 0$ we can find $A_0 \gg 1$ sufficiently large so that $P(y) > c|y^{(2)}|^2$ for all $A \geq A_0$, $\epsilon > 0$ sufficiently small, $|\check{x}| \leq c_1$, and $y \in \mathcal{T}_{x_\epsilon}$ as long as $y^{(2)} \in \Omega_2$. It follows from (7.8) that the integral in (10.4) admits a uniform (i.e., independent of $\epsilon > 0$ sufficiently small, $A \gg 1$ sufficiently large, and $\check{x}$ confined to a bounded set) integrable bound:

$$\begin{aligned} \int_{|y^{(2)}| \leq O(1)} P(y)^{s_0-\beta_0+c} dy^{(2)} &\leq c' \int_{|y^{(2)}| \leq O(1)} |y^{(2)}|^{2(s_0-\beta_0+c)} dy^{(2)} \\ &\leq c' \int_0^{O(1)} r^{2(s_0-\beta_0+c)} r^{N-1} dr < \infty, \quad \beta_0 - s_0 = N/2. \end{aligned} \quad (10.9)$$

In (10.9), the constant c in the exponent is the same as the one in (7.8). Therefore, we can compute the limit of $f^{(2)}(x_\epsilon)$ as $\epsilon \rightarrow 0$ by taking the pointwise limit of the

integrand in (10.4). This limit is independent of $A \gg 1$. Shrinking $\mathcal{V}$ if necessary, by (6.3) we can ensure that $P(y) > 0$ for any $y \in \mathcal{T}_{x_0}$, $y \neq 0$. Hence

$$\lim_{\epsilon \rightarrow 0} f^{(2)}(x_\epsilon) = \int_{\mathcal{T}_{x_0}} (\mathcal{B}g)(y)w(x_0, y)dy. \quad (10.10)$$

Thus, the limit is independent of $\check{x}$.

A slightly more general argument holds as well. Let $x = (x_1, 0, \dots, 0)^T \in \mathcal{U}$ be a point with $x_1 > 0$ sufficiently small, and let $y \in \mathcal{T}_x$ be arbitrary. It follows from (6.3) and $\partial Y_1/\partial x_1 > 0$ (see (11.8)) that $P(Y(y^{(2)}, x)) \geq c(x_1 + |y^{(2)}|^2)$ for some $c > 0$. This is easy to understand geometrically. If $x_1 > 0$, i.e. x is on the interior side of $\mathcal{S}$, no curve $\mathcal{S}_y$, $y \in \mathcal{V}$, is tangent to $\mathcal{S}$. Hence the curve $\mathcal{T}_x$ does not intersect $\Gamma = \mathcal{T}_{\mathcal{S}}$, and $P(Y(y^{(2)}, x))$ is bounded away from zero. For any such x , we still have the same lower bound $P(Y(y^{(2)}, x)) \geq c|y^{(2)}|^2$. In view of (4.21) and (10.9), we can use dominated convergence to conclude

$$\int_{\mathcal{T}_{x_0}} (\mathcal{B}g)(y)w(x_0, y)dy = \lim_{\substack{x=(x_1, x_1^\perp=0) \\ x_1 \rightarrow 0^+}} \int_{y \in \mathcal{T}_x} (\mathcal{B}g)(y)w(x, y)dy = (\mathcal{R}^* \mathcal{B}g)(x_0^{\text{int}}). \quad (10.11)$$

Comparing (10.10) with (10.11) proves (10.2).

11 Computation of the DTB: End of the Proof of Theorem 4.7

In this section we evaluate the right side of (10.1) and show that it equals to (4.22) if $\kappa > 0$ and to (4.23) if $\kappa = 0$.

The right side of (10.1) simplifies to the expression

$$\frac{2^{N/2}}{|\det Q|^{1/2}} \int_{\mathbb{R}^N} \Upsilon(h + |v|^2) dv = \frac{2^{N/2} |S^{N-1}|}{|\det Q|^{1/2}} \int_0^\infty \Upsilon(h + q^2) q^{N-1} dq, \\ h := (\partial Y_1/\partial x_1)\check{x}_1, \quad (11.1)$$

where $|S^{N-1}|$ is the area of the unit sphere in $\mathbb{R}^N$. Set (see (C.3))

$$J(h) := \frac{1}{2} \int \Upsilon(h+q) q_+^{(N-2)/2} dq = \frac{e(-N/2)\Gamma(N/2)}{2} \mathcal{F}_{1d}^{-1}(\tilde{\Upsilon}(\lambda)(\lambda - i0)^{-N/2}). \quad (11.2)$$

As is shown in (C.1)-(C.7), the leading singular term of g (cf. (4.15)):

$$\mathcal{F}_{1d}^{-1}(\tilde{v}^+(y)\lambda_+^{-(s_0+1)} + \tilde{v}^-(y)\lambda_-^{-(s_0+1)})(P(y)) \quad (11.3)$$

is indeed of the form (8.3) (where $a^\pm(y)$ are linear combinations of $\tilde{v}^\pm(y)$). Hence $\tilde{\mathcal{A}}(\lambda) = \tilde{v}^+(y)\lambda_+^{-(s_0+1)} + \tilde{v}^-(y)\lambda_-^{-(s_0+1)}$, and (8.6), (11.2) yield

$$J(h) = \frac{\Gamma(N/2)}{2} \mathcal{F}_{1d}^{-1}(\tilde{\varphi}(\lambda\Theta_0)\mu(\lambda))(h), \quad (11.4)$$

$$\mu(\lambda) := \tilde{B}_0(y_0, \Theta_0)\tilde{v}^+(y_0)e(-N/2)\lambda_+^{\kappa-1} + \tilde{B}_0(y_0, -\Theta_0)\tilde{v}^-(y_0)e(N/2)\lambda_-^{\kappa-1}.$$

The function $J(h)$ is identical to the one introduced in [19, Eq. (4.6)] if we replace n in the latter with $N+1$. See [19, Sect. 4.5] for additional information about this function. In particular, $\Upsilon(p) = O(|p|^{-(\beta_0-s_0)})$, $p \rightarrow \infty$ (this follows from (8.6)), hence the integral in (11.2) is absolutely convergent if $\kappa > 0$. Also, $\mu(\lambda)$ is the product of three distributions $\tilde{b}(\lambda)\tilde{\mathcal{A}}(\lambda)(\lambda-i0)^{-N/2}$, which is well-defined as a locally integrable function if $\kappa > 0$.

Combine with the factor $w(x_0, y_0)$ in (10.1) and the factor in front of the integral in (11.1) and compute the inverse Fourier transform (cf. (C.3))

$$\text{DTB}(h) = C_1 \int \hat{\varphi}(\Theta_0, h-p) \left(c_1^+(p-i0)^{-\kappa} + c_1^-(p+i0)^{-\kappa} \right) dp, \quad \kappa > 0, \quad (11.5)$$

$$C_1 = (2\pi)^{N/2} w(x_0, y_0) |\det Q|^{-1/2}, \quad c_1^\pm = \frac{\Gamma(\kappa)}{2\pi} \tilde{B}_0(y_0, \pm\Theta_0) \tilde{v}^\pm(y_0) e(\mp(\beta_0-s_0)).$$

If $\kappa = 0$, a more careful analysis of $J(h)$ is required (see Sect. 4.6 of [19]). Similarly to (C.16), (C.17), condition (4.20) implies that $\Upsilon(p) \equiv 0$, $p > c$, for some $c > 0$, hence the integral in (11.2) is still absolutely convergent. Straightforward multiplication of the distributions to obtain $\mu(\lambda)$ no longer works, because $\mu(\lambda)$ is not a locally integrable function if $\kappa = 0$. Fortunately, in this case $\mu(\lambda)$ is computed in [19, Eq. (4.45)] (see [19, Eqs. (4.45)–(4.47)]). Observe that condition (4.20) in this paper is equivalent to [19, condition (4.6)]. At first glance the two conditions differ by a sign, but s_0 in [19] corresponds to s_0+1 here, which eliminates the discrepancy. Then $\mu(\lambda) = \tilde{B}_0(y_0, \Theta_0)\tilde{v}^+(y_0)e(-N/2)(\lambda-i0)^{-1}$, and (11.5) becomes

$$\text{DTB}(h) = C_1 c_1 \int \hat{\varphi}(\Theta_0, h-p) p_-^0 dp = C_1 c_1 \int_{-\infty}^0 \hat{\varphi}(\Theta_0, h-p) dp, \quad \kappa = 0, \quad (11.6)$$

$$c_1 := i \tilde{B}_0(y_0, \Theta_0) \tilde{v}^+(y_0) e(-(\beta_0-s_0)),$$

where C_1 is the same as in (11.5).

Let us now compute $|\det Q|^{1/2}$. The following lemma is proven in Appendix 1.

Lemma 11.1 *One has*

$$|\det Q|^{1/2} = \frac{(\partial\Phi_1/\partial y_1)^{N/2}}{|\det \Delta H_S|^{1/2}} \left| \det \frac{\partial^2 \Phi_1}{\partial x^{(2)} \partial y^{(2)}} \right| \quad (11.7)$$

and

$$\partial Y_1 / \partial x_1 = (\partial\Phi_1/\partial y_1)^{-1}, \quad \partial Y_1 / \partial x_1, \partial\Phi_1/\partial y_1 > 0, \quad \partial Y_1 / \partial x^\perp = 0, \quad (11.8)$$

where $y = Y(y^{(2)}, x)$ is the function constructed in the proof of Lemma 3.7.

Substituting (11.7) into (11.5) and comparing (11.5), (11.6) with (4.22), (4.23), respectively, we finish the proof of Theorem 4.7.

12 Proof of Theorem 5.4

From (C.13) in the proof of Lemma 7.2 and (4.1) it follows that

$$\begin{aligned} (\mathcal{R}^* \mathcal{B}g)(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}^N} J(y, \lambda) e^{-i\lambda P(y)} w(x, y) (\det G^T(y^{(2)}, x))^{1/2} dy^{(2)} d\lambda, \\ J(y, \lambda) &= J^{(1)}(y, \lambda) + J^{(2)}(y, \lambda), \quad y = Y(y^{(2)}, x) \in \mathcal{T}_x, \\ J^{(1)}(y, \lambda) &= \tilde{B}_0(y, dP(y)) \tilde{v}^+(y) \lambda_+^{\beta_0 - s_0 - 1} + \tilde{B}_0(y, -dP(y)) \tilde{v}^-(y) \lambda_-^{\beta_0 - s_0 - 1}, \\ J^{(2)} &\in S^{c-1}(\mathcal{V} \times \mathbb{R}), \quad c = \max(\beta_0 - s_0 - 1, \beta_1 - s_0, \beta_0 - s_1). \end{aligned} \quad (12.1)$$

By construction, $P(Y_0(y^{(2)})) = \Theta_0 \cdot (Y_0(y^{(2)}) - Z(y^{(2)}))$. By (6.3),

$$\partial_{y^{(2)}} P(Y_0(y^{(2)}))|_{y^{(2)}=0} = 0, \quad \frac{\partial^2 P(Y_0(y^{(2)}))}{(\partial y^{(2)})^2} \Big|_{y^{(2)}=0} = -Q, \quad \det Q \neq 0. \quad (12.2)$$

Therefore the stationary point $y_*^{(2)}(x)$ of the phase $P(Y(y^{(2)}, x))$ is a smooth function of x in a neighborhood of $x = x_0$. Set $x = x_\epsilon = x_0 + \epsilon \check{x}$.

By (6.17), (6.18), and (12.2), application of the stationary phase method to the integral with respect to $y^{(2)}$ in (12.1) yields:

$$\begin{aligned} W_\epsilon(\check{x}, \lambda) &:= \int_{\mathbb{R}^N} J(y, \lambda) e^{-i\lambda P(y)} w(x_\epsilon, y) (\det G^T(y^{(2)}, x_\epsilon))^{1/2} dy^{(2)} \\ &= (\tilde{R}_\epsilon^{(1)}(\check{x}, \lambda) + \tilde{R}_\epsilon^{(2)}(\check{x}, \lambda)) e^{-i\lambda \epsilon(h + O(\epsilon))}, \quad y = Y(y^{(2)}, x_\epsilon), \\ \tilde{R}_\epsilon^{(1)}(\check{x}, \lambda) &= (2\pi)^{\frac{N}{2}} \left(\frac{w(x_0, y_0)}{|\det Q|^{1/2}} + O(\epsilon) \right) \left((\tilde{B}_0(y_0, \Theta_0) \tilde{v}^+(y_0) + O(\epsilon)) e \left(-\frac{N}{2} \right) \lambda_+^{\kappa-1} \right. \\ &\quad \left. + (\tilde{B}_0(y_0, -\Theta_0) \tilde{v}^-(y_0) + O(\epsilon)) e \left(\frac{N}{2} \right) \lambda_-^{\kappa-1} \right) \\ \tilde{R}_\epsilon^{(2)} &\in S^{\kappa-1-c}(U \times \mathbb{R}), \quad h = (\partial Y_1 / \partial x_1) \check{x}_1, \quad \kappa > 0, \end{aligned} \quad (12.3)$$

for some $c > 0$ and any open, bounded set $U \subset \mathbb{R}^n$. Here we have used also that Q is negative definite, and $\det G^T = 1$.

The function $\tilde{R}_\epsilon^{(1)}(\check{x}, \lambda)$ is the leading (as $\lambda \rightarrow \infty$) term of $W_\epsilon(\check{x}, \lambda)$. It is obtained by replacing $J(y, \lambda)$ with $J^{(1)}(y, \lambda)$ (cf. (12.1)) in (12.3) and retaining the top order terms after the stationary phase method is applied. Thus, all the $O(\epsilon)$ terms in the formula for $\tilde{R}_\epsilon^{(1)}(\check{x}, \lambda)$ are independent of λ , they are smooth functions of $\check{x}$, and remain $O(\epsilon)$ when differentiated any number of times with respect to $\check{x}$. All the remaining lower order (as $\lambda \rightarrow \infty$) terms are absorbed into $\tilde{R}_\epsilon^{(2)}(\check{x}, \lambda)$.

The expression $\epsilon(h + O(\epsilon))$ in the exponent arises by evaluating $P(y)$ at the stationary point:

$$\begin{aligned} P(Y(y_*^{(2)}(x_\epsilon), x_\epsilon)) &= P(Y(O(\epsilon), \epsilon\check{x})) = Y_1(O(\epsilon), \epsilon\check{x}) + O(\epsilon^2) \\ &= \epsilon[(\partial Y_1/\partial x_1)\check{x}_1 + O(\epsilon)], \end{aligned} \quad (12.4)$$

where we have used that $x_0 = 0$, $y_0 = 0$, $\partial_{y^{(2)}} Y_1 = 0$ (cf. (3.17)), $P(y) = y_1 - \psi(y^\perp)$, $\psi(0) = 0$, and $\psi'(0) = 0$, and (8.35).

Now we show that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \epsilon^\kappa (\mathcal{R}^* \mathcal{B}g)(x_\epsilon) \\ = C_1 \mathcal{F}_{1d}^{-1}(\mu(\lambda))(h) = C_1 (c_1^+(h - i0)^{-\kappa} + c_1^-(h + i0)^{-\kappa}), \quad \kappa > 0, \end{aligned} \quad (12.5)$$

where $\mu(\lambda)$ is the same as in (11.4), and C_1 and $c_1^\pm$ are the same as in (11.5). The limit in (12.5) is understood in the sense of distributions with test functions $\omega \in C_0^\infty(\mathbb{R}^n)$ (cf. [16] and (5.6)).

When $a = -\kappa < 0$, $m = 0$, and it is not necessary to require derivatives in (5.6). Thus, the limit in (12.5) is understood in the sense of taking the limit as $\epsilon \rightarrow 0$ on both sides of the following equality:

$$\int_{\mathbb{R}^n} \epsilon^\kappa (\mathcal{R}^* \mathcal{B}g)(x_0 + \epsilon\check{x}) \omega(\check{x}) d\check{x} = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}^n} \epsilon^{\kappa-1} W_\epsilon(\check{x}, \sigma/\epsilon) \omega(\check{x}) d\check{x} d\sigma. \quad (12.6)$$

The right-hand side of (12.5) follows from (12.3), the dominated convergence theorem, and (C.3). The dominated convergence theorem can be applied because

- (1) The function $\tilde{R}_\epsilon(\check{x}, \lambda)$, where $\tilde{R}_\epsilon := \tilde{R}_\epsilon^{(1)} + \tilde{R}_\epsilon^{(2)}$, is absolutely integrable at $\lambda = 0$ since $\kappa > 0$; and
- (2) The integrand in (12.6) is rapidly decreasing as $\sigma \rightarrow \infty$ uniformly in ϵ . This follows from integration by parts with respect to $\check{x}_1$ on the right in (12.6) and using that $\tilde{R}_\epsilon \in S^{\kappa-1}(U \times \mathbb{R})$ (cf. (12.3)), $\omega \in C_0^\infty(\mathbb{R}^n)$, and $\partial Y_1/\partial x_1 \neq 0$ (cf. (11.8)). Clearly, the constants c_m and c_{m_1, m_2} that control the derivatives of $\tilde{R}_\epsilon$ in (4.7) can be selected independently of ϵ for all $\epsilon > 0$ sufficiently small.

If $\kappa = 0$, the function $\mu(\lambda)$ is no longer integrable at the origin, and $|m| = 1$ in (5.6). Hence we use test functions of the form $\partial_{\check{x}_j} \omega(\check{x})$, $1 \leq j \leq n$. This makes the same argument as in the case $\kappa > 0$ to work, but the price to pay is that the CTB is determined up to a constant. See the paragraph following Theorem 4.6 of [19] for a similar phenomenon. Using condition (4.20) in (12.3) implies

$$\lim_{\epsilon \rightarrow 0} (\mathcal{R}^* \mathcal{B}g)(x_\epsilon) - C_1 c_1 (-1/2) \operatorname{sgn}(\check{x}_1) = \text{const}, \quad \kappa = 0, \quad (12.7)$$

where c_1 is the same as in (11.6).

Comparing (11.5) and (11.6) with (12.5) and (12.7), respectively, we see that the DTB is the convolution of the CTB with the scaled classical Radon transform of the

interpolating kernel. The difference between p_-^0 in (11.6) and $(-1/2)\text{sgn}(p)$ in (12.7) is due to the nonuniqueness (up to a constant).

Appendix A. Proof of Lemma 3.5

We begin by constructing an orthogonal matrix $\check{U}$ such that the intermediate coordinates $\check{y} = \check{U}^T(\tilde{y} - \tilde{y}_0)$ and the intermediate function $\check{\Phi}(t, \check{y}) = \tilde{\Phi}(t, \check{U}\check{y} + \tilde{y}_0)$ satisfy (3.10). Here $\check{y} = (z^{(1)}, y^{(2)})^T$. The final coordinates y and the intermediate coordinates $\check{y}$ will have the same $y^{(2)}$ component, this is why we wrote $y^{(2)}$ in the definition of $\check{y}$.

Let $V_1 \Sigma V_2^T$ be the SVD of the Jacobian matrix $\tilde{\Phi}_{\tilde{y}}^{(1)}$. To remind the reader, $\tilde{\Phi}_{\tilde{y}}^{(1)}$ stands for the matrix of partial derivatives $\partial_{\tilde{y}} \tilde{\Phi}^{(1)}(x^{(2)}, \tilde{y})$ evaluated at $(x_0^{(2)}, \tilde{y}_0)$. Here $V_1 \in O(n - N)$ and $V_2 \in O(n)$ are orthogonal matrices, and Σ is a rectangular $(n - N) \times n$ matrix with $\Sigma_{ij} = 0$, $i \neq j$, and $\Sigma_{ii} > 0$, $1 \leq i \leq n - N$. The latter property follows from Assumption 3.1(G2) and $\tilde{\Phi}_{x^{(2)}}^{(1)} = 0$, which yield that $\text{rank} \tilde{\Phi}_{\tilde{y}}^{(1)} = n - N$. Then we can take $\check{U} = V_2$, so $\check{y} = V_2^T(\tilde{y} - \tilde{y}_0)$. Indeed,

$$\frac{\partial \check{\Phi}^{(1)}}{\partial y^{(2)}} = \frac{\partial \tilde{\Phi}^{(1)}}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial y^{(2)}} = V_1 \Sigma V_2^T V_2^{(2)} = V_1 \Sigma \begin{pmatrix} 0 \\ I_N \end{pmatrix} = 0. \quad (\text{A.1})$$

Here $V_2^{(2)}$ is the $n \times N$ matrix consisting of the last N columns of V_2 . Likewise,

$$\frac{\partial \check{\Phi}^{(1)}}{\partial z^{(1)}} = \frac{\partial \tilde{\Phi}^{(1)}}{\partial \tilde{y}} \frac{\partial \tilde{y}}{\partial z^{(1)}} = V_1 \Sigma V_2^T V_2^{(1)} = V_1 \Sigma \begin{pmatrix} I_{n-N} \\ 0 \end{pmatrix}, \quad \det \frac{\partial \check{\Phi}^{(1)}}{\partial z^{(1)}} \neq 0. \quad (\text{A.2})$$

The final coordinates y and the final orthogonal matrix U can be found as follows. As was already mentioned, we keep the coordinates $y^{(2)}$ the same and rotate the $z^{(1)}$ coordinates: $z^{(1)} \rightarrow y^{(1)}$. Hence (A.1), (A.2) still hold with $\check{\Phi}$ and $z^{(1)}$ replaced by Φ and $y^{(1)}$, respectively, and the new y coordinates satisfy (3.10).

By (3.5), the rotation $z^{(1)} \rightarrow y^{(1)}$ should be selected so that

$$\partial \Phi_1 / \partial y_j = (\partial \check{\Phi}_1 / \partial z^{(1)}) (\partial z^{(1)} / \partial y_j) = 0, \quad j = 2, \dots, n - N. \quad (\text{A.3})$$

If $V \in O(n - N)$ is such that $z^{(1)} = Vy^{(1)}$, then (A.3) implies that the second through the last columns of V form an orthonormal basis of the subspace of $\mathbb{R}^{n-N}$, which consists of vectors orthogonal to $\partial \check{\Phi}_1 / \partial z^{(1)}$. It is clear that such a basis can be found. Then the matrix U becomes

$$U = \check{U} \begin{pmatrix} V & 0 \\ 0 & I_N \end{pmatrix}. \quad (\text{A.4})$$

Our construction ensures that all components of the vector $\partial_y \Phi_1$, except, possibly, the first one, are zero. By (A.2), the first component is not zero. Multiplying $\Psi(y)$ by a constant, we can make sure that $\partial_{y_1}(\Psi \circ \Phi) = 1$.

Appendix B. Behavior of $\mathcal{R}f$ Near Γ

Suppose $f \in \mathcal{E}'(\mathcal{U})$ is given by

$$f(x) = \frac{1}{2\pi} \int \tilde{f}(x, \lambda) e^{-i\lambda\Psi(x)} d\lambda, \quad (\text{B.1})$$

where Ψ is the same as in Sects. 3, 6, and $\tilde{f}$ satisfies

$$\begin{aligned} \tilde{f}(x, \lambda) &= \tilde{f}^+(x) \lambda_+^{-(s_0+1)+\frac{N}{2}} + \tilde{f}^-(x) \lambda_-^{-(s_0+1)+\frac{N}{2}} + \tilde{R}(x, \lambda), \forall x \in \mathcal{U}, |\lambda| \geq 1; \\ \tilde{f}(x, \lambda) &\equiv 0 \quad \forall x \in \mathcal{U} \setminus K \text{ for some compact } K \subset \mathcal{U}; \\ |\partial_x^m \tilde{f}(x, \lambda)| &\leq c_m |\lambda|^a, \quad a > -1, \quad \forall x \in \mathcal{U}, 0 < |\lambda| \leq 1, m \in \mathbb{N}_0^n; \\ \tilde{R} &\in S^{-(s_1+1)+\frac{N}{2}}(\mathcal{U} \times \mathbb{R}), \quad \tilde{f}^\pm \in C_0^\infty(\mathcal{U}), \quad (N/2) - 1 < s_0 < s_1, \end{aligned} \quad (\text{B.2})$$

for some $a, c_m, s_0, s_1, \tilde{R}$, and $\tilde{f}^\pm$. If $g = \mathcal{R}f$ for a sufficiently regular f , then g should have more regularity ($s_0 > (N/2) - 1$) than in the general case (4.15) ($s_0 > 0$).

From (2.3), after changing variables and the defining function ($t \rightarrow x^{(2)}, \tilde{y} \rightarrow y, \tilde{\Phi} \rightarrow \Phi$), the GRT of f is given by

$$\begin{aligned} \mathcal{R}f(y) &= \int_{\mathbb{R}^N} f(x) b(x, y) (\det G^{\mathcal{S}}(x^{(2)}, y))^{1/2} dx^{(2)} \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}^N} \tilde{f}(x, \lambda) e^{-i\lambda\Psi(x)} b(x, y) (\det G^{\mathcal{S}}(x^{(2)}, y))^{1/2} dx^{(2)} d\lambda, \end{aligned} \quad (\text{B.3})$$

where $x = \Phi(x^{(2)}, y)$.

Consider the second equation for $\mathcal{T}_{\mathcal{S}}$ in (6.4) and solve it for $x^{(2)}$. Since $\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}} \neq 0$, the solution $x_*^{(2)} = x_*^{(2)}(y)$ is a smooth function. The function $x_*^{(2)}(y)$ here is different from $x^{(2)}(y^\perp)$ in the paragraph following (3.16), because now we solve only the second of the two equations that define $\mathcal{T}_{\mathcal{S}}$. The asymptotics as $\lambda \rightarrow \infty$ of the integral with respect to $x^{(2)}$ in (B.3) is computed with the help of the stationary phase method [33, Chapter VIII, Eqs. (2.14)–(2.20)]

$$\begin{aligned} &\int \tilde{f}(x, \lambda) e^{-i\lambda\Psi(x)} b(x, y) (\det G^{\mathcal{S}}(x^{(2)}, y))^{1/2} dx^{(2)} \\ &= \left(\tilde{f}(x_*, \lambda) b(x_*, y) \left| \frac{\det G^{\mathcal{S}}(x_*^{(2)}, y)}{\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}(x_*^{(2)}, y)} \right|^{1/2} \left(\frac{2\pi}{|\lambda|} \right)^{N/2} + \tilde{R}(y, \lambda) \right) \\ &\quad \times e^{-i\frac{\pi}{4} \operatorname{sgn}(\lambda(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}(x_*^{(2)}, y))} e^{-i\lambda\Psi(x_*)}, \quad |\lambda| \geq 1, \quad \tilde{R} \in S^{-(s_0+2)}(\mathcal{V} \times \mathbb{R}), \end{aligned} \quad (\text{B.4})$$

for some $\tilde{R}$. Here $x_* = \Phi(x_*^{(2)}(y), y)$, and $\operatorname{sgn} M$ for a symmetric matrix M denotes the signature of M , i.e. the number of positive eigenvalues of M minus the number of negative eigenvalues.

Introduce the function

$$P_1(y) := (\Psi \circ \Phi)(x_*^{(2)}(y), y). \quad (\text{B.5})$$

Then $\mathcal{R}f$ can be written as

$$\mathcal{R}f(y) = \frac{1}{2\pi} \int \tilde{v}(y, \lambda) e^{-i\lambda P_1(y)} d\lambda, \quad (\text{B.6})$$

and, with the same a, s_0, s_1 as in (B.2) and some $c_m, \tilde{R}$,

$$\begin{aligned} \tilde{v}(y, \lambda) &= \tilde{v}^+(y) \lambda_+^{-(s_0+1)} + \tilde{v}^-(y) \lambda_-^{-(s_0+1)} + \tilde{R}(y, \lambda), \quad |\lambda| \geq 1; \\ |\partial_y^m \tilde{v}(y, \lambda)| &\leq c_m |\lambda|^a, \quad \forall y \in \mathcal{V}, 0 < |\lambda| \leq 1, m \in \mathbb{N}_0^n; \\ \tilde{R} &\in S^{-\min(s_1+1, s_0+2)}(\mathcal{V} \times \mathbb{R}); \end{aligned} \quad (\text{B.7})$$

$$\tilde{v}^\pm(y) = (2\pi)^{N/2} \tilde{f}^\pm(x_*) b(x_*, y) \left| \frac{\det G^S(x_*^{(2)}, y)}{\det(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}(x_*^{(2)}, y)} \right|^{1/2} \times e^{\pm i \frac{\pi}{4} N} \in C^\infty(\mathcal{V});$$

where $x_* = \Phi(x_*^{(2)}(y), y)$, and we have used that $(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}$ is negative definite.

By construction, $P_1(y) = 0$ is another equation for $\mathcal{T}_S$. Since $(\Psi \circ \Phi)_{x^{(2)}} = 0$, equation (6.6) does not determine $x_*^{(2)}$. Therefore, to first order, $x_*^{(2)}$ is determined by solving (6.7):

$$x_*^{(2)}(y) = -(\Psi \circ \Phi)_{x^{(2)}x^{(2)}}^{-1}(\Psi \circ \Phi)_{x^{(2)}y} y + O(|y|^2), \quad (\text{B.8})$$

and

$$P_1(y) = d_y(\Psi \circ \Phi)y + O(|y|^2) = \Theta_0 \cdot y + O(|y|^2). \quad (\text{B.9})$$

Remark B.1 We are now in a position to discuss the implications of Assumption 4.5(g4). Suppose $g = \mathcal{R}f$ and $s_0 \in \mathbb{N}$. From (4.17) and (B.7),

$$\frac{\tilde{v}^-(y)}{\tilde{v}^+(y)} = (-1)^{s_0+1}, \quad \frac{\tilde{f}^-(x)}{\tilde{f}^+(x)} = e(-2a), \quad a := s_0 - (N/2) + 1, \quad \forall y \in \Gamma, x \in \mathcal{S}. \quad (\text{B.10})$$

Here we have used that $x_*(y) \in \mathcal{S}$ if $y \in \Gamma$. Recall that the function $e(a)$ is defined in (4.18).

Suppose first that N is odd, i.e., $a \notin \mathbb{N}$. Substituting (B.10) into (B.2) gives to leading order:

$$\begin{aligned} \tilde{f}(x, \lambda) &\sim \tilde{f}^+(x) \lambda_+^{-a} + \tilde{f}^-(x) \lambda_-^{-a} = \tilde{f}^+(x) (\lambda_+^{-a} + e(-2a) \lambda_-^{-a}) \\ &= \tilde{f}^+(x) (\lambda + i0)^{-a}, \quad \lambda \rightarrow \infty, \quad \forall x \in \mathcal{S}. \end{aligned} \quad (\text{B.11})$$

Using (B.1) and computing the inverse Fourier transform, we approximate f to leading order:

$$f(x + \Delta x) \sim c \tilde{f}^+(x) (\mathrm{d}\Psi(x) \Delta x)_+^{a-1}, \quad |\Delta x| \rightarrow 0, \quad \forall x \in \mathcal{S}, \quad (\text{B.12})$$

for some $c \neq 0$. Thus, if N is odd, Assumption 4.5(g4) means that, to leading order, the nonsmooth part of f is supported on the positive side of $\mathcal{S}$.

Suppose next that N is even, i.e. $a \in \mathbb{N}$. Substituting (B.10) into (B.2) gives:

$$\begin{aligned} \tilde{f}(x, \lambda) &\sim \tilde{f}^+(x) \lambda^{-a}, \quad \lambda \rightarrow \infty, \quad \forall x \in \mathcal{S}, \\ f(x + \Delta x) &\sim c \tilde{f}^+(x) h^{a-1} \operatorname{sgn} h, \quad h := \mathrm{d}\Psi(x) \Delta x, \quad |\Delta x| \rightarrow 0, \quad \forall x \in \mathcal{S}, \end{aligned} \quad (\text{B.13})$$

for some $c \neq 0$. Thus, if N is even, Assumption 4.5(g4) means that, to leading order, the nonsmooth part of f is symmetric about $\mathcal{S}$: $f(x + \Delta x) \sim f(x - \Delta x)$ if a is even, and $f(x + \Delta x) \sim -f(x - \Delta x)$ if a is odd.

Remark B.2 The behavior of f near $\mathcal{S}$ can be obtained in the same way even if $s_0 \notin \mathbb{N}$, and Assumption 4.5(g4) does not apply. Taking the inverse Fourier transform of the first (asymptotic) equality in (B.11) using [9, p. 360, Eqs. 25, 26] shows that

$$\begin{aligned} f(x + \Delta x) &\sim c_+ \tilde{f}^+(x) (h - i0)^{a-1} + c_- \tilde{f}^-(x) (h + i0)^{a-1}, \\ h &:= \mathrm{d}\Psi(x) \Delta x, \quad |\Delta x| \rightarrow 0, \quad \forall x \in \mathcal{S}, \quad a \notin \mathbb{N}, \end{aligned} \quad (\text{B.14})$$

for some constants $c_{\pm} \neq 0$. If $a \in \mathbb{N}$, then the leading singularity of f may contain logarithmic terms [9, Chapter II, Sect. 2.4, Eqs. (14) and (20)]. Computing the corresponding explicit expressions is fairly straightforward and is outside the scope of the paper.

Appendix C. Proofs of Lemmas 7.1–7.4

C.1. Proof of Lemma 7.1

The following expression for g (modulo a $C^\infty(\mathcal{V})$ function) is obtained directly from (4.12), (4.15):

$$\begin{aligned} g(y) - G(y, P(y)) &\in C^\infty(\mathcal{V}), \\ G(y, p) &:= \mathcal{F}_{1d}^{-1} \left(\tilde{v}^+(y) \lambda_+^{-(s_0+1)} + \tilde{v}^-(y) \lambda_-^{-(s_0+1)} + \tilde{R}(y, \lambda) \right) (p), \end{aligned} \quad (\text{C.1})$$

where $\tilde{R} \in S^{-(s_1+1)}(\mathcal{V} \times \mathbb{R})$, and $\mathcal{F}_{1d}^{-1}$ is the one-dimensional inverse Fourier transform acting with respect to λ . The inverse transforms $\mathcal{F}_{1d}^{-1}(\lambda_{\pm}^{-(s_0+1)})$ are understood in the sense of distributions [9, Chapter II, Sect. 2.3]. By the properties of $\tilde{R}$, we get by computing the inverse Fourier transform if $s_0 \notin \mathbb{N}$:

$$G(y, p) = \tilde{v}^+(y) \Psi_{-s_0}^+(p) + \tilde{v}^-(y) \Psi_{-s_0}^-(p) + R(y, p), \quad \forall y \in \mathcal{V}, \quad p \in \mathbb{R}. \quad (\text{C.2})$$

Here [9, p. 360]

$$\Psi_a^\pm(p) = \mathcal{F}_{1d}^{-1}(\lambda_{\pm}^{a-1})(p) = \frac{\Gamma(a)}{2\pi} e(\mp a)(p \mp i0)^{-a}, \quad a \neq 0, -1, -2, \dots, \quad (\text{C.3})$$

and $R(y, p) = \mathcal{F}_{1d}^{-1}(\tilde{R}(y, \lambda))(p)$. By [1, Theorem 5.12], R satisfies

$$|\partial_y^m \partial_p^l R(y, p)| \leq c_{m,l} \begin{cases} |p|^{s_1-l}, & s_1 < l, \\ 1 + |\log |p||, & s_1 = l, \quad \forall m \in \mathbb{N}_0^n, l \in \mathbb{N}_0, y \in \mathcal{V}, p \neq 0, \\ 1, & s_1 > l, \end{cases} \quad (\text{C.4})$$

for some $c_{m,l} > 0$. Recall that $P(y + p\Theta_0) \equiv p$ for any $y \in \Gamma$ and p such that $y + p\Theta_0 \in \mathcal{V}$. Combining (C.1)–(C.4) gives the leading singular behavior of g :

$$\begin{aligned} g(y + p\Theta_0) &\sim a^+(y)p_+^{s_0} + a^-(y)p_-^{s_0}, \quad p \rightarrow 0, \quad \forall y \in \Gamma, \\ a^\pm(y) &= \frac{\Gamma(-s_0)}{2\pi} (\tilde{v}^+(y)e(\pm s_0) + \tilde{v}^-(y)e(\mp s_0)), \quad s_0 \notin \mathbb{N}. \end{aligned} \quad (\text{C.5})$$

See [2, 28] for a characterization of the singularities of the classical Radon transform $g = \mathcal{R}f$ for more general surfaces $\mathcal{S}$.

If $s_0 \in \mathbb{N}$, condition (4.17) implies $\tilde{v}^+(y)\lambda_+^{-(s_0+1)} + \tilde{v}^-(y)\lambda_-^{-(s_0+1)} \equiv \tilde{v}^+(y)\lambda^{-(s_0+1)}$, $y \in \Gamma$, so [9, p. 360]

$$\begin{aligned} G(y, p) &= \tilde{v}^+(y)\Psi_{-s_0}(p) + R(y, p), \quad \forall y \in \mathcal{V}, \\ \Psi_{-s_0}(p) &= \mathcal{F}_{1d}^{-1}(\lambda^{-(s_0+1)})(p) = \frac{1}{2} \frac{(-i)^{s_0+1}}{s_0!} p^{s_0} \text{sgn}(p), \quad s_0 \in \mathbb{N}. \end{aligned} \quad (\text{C.6})$$

An equation of the kind (C.5) still holds:

$$\begin{aligned} g(y + p\Theta_0) &\sim a^+(y)p_+^{s_0} + a^-(y)p_-^{s_0}, \quad p \rightarrow 0, \quad \forall y \in \Gamma, \\ a^\pm(y) &= \frac{\tilde{v}^+(y)}{2s_0!} e(\mp(s_0 + 1)), \quad s_0 \in \mathbb{N}. \end{aligned} \quad (\text{C.7})$$

Combining (C.1)–(C.6) and using that (C.2) and (C.6) can be differentiated proves (7.5).

From the second equation in (C.1), (C.6), and (7.3) we get also

$$\partial_y^m G(y, \cdot) \in \begin{cases} C_*^{s_0}(\mathbb{R}) \quad \forall m \in \mathbb{N}_0^n, \\ C_0^{s_0}(\mathbb{R}) \quad \forall m \in \mathbb{N}_0^n \text{ if } s_0 \in \mathbb{N}. \end{cases} \quad (\text{C.8})$$

Together with the first equation in (C.1) this proves (7.6).

If $\tilde{v}^\pm \equiv 0$, the result follows from the properties of $\tilde{R}(y, \lambda)$ and (7.3), because $s_1 \notin \mathbb{N}$.

C.2. Proof of Lemma 7.2

From (4.3) and (4.12),

$$(\mathcal{B}g)(y) = \frac{1}{(2\pi)^{n+1}} \int_{\mathbb{R}^n} \tilde{B}(y, \eta) \int_{\mathcal{V}} \int_{\mathbb{R}} \tilde{v}(z, \lambda) e^{-i\lambda P(z) + i\eta(z-y)} d\lambda dz d\eta. \quad (\text{C.9})$$

As is standard (see e.g., [33]), set $u = \eta/\lambda$ and consider the phase function

$$W(z, u, y) := P(z) - u(z - y). \quad (\text{C.10})$$

The only critical point (z_0, u_0) and the corresponding Hessian H are given by

$$z_0 = y, \quad u_0 = dP(y), \quad H(y) = \begin{pmatrix} P_{yy}(y) & -I_n \\ -I_n & 0 \end{pmatrix}. \quad (\text{C.11})$$

Clearly, $|\det H(y)| = 1$ and $\text{sgn } H(y) = 0$ for any $y \in \mathcal{V}$. By the stationary phase method (see [33, Chapter VIII, Eqs. (2.14)–(2.20)]) we get using (4.8) and (4.15)

$$\begin{aligned} J(y, \lambda) &:= \frac{|\lambda|^n}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathcal{V}} \tilde{B}(y, \lambda u) \tilde{v}(z, \lambda) e^{-i\lambda(P(z) - P(y) - u(z-y))} dz du \\ &= \tilde{B}(y, \lambda dP(y)) \tilde{v}(y, \lambda) + \tilde{R}(y, \lambda), \\ J &\in S^{\beta_0 - s_0 - 1}(\mathcal{V} \times \mathbb{R}), \quad \tilde{R} \in S^{\beta_0 - s_0 - 2}(\mathcal{V} \times \mathbb{R}). \end{aligned} \quad (\text{C.12})$$

The fact that the u -integration is over an unbounded domain does not affect the result, because integrating by parts with respect to z we obtain a function that decreases rapidly as $|u| \rightarrow \infty$ (i.e., when $|u| > \sup_{y \in \mathcal{V}} |dP(y)|$) and $\lambda \rightarrow \infty$.

Substituting (C.12) into (C.9) and using (4.8), (4.15) leads to

$$\begin{aligned} (\mathcal{B}g)(y) &= \frac{1}{2\pi} \int J(y, \lambda) e^{-i\lambda P(y)} d\lambda \\ &= \mathcal{F}_{1d}^{-1} \left(\tilde{B}_0(y, dP(y)) \tilde{v}^+(y) \lambda_+^{\beta_0 - s_0 - 1} \right. \\ &\quad \left. + \tilde{B}_0(y, -dP(y)) \tilde{v}^-(y) \lambda_-^{\beta_0 - s_0 - 1} + \tilde{R}(y, \lambda) \right) (P(y)), \\ \tilde{R} &\in S^{c-1}(\mathcal{V} \times \mathbb{R}), \quad c := \max(\beta_0 - s_0 - 1, \beta_1 - s_0, \beta_0 - s_1) < \beta_0 - s_0. \end{aligned} \quad (\text{C.13})$$

The extra factor $|\lambda|^n$ in (C.12) cancels because $|\lambda|^n du = d\eta$. Computing the asymptotics of the inverse Fourier transform as $p = P(y) \rightarrow 0$ and using that $\tilde{B}_0(y, \pm dP(y)) \in C_0^\infty(\mathcal{V})$ and $P(y + p\Theta_0) \equiv p$ if $y \in \Gamma$ gives

$$\begin{aligned} (\mathcal{B}g)(y + p\Theta_0) &= c_1^+ \Psi_{\beta_0 - s_0}^+(p) + c_1^- \Psi_{\beta_0 - s_0}^-(p) + R(y + p\Theta_0, p), \\ c_1^\pm &= \tilde{B}_0(y, \pm dP(y)) \tilde{v}^\pm(y), \quad \forall y \in \Gamma. \end{aligned} \quad (\text{C.14})$$

Recall that $\Psi_a^\pm$ are defined in (C.3), and $R(y, p) = \mathcal{F}_{1d}^{-1}(\tilde{R}(y, \lambda))(p)$. By [1, Theorem 5.12] and (C.13), the remainder satisfies

$$|\partial_y^m \partial_p^l R(y, p)| \leq c_{m,l} \begin{cases} |p|^{c-l}, & c < l, \\ 1 + |\log |p||, & c = l, \quad \forall m \in \mathbb{N}_0^n, l \in \mathbb{N}_0, y \in \mathcal{V}, p \neq 0, \\ 1, & c > l, \end{cases} \quad (\text{C.15})$$

for some $c_{m,l} > 0$. The constant c here is the same as in (C.13). The estimate (7.7) follows from (C.14), (C.15).

If $\kappa = 0$, condition (4.20) implies

$$\begin{aligned} \tilde{B}_0(y, dP(y)) \tilde{v}^+(y) \lambda_+^{\beta_0 - s_0 - 1} + \tilde{B}_0(y, -dP(y)) \tilde{v}^-(y) \lambda_-^{\beta_0 - s_0 - 1} \\ = \tilde{B}_0(y, dP(y)) \tilde{v}^+(y) (\lambda - i0)^{\beta_0 - s_0 - 1} \quad \forall y \in \Gamma, \end{aligned} \quad (\text{C.16})$$

and

$$\begin{aligned} (\mathcal{B}g)(y + p\Theta_0) = \tilde{B}_0(y, dP(y)) \tilde{v}^+(y) \begin{cases} \frac{e(-\beta_0 + s_0 + 1)}{\Gamma(-\beta_0 + s_0 + 1)} P_-^{-(\beta_0 - s_0)}, & \beta_0 - s_0 \notin \mathbb{N} \\ 0, & \beta_0 - s_0 \in \mathbb{N} \end{cases} \\ + R(y + p\Theta_0, p) \quad \forall y \in \Gamma, p \neq 0. \end{aligned} \quad (\text{C.17})$$

This proves (7.8).

C.3. Proof of Lemma 7.3

Using that $l > s_0$ and φ has $\lceil \beta_0^+ \rceil$ bounded derivatives and φ is exact to the degree $\lceil \beta_0 \rceil$, we get with any $0 \leq M \leq l$:

$$\begin{aligned} g_\epsilon^{(l)}(y) &= \sum_{|j| \leq \partial/\epsilon} \partial_{y_1}^l \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) \left(g(\hat{y}^j) - \sum_{|m| \leq M-1} \frac{(\hat{y}^j - y)^m}{m!} g^{(m)}(y) \right) \\ &= \epsilon^{-l} \sum_{|j| \leq \partial/\epsilon} (\partial_{v_1}^l \varphi) \left(\frac{y - \hat{y}^j}{\epsilon} \right) \sum_{|m|=M} R_m(\hat{y}^j, y) (\hat{y}^j - y)^m. \end{aligned} \quad (\text{C.18})$$

Here $(\partial_{v_1} \varphi)(\cdot)$ is the derivative of $\varphi(v)$ with respect to v_1 evaluated at the indicated point, and the remainder satisfies

$$\begin{aligned} R_m(\hat{y}^j, y) &= \frac{|m|}{m!} \int_0^1 (1-t)^{|m|-1} g^{(m)}(y + t(\hat{y}^j - y)) dt, \\ |R_m(\hat{y}^j, y)| &\leq \frac{1}{m!} \max_{|m'|=|m|, v \in \text{supp } \varphi} |g^{(m')}(y + \epsilon v)|. \end{aligned} \quad (\text{C.19})$$

To prove the top case in (7.10) select $\varkappa_1 > 0$ so that $|P(y)| \geq \varkappa_1 \epsilon$ and $\varphi((y - w)/\epsilon) \neq 0$ implies $|P(w)| \geq \epsilon$. By Lemma 7.1, this ensures that for each $m \in \mathbb{N}_0^n$ there exists $c(m) > 0$ such that

$$\max_{v \in \text{supp } \varphi} |g^{(m)}(y + \epsilon v)| \leq c(m) \begin{cases} |P(y)|^{s_0 - |m|}, & |m| > s_0, \\ 1, & |m| \leq s_0, \end{cases} \quad |P(y)| \geq \varkappa_1 \epsilon. \quad (\text{C.20})$$

Set $M = l$ in (C.18). Then (C.20) together with the bottom line in (C.19) prove the result.

To prove the middle case in (7.10), set $M = \lfloor s_0 \rfloor$ in (C.18). If $s_0 \in \mathbb{N}$, (7.6) and (C.19) imply that $R_m = O(1)$, thereby proving the assertion. If $s_0 \notin \mathbb{N}$, the remainder can be modified as follows

$$\tilde{R}_m(\hat{y}^j, y) = \frac{|m|}{m!} \int_0^1 (1-t)^{|m|-1} [g^{(m)}(y + t(\hat{y}^j - y)) - g^{(m)}(y)] dt = O(\epsilon^{\{s_0\}}). \quad (\text{C.21})$$

Here we have used (7.4) with $r = s_0$. Since $l > M$, we can replace R_m with $\tilde{R}_m$ in (C.18) without changing the equality, and the desired inequality follows.

The bottom case in (7.10) follows by setting $M = l$ in (C.18) and noticing that (7.6) and (C.19) imply $R_m = O(1)$.

If $\tilde{v}^\pm \equiv 0$, the same argument as above applies with s_0 replaced by s_1 . The only change is that there is no need to consider the case $s_1 \in \mathbb{N}$.

C.4. Proof of Lemma 7.4

Since $\varkappa_1 > 0$ is the same as in the proof of Lemma 7.3, $|P(y)| \geq \varkappa_1 \epsilon$ and $\varphi((y - w)/\epsilon) \neq 0$ imply $|P(w)| \geq \epsilon$. Similarly to (C.18), using the properties of φ we obtain

$$\begin{aligned} g_\epsilon^{(l)}(y) &= \sum_{|j| \leq \vartheta/\epsilon} \partial_{y_1}^l \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) \left(\sum_{|m| \leq M-1} \frac{(\hat{y}^j - y)^m}{m!} g^{(m)}(y) \right. \\ &\quad \left. + \sum_{|m|=M} R_m(\hat{y}^j, y) (\hat{y}^j - y)^m \right) \\ &= g^{(l)}(y) + \sum_{|j| \leq \vartheta/\epsilon} \partial_{y_1}^l \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) \sum_{|m|=M} R_m(\hat{y}^j, y) (\hat{y}^j - y)^m, \quad l < M \leq \lceil \beta_0 \rceil. \end{aligned} \quad (\text{C.22})$$

The term $g^{(l)}(y)$ on the right in (C.22) is the only term from the Taylor polynomial that remains after the summation with respect to j . In particular, all the terms corresponding to $l < |m| \leq M - 1$ are converted to zero, because φ is exact to the degree $\lceil \beta_0 \rceil$, and

$$\sum_{|j| \leq \vartheta/\epsilon} \partial_{y_1}^l \varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) (\hat{y}^j - y)^m = \partial_{w_1}^l (w - y)^m|_{w=y} = 0, \quad (\text{C.23})$$

$$\forall m \in \mathbb{N}_0^n : l < |m| \leq M - 1, \quad y \in \mathcal{V}.$$

Using (C.22) with $M = l + 1$ and appealing to (C.19), (C.20) proves (7.11). Indeed, recall that $l \geq \lfloor s_0^- \rfloor$, so $M = l + 1 \geq s_0$. If $s_0 \notin \mathbb{N}$, then $M > s_0$, and the top case in (C.20) applies when estimating R_m , $|m| = M$. If $s_0 \in \mathbb{N}$, then $M = s_0$, and the bottom case in (C.20) applies when estimating R_m , $|m| = M$.

To prove (7.12), we use (C.22) with $M = \lfloor s_0 \rfloor$. If $s_0 \in \mathbb{N}$, then $l < \lfloor s_0 \rfloor = s_0$ (by assumption, $l \leq \lfloor s_0^- \rfloor$), and (C.22), (7.6) prove (7.12).

If $s_0 \notin \mathbb{N}$, we replace R_m with $\tilde{R}_m$ in (C.22) as this was done in the proof of Lemma 7.3. As before, this does not invalidate the equality and extends its applicability to the case $l = M$. Note, however, that if $l = M$, then the term $g^{(l)}(y)$ on the right in (C.22) comes not from the Taylor polynomial, but from the modification of the remainder. The desired assertion follows from (C.21) and the modified (C.22).

If $\tilde{v}^\pm \equiv 0$, the same argument as above applies with s_0 replaced by s_1 . The only change is that there is no need to consider the case $s_1 \in \mathbb{N}$.

Appendix D. Proof of Lemma 8.2

Throughout the proof, c denotes various positive constants that can vary from one place to the next. To simplify notations, in this proof we drop the subscripts from β_0 and s_0 : $\beta = \beta_0$, $s = s_0$. By the choice of y coordinates (see (3.9)) and by (3.18), $y_1 = \Theta_0 \cdot y$ (recall that $|\Theta_0| = 1$).

Using (4.6), (8.3), and that the symbol of $\mathcal{B}$ is homogeneous of degree β we have

$$\epsilon^\beta (\mathcal{B}g_\epsilon)(y) = \sum_{|j| \leq \vartheta/\epsilon} \left(\mathcal{B}\varphi \left(\cdot - (\hat{y}^j/\epsilon) \right) \right) (y/\epsilon) \left[a^+(\hat{y}^j) P_+^{s_0}(\hat{y}^j) + a^-(\hat{y}^j) P_-^{s_0}(\hat{y}^j) \right], \quad (\text{D.1})$$

where

$$(\mathcal{B}\varphi(\cdot - a))(u) := (\mathcal{B}\varphi_1)(u), \quad \varphi_1(u) := \varphi(u - a). \quad (\text{D.2})$$

Also (cf. (8.11)):

$$\epsilon^s \mathcal{A} \left(\Theta_0 \cdot \frac{\hat{y}^j - z}{\epsilon} \right) = [a^+(y_0)(\hat{y}_1^j - z_1)_+^s + a^-(y_0)(\hat{y}_1^j - z_1)_-^s]. \quad (\text{D.3})$$

We start by estimating the difference between the terms with the subscript ‘+’ inside the brackets in (D.1) and (D.3)

$$\begin{aligned} & \left| a^+(\hat{y}^j) P_+^s(\hat{y}^j) - a^+(y_0)(\hat{y}_1^j - z_1)_+^s \right| \\ & \leq \left| P_+^s(\hat{y}^j) - (\hat{y}_1^j - z_1)_+^s \right| |a^+(\hat{y}^j)| + |\hat{y}_1^j - z_1|^s \left| a^+(\hat{y}^j) - a^+(y_0) \right|, \quad (\text{D.4}) \\ & \hat{y}^j = U^T(\epsilon D j - \tilde{y}_0). \end{aligned}$$

The following inequalities can be shown to hold. For all $q, r \in \mathbb{R}$ one has

$$\begin{aligned} |(q+r)_{\pm}^s - q_{\pm}^s| &\leq 2^{s-1}(|r|^s + s|q|^{s-1}|r|), \quad s > 1, \\ |(q+r)_{\pm}^s - q_{\pm}^s| &\leq |r|^s, \quad 0 < s \leq 1. \end{aligned} \quad (\text{D.5})$$

Consider the top inequality. The case $q, q+r \leq 0$ is trivial. The cases $q+r \leq 0 \leq q$ and $q \leq 0 \leq q+r$ can be verified directly. By a change of variables and convexity, it is easily seen that the case $r < 0 < q$ follows from the case $q, r > 0$. To prove the latter, divide by q^s and set $x = r/q$. Both sides equal zero when $x = 0$. Differentiating with respect to x , we see that the inequality is proven because $(1+x)^{s-1} \leq 2^{s-1}(x^{s-1} + 1)$ (consider $0 < x \leq 1$ and $x \geq 1$). The second inequality in (D.5) is obvious.

The assumption $z \in \Gamma$ implies $z_1 = \psi(z^{\perp})$, so

$$P(\hat{y}^j) = \hat{y}_1^j - \psi((\hat{y}^j)^{\perp}) = \hat{y}_1^j - z_1 + [\psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})]. \quad (\text{D.6})$$

Setting $q = \hat{y}_1^j - z_1$ and $r = \psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})$ in (D.5) and using (D.6) and that $a^+(y)$ is bounded, we estimate the first term on the right in (D.4) as follows

$$\begin{aligned} &|P_+^s(\hat{y}^j) - (\hat{y}_1^j - z_1)_+^s| |a^+(\hat{y}^j)| \\ &\leq c \left(|\psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})|^s + \begin{cases} |\hat{y}_1^j - z_1|^{s-1} |\psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})|, & s > 1 \\ 0, & 0 < s \leq 1 \end{cases} \right). \end{aligned} \quad (\text{D.7})$$

Recall that in this lemma we assume that the amplitude of $\mathcal{B}$ satisfies $\tilde{B}(y, \eta) \equiv \tilde{B}_0(y, \eta)$. By (4.8), the fact that the amplitude of $\mathcal{B}$ is homogeneous in the frequency variable (and, therefore, the Schwartz kernel $K(y, w)$ of $\mathcal{B}$ is homogeneous in w), and Assumption 4.3(IK1),

$$|\mathcal{B}\varphi(u)| \leq c(1 + |u|)^{-(\beta+n)}, \quad u \in \mathbb{R}^n. \quad (\text{D.8})$$

Therefore, by (8.11) and (D.1)–(D.3), we have to estimate the following two sums

$$\begin{aligned} J_1 &:= \sum_{|j| \leq \vartheta/\epsilon} \frac{|\psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})|/\epsilon^s}{(1 + |(y - \hat{y}^j)/\epsilon|)^{\beta+n}}, \\ J_2 &:= \sum_{|j| \leq \vartheta/\epsilon} \frac{|\hat{y}_1^j - z_1|/\epsilon^{s-1} |\psi(z^{\perp}) - \psi((\hat{y}^j)^{\perp})|/\epsilon}{(1 + |(y - \hat{y}^j)/\epsilon|)^{\beta+n}}. \end{aligned} \quad (\text{D.9})$$

The second sum is required if $s > 1$.

Note that the quantities $J_{1,2}$ include the factor ϵ^{-s} , which appears on the left in (8.11) and has been unaccounted for until now. The remaining factor ϵ^{β} has been accounted for in (D.1). In (8.11), $\mathcal{B}_0$ already acts with respect to the rescaled variable y/ϵ , so the factor ϵ^{β} is not needed on the right in (8.11). Since $\mathcal{B}_0$ is shift-invariant, it is not necessary to represent its action in the form (D.2).

Assumptions of the lemma imply

$$|\psi(z^\perp) - \psi((\hat{y}^j)^\perp)| \leq |\psi'(y_*^\perp)| |z^\perp - (\hat{y}^j)^\perp| \leq c(\epsilon^{1/2} + |z^\perp - (\hat{y}^j)^\perp|) |z^\perp - (\hat{y}^j)^\perp| \quad (\text{D.10})$$

for some $c > 0$. Here $y_*^\perp \in \mathbb{R}^{n-1}$ is some point on the line segment with the endpoints $z^\perp$, $(\hat{y}^j)^\perp$, and we have used that $|\psi'(y_*^\perp)| \leq c(|z^\perp| + |z^\perp - (\hat{y}^j)^\perp|)$, which follows from $\psi'(y_0^\perp) = 0$.

Let $m = m(z, \epsilon) \in \mathbb{Z}^n$ be such that $|(z + U^T \tilde{y}_0)/\epsilon - U^T Dm| < c$. The dependence of m on z and ϵ is omitted from notations. This implies

$$\begin{aligned} \max(|z_1 - \hat{y}_1^j|, |z^\perp - (\hat{y}^j)^\perp|) &\leq |z - \hat{y}^j| \leq c\epsilon \left| \frac{z + U^T \tilde{y}_0}{\epsilon} - U^T Dm - U^T D(j - m) \right| \\ &\leq c\epsilon(c + |j - m|). \end{aligned} \quad (\text{D.11})$$

Also, using that $|y - z| = O(\epsilon)$ gives

$$\left| \frac{y - \hat{y}^j}{\epsilon} \right| = \left| \frac{(y - z) + (z - \hat{y}^j)}{\epsilon} \right| \geq c|j - m| \text{ if } |j - m| \gg 1. \quad (\text{D.12})$$

Substitute (D.10) into the expression for J_1 in (D.9), shift the index $j \rightarrow j - m$, and use (D.11), (D.12):

$$J_1 \leq c \sum_{|j| \leq \vartheta/\epsilon} \frac{(\epsilon^{1/2} + \epsilon(c + |j|))^s (c + |j|)^s}{(1 + c|j|)^{\beta+n}} + O(\epsilon^{s/2}). \quad (\text{D.13})$$

Here we have used that we can ignore any finite number of terms (their contribution is $O(\epsilon^{s/2})$), and (D.12) applies to the remaining terms. This gives

$$\begin{aligned} J_1 &\leq c \sum_{0 < |j| \leq \vartheta/\epsilon} \frac{(\epsilon^{1/2} + \epsilon|j|)^s}{|j|^{\beta+n-s}} + O(\epsilon^{s/2}) \\ &\leq c \int_1^{\vartheta/\epsilon} \frac{(\epsilon^{1/2} + \epsilon r)^s}{r^{\beta+1-s}} dr + O(\epsilon^{s/2}) = O(\epsilon^{\min(\beta-s, s/2)}). \end{aligned} \quad (\text{D.14})$$

To estimate J_2 , we use the same approach as in (D.10)–(D.14):

$$\begin{aligned} J_2 &\leq c \sum_{|j| \leq \vartheta/\epsilon} \frac{(\epsilon^{1/2} + \epsilon|j|)(c + |j|)^s}{(1 + c|j|)^{\beta+n}} + O(\epsilon^{1/2}) \\ &\leq c \sum_{0 < |j| \leq \vartheta/\epsilon} \frac{\epsilon^{1/2} + \epsilon|j|}{|j|^{\beta+n-s}} + O(\epsilon^{1/2}) \\ &\leq c \int_1^{\vartheta/\epsilon} \frac{\epsilon^{1/2} + \epsilon r}{r^{\beta+1-s}} dr + O(\epsilon^{1/2}) = O(\epsilon^{\min(\beta-s, 1/2)}) = O(\epsilon^{1/2}). \end{aligned} \quad (\text{D.15})$$

Here we have used that $\beta - s \geq N/2 \geq 1/2$.

The second term on the right in (D.4) is estimated as follows:

$$\begin{aligned} |\hat{y}_1^j - z_1|^s \left| a^+(\hat{y}^j) - a^+(y_0) \right| &\leq |\hat{y}_1^j - z_1|^s \left| (a^+(\hat{y}^j) - a^+(z)) + (a^+(z) - a^+(y_0)) \right| \\ &\leq c[\epsilon(c + |j - m|)]^s (\epsilon(c + |j - m|) + \epsilon^{1/2}). \end{aligned} \quad (\text{D.16})$$

Shifting the j index as before and estimating a finite number of terms by $O(\epsilon^{1/2})$ gives an upper bound

$$\sum_{0 < |j| \leq \vartheta/\epsilon} \frac{\epsilon^{1/2} + \epsilon|j|}{|j|^{\beta+n-s}} + O(\epsilon^{1/2}) = O(\epsilon^{1/2}). \quad (\text{D.17})$$

The terms with the subscript $'-'$ in (8.11) are estimated analogously. Our argument proves (8.11) with $\mathcal{B}$ instead of $\mathcal{B}_0$ on the right. This implies, in particular, that the sum on the right in (8.11) is restricted to $|j| \leq \vartheta/\epsilon$.

The left-hand side of (8.11) is bounded, because

$$\begin{aligned} |P(\hat{y}^j)| &\leq |\hat{y}_1^j - z_1| + |\psi(z^\perp) - \psi((\hat{y}^j)^\perp)| \\ &\leq c\epsilon(c + |j - m|)(1 + (\epsilon^{1/2} + \epsilon(c + |j - m|))) \\ &\leq c\epsilon(1 + |j - m|) \end{aligned} \quad (\text{D.18})$$

by (D.6), (D.10), (D.11), and $|j| \leq \vartheta/\epsilon$, and

$$\frac{|P(\hat{y}^j)/\epsilon|^s}{(1 + |(y - \hat{y}^j)/\epsilon|)^{\beta+n}} \leq c \sum_{0 < |j| \leq \vartheta/\epsilon} \frac{1}{|j|^{\beta+n-s}} + O(1) < \infty. \quad (\text{D.19})$$

It is easy to see that

$$\epsilon^\beta \left| \left(\mathcal{B}\varphi \left(\cdot - \frac{\hat{y}^j}{\epsilon} \right) \right) (y) - \mathcal{B}_0\varphi \left(\frac{y - \hat{y}^j}{\epsilon} \right) \right| \leq c \frac{\epsilon^{1/2}}{(1 + |(y - \hat{y}^j)/\epsilon|)^{\beta+n}}. \quad (\text{D.20})$$

This follows from $\varphi \in C_0^{\lceil \beta_0^+ \rceil}$, $|y - y_0| = O(\epsilon^{1/2})$, and

$$|\partial_\eta^m (\tilde{B}_0(y, \eta) - \tilde{B}_0(y_0, \eta))| \leq c_m |y - y_0| |\eta|^{\beta-|m|}, \quad |\eta| \geq 1, \quad m \in \mathbb{N}_0^n. \quad (\text{D.21})$$

Together with (D.19) this implies that replacing y with y_0 in the amplitude of the $\Psi\text{DO } \mathcal{B}$ (i.e., replacing $\tilde{B}_0(y, \eta)$ with $\tilde{B}_0(y_0, \eta)$) introduces an error of the magnitude $O(\epsilon^{1/2})$, while keeping the sum restricted to $|j| \leq \vartheta/\epsilon$.

Using that $|y - z| = O(\epsilon)$, (D.3) and (D.8) imply that the terms of the series on the right in (8.11) are bounded by $O((1 + |j|)^{s-(\beta+n)})$. Hence the series is absolutely

convergent. Contribution of the terms corresponding to $|j| > \vartheta/\epsilon$ is bounded by $c \sum_{|j| > \vartheta/\epsilon} |j|^{s-(\beta+n)} = O(\epsilon^{\beta-s}) \rightarrow 0$ for some $c > 0$, and the lemma is proven.

Appendix E. Proof of Lemma 8.3

Pick some sufficiently large $J \gg 1$. Then, with $D_1 := U^T D$,

$$\begin{aligned} \phi(v + \Delta v, p) - \phi(v, p) &= \sum_{j \in \mathbb{Z}^n} [\mathcal{B}_0 \varphi(v + \Delta v - D_1 j) - \mathcal{B}_0 \varphi(v - D_1 j)] \mathcal{A}(\Theta_0 \cdot D_1 j - p) \\ &= \sum_{|j| \leq J} (\cdot) + \sum_{|j| > J} (\cdot) =: J_1 + J_2. \end{aligned} \quad (\text{E.1})$$

Because $\mathcal{B}_0 \in S^{\beta_0}(\mathbb{R}^n \times \mathbb{R}^n)$, [1, Theorem 6.19] implies that $\mathcal{B}_0 : C_*^{\lceil \beta_0^+ \rceil} \rightarrow C_*^a$, $a = \lceil \beta_0^+ \rceil - \beta_0 = 1 - \{\beta_0\} > 0$, is continuous. Nonsmoothness of the symbol at the origin, which is not allowed by the assumptions of the theorem, is irrelevant. By assumption, $\varphi \in C_0^{\lceil \beta_0^+ \rceil}(\mathbb{R}^n)$, so $J_1 = O(|\Delta v|^a)$. In the second term J_2 , the arguments of $\mathcal{B}_0 \varphi$ are bounded away from zero, and the factor in brackets is smooth. Moreover, using again that the Schwartz kernel $K(y, w)$ of $\mathcal{B}_0$ is homogeneous in w , we have,

$$|\partial_{u_l} \mathcal{B}_0 \varphi(u)| = O(|u|^{-(n+\beta_0+1)}), \quad |u| \rightarrow \infty, \quad 1 \leq l \leq n. \quad (\text{E.2})$$

Using the argument analogous to the one in (D.19), we easily see that $J_2 = O(|\Delta v|)$. This proves the first line in (8.22).

The second line in (8.22) is proven analogously:

$$\begin{aligned} \phi(v, p + \Delta p) - \phi(v, p) &= \sum_{j \in \mathbb{Z}^n} \mathcal{B}_0 \varphi(v - D_1 j) [\mathcal{A}(\Theta_0 \cdot D_1 j - (p + \Delta p)) - \mathcal{A}(\Theta_0 \cdot D_1 j - p)] \\ &= \sum_{|\Theta_0 \cdot D_1 j| \leq J} (\cdot) + \sum_{|\Theta_0 \cdot D_1 j| > J} (\cdot) =: J_1 + J_2. \end{aligned} \quad (\text{E.3})$$

Clearly, $\mathcal{A}(q + \Delta p) - \mathcal{A}(q) = O(|\Delta p|^{\min(s_0, 1)})$ uniformly in q confined to any bounded set. Using in addition that $\mathcal{B}_0 \varphi(u)$ is bounded and $\mathcal{B}_0 \varphi(u) = O(|u|^{-(n+\beta_0)})$ as $|u| \rightarrow \infty$, we get that $J_1 = O(|\Delta p|^{\min(s_0, 1)})$.

In J_2 , the argument of $\mathcal{A}$ is bounded away from zero. In view of $\mathcal{A}'(q) = O(|q|^{s_0-1})$, $|q| \rightarrow \infty$, we finish the proof by noticing that

$$|J_2| \leq O(|\Delta p|) \sum_{|j| > 0} \frac{|j|^{s_0-1}}{|j|^{n+\beta_0+1}} = O(|\Delta p|). \quad (\text{E.4})$$

The fact that both estimates are uniform with respect to v and p confined to bounded sets is obvious.

Appendix F. Proof of Lemma 10.3

As usual, c denotes various positive constants that may have different values in different places. Recall that $\beta_0 - s_0 > 0$. Set $k := \lceil \beta_0 \rceil$, $\nu := k - \beta_0$. Thus, $0 \leq \nu < 1$, and $\nu = 0$ if $\beta_0 \in \mathbb{N}$. Similarly to (9.2),

$$\mathcal{B} = \mathcal{W}_1 \partial_{y_1}^k + \mathcal{W}_2, \quad (\text{F.1})$$

for some $\mathcal{W}_1 \in S^{-\nu}(\mathcal{V} \times \mathbb{R}^n)$, $\mathcal{W}_2 \in S^{-\infty}(\mathcal{V} \times \mathbb{R}^n)$.

F.1. Proof in the Case $\beta_0 \notin \mathbb{N}$

Let $K(y, w)$ be the Schwartz kernel of $\mathcal{W}_1$. Suppose, for example, that $P := P(y) > 0$. The case $P < 0$ is completely analogous. Initially, as $\varkappa_2$ in Lemma 10.3, we can pick any constant that satisfies $\varkappa_2 \geq 2\varkappa_1$, where $\varkappa_1$ is the same as in (7.11). This implies that $P/2 \geq \varkappa_1 \epsilon$. Later (see the beginning of the proof of Lemma F.1), we update the choice of $\varkappa_2$. Denote (cf. (10.5))

$$J_\epsilon(y) := (\mathcal{B}g_\epsilon)(y) - (\mathcal{B}g)(y) = (\mathcal{B}\Delta g_\epsilon)(y). \quad (\text{F.2})$$

Then

$$\begin{aligned} J_\epsilon(y) &= J_\epsilon^{(1)}(y) + J_\epsilon^{(2)}(y) + O(\epsilon^{s_0}), \\ J_\epsilon^{(1)}(y) &:= \int_{|P(w)| \geq P/2} K(y, y-w) \Delta g_\epsilon^{(k)}(w) dw, \\ J_\epsilon^{(2)}(y) &:= \int_{|P(w)| \leq P/2} K(y, y-w) \Delta g_\epsilon^{(k)}(w) dw. \end{aligned} \quad (\text{F.3})$$

The big- O term in (F.3) appears because of the Ψ DO $\mathcal{W}_2$ in (F.1), and the magnitude of the term follows from (7.12) with $l = 0$. From (7.9) and (7.11) with $l = k$, (F.3), and (9.4) with $l = 0$, it follows that

$$\begin{aligned} |J_\epsilon^{(1)}(y)| &\leq c\epsilon \int_{|P(w)| \geq P/2} \frac{|w_1 - \psi(w^\perp)|^{s_0-1-k}}{|y-w|^{n-\nu}} dw \\ &= c\epsilon \int_{|p| \geq P/2} \int \frac{|p|^{s_0-1-k}}{|([P-p] + \psi(y^\perp) - \psi(w^\perp), y^\perp - w^\perp)|^{n-\nu}} dw^\perp dp. \end{aligned} \quad (\text{F.4})$$

Hence, we obtain similarly to (9.8)

$$\begin{aligned} |J_\epsilon^{(1)}(y)| &\leq c\epsilon \int \int_{|p| \geq P/2} \frac{|p|^{s_0-1-k}}{(|P-p| + |w^\perp|)^{n-\nu}} dp dw^\perp \\ &\leq c\epsilon \int_{|p| \geq P/2} \frac{|p|^{s_0-1-k}}{|P-p|^{1-\nu}} dp = c\epsilon P^{s_0-1-\beta_0}. \end{aligned} \quad (\text{F.5})$$

To estimate $J_\epsilon^{(2)}(y)$, integrate by parts with respect to w_1 in (F.3):

$$|J_\epsilon^{(2)}(y)| \leq c \left(J_k + \sum_{l=l_0}^{k-1} (J_l^- + J_l^+) \right), \quad l_0 := \lfloor s_0^- \rfloor, \quad (F.6)$$

$$J_k := \int_{|P(w)| \leq P/2} \left| \partial_{w_1}^{k-l_0} K(y, y-w) \Delta g_\epsilon^{(l_0)}(w) \right| dw,$$

$$J_l^\pm := \int_{\mathbb{R}^{n-1}} \left| \partial_{w_1}^{k-1-l} K(y, y-w) \Delta g_\epsilon^{(l)}(w) \right|_{w=(\psi(w^\perp) \pm P/2, w^\perp)} dw^\perp.$$

By construction, $P/2 \geq \varkappa_1 \epsilon$. Using (7.11), (7.12) with $l = l_0$ (both inequalities apply when $l = l_0 = \lfloor s_0^- \rfloor$), and arguing similarly to (F.4), (F.5), gives

$$\begin{aligned} J_k &\leq c \int_{|p| \leq \varkappa_1 \epsilon} \frac{\epsilon^{s_0-l_0}}{|P-p|^{\beta_0+1-l_0}} dp + c\epsilon \int_{\varkappa_1 \epsilon \leq |p| \leq P/2} \frac{|p|^{s_0-l_0-1}}{|P-p|^{\beta_0+1-l_0}} dp \\ &\leq c\epsilon^{s_0-l_0+1} P^{-(\beta_0+1-l_0)} + c\epsilon P^{s_0-1-\beta_0} \int_{\varkappa_1(\epsilon/P) \leq |\check{p}| \leq 1/2} \frac{|\check{p}|^{s_0-l_0-1}}{|1-\check{p}|^{\beta_0+1-l_0}} d\check{p} \\ &\leq c\epsilon^{s_0-l_0+1} P^{-(\beta_0+1-l_0)} + c\epsilon P^{s_0-1-\beta_0} \left(1 + (\epsilon/P)^{s_0-l_0} \right), \end{aligned} \quad (F.7)$$

where we have used that $l_0 < s_0$. Using again that $\epsilon/P \leq 1/(2\varkappa_1)$ gives $J_k \leq c\epsilon P^{s_0-1-\beta_0}$.

Next we estimate the boundary terms in (F.6). By (7.11) (using that $\lfloor s_0^- \rfloor = l_0 \leq l \leq k-1$) and (9.4),

$$J_l^\pm \leq c\epsilon \int_{\mathbb{R}^{n-1}} \frac{|w_1 - \psi(w^\perp)|^{s_0-1-l}}{|y-w|^{n+\beta_0-1-l}} dw^\perp, \quad w_1 = \psi(w^\perp) \pm P/2. \quad (F.8)$$

Appealing to (9.7) gives

$$J_l^\pm \leq c\epsilon P^{s_0-1-l} \int_{\mathbb{R}^{n-1}} \frac{dw^\perp}{(P \pm (P/2) + |w^\perp|)^{n+\beta_0-l-1}} = c\epsilon P^{s_0-1-\beta_0}, \quad (F.9)$$

which finishes the proof. As easily checked, the integral in (F.9) converges because $l \leq k-1 < \beta_0$.

F.2. Proof in the Case $\beta_0 \in \mathbb{N}$

Suppose now $\beta_0 \in \mathbb{N}$, i.e. $k = \beta_0$ and $\nu = 0$. All the terms that do not involve integration over a neighborhood of the set $\{w \in \mathcal{V} : P(w) = P\}$ are estimated the same way as before. For example, estimation of $J_\epsilon^{(2)}(y)$ is completely analogous to (F.6)–(F.9), and we obtain the same bound $|J_\epsilon^{(2)}(y)| \leq c\epsilon P^{s_0-1-\beta_0}$. Estimating of

$J_\epsilon^{(1)}$ is much more involved now, because the singularity at $P(w) = P$ is no longer integrable. We have with some $c_1 > 0$, which is to be selected later:

$$\begin{aligned} J_\epsilon^{(1)}(y) &= J_\epsilon^{(1a)}(y) + J_\epsilon^{(1b)}(y) + J_\epsilon^{(1c)}(y), \\ J_\epsilon^{(1a)}(y) &:= \int_{P/2 \leq P(w) \leq P-c_1\epsilon} K(y, y-w) \Delta g_\epsilon^{(\beta_0)}(w) dw, \\ J_\epsilon^{(1b)}(y) &:= \int_{P-c_1\epsilon \leq P(w) \leq P+c_1\epsilon} K(y, y-w) \Delta g_\epsilon^{(\beta_0)}(w) dw, \\ J_\epsilon^{(1c)}(y) &:= \int_{P+c_1\epsilon \leq P(w)} K(y, y-w) \Delta g_\epsilon^{(\beta_0)}(w) dw. \end{aligned} \quad (\text{F.10})$$

We do not estimate the integral $\int_{P(w) \leq P/2} (\cdot) dw$, because the domain of integration is bounded away from the set $\{w \in \mathcal{V} : P(w) = P\}$, and this integral admits the same bound as in the previous subsection (cf. (F.5)). Similarly to (F.5), by (7.11) with $l = \beta_0$,

$$\begin{aligned} |J_\epsilon^{(1a)}(y)| &\leq c\epsilon \int_{P/2 \leq p \leq P-c_1\epsilon} \frac{p^{s_0-1-\beta_0}}{P-p} dp = c\epsilon P^{s_0-1-\beta_0} \ln(P/\epsilon), \\ |J_\epsilon^{(1c)}(y)| &\leq c\epsilon \int_{P+c_1\epsilon \leq p} \frac{p^{s_0-1-\beta_0}}{p-P} dp = c\epsilon P^{s_0-1-\beta_0} \ln(P/\epsilon). \end{aligned} \quad (\text{F.11})$$

The term $J_\epsilon^{(1b)}$ is split further as follows:

$$\begin{aligned} J_\epsilon^{(1b)}(y) &= J_\epsilon^{(1b1)}(y) + J_\epsilon^{(1b2)}(y) + J_\epsilon^{(1b3)}(y), \\ J_\epsilon^{(1b1)}(y) &:= \int_{\substack{|P-P(w)| \leq c_1\epsilon \\ |y^\perp - w^\perp| \geq c_1P}} K(y, y-w) \Delta g_\epsilon^{(\beta_0)}(w) dw, \\ J_\epsilon^{(1b2)}(y) &:= \int_{\substack{|P-P(w)| \leq c_1\epsilon \\ |y^\perp - w^\perp| \leq c_1P}} K(y, y-w) (\Delta g_\epsilon^{(\beta_0)}(w) - \Delta g_\epsilon^{(\beta_0)}(y)) dw, \\ J_\epsilon^{(1b3)}(y) &:= \Delta g_\epsilon^{(\beta_0)}(y) I, \quad I := \int_{\substack{|P-P(w)| \leq c_1\epsilon \\ |y^\perp - w^\perp| \leq c_1P}} K(y, y-w) dw. \end{aligned} \quad (\text{F.12})$$

Similarly to (F.5), by (7.11) with $l = \beta_0$,

$$|J_\epsilon^{(1b1)}(y)| \leq c\epsilon \int_{|w^\perp| \geq c_1P} \int_{|P-p| \leq c_1\epsilon} \frac{p^{s_0-1-\beta_0}}{(|P-p| + |w^\perp|)^n} dp dw^\perp \leq c\epsilon^2 P^{s_0-2-\beta_0}. \quad (\text{F.13})$$

The second part is estimated by rearranging the Δg terms:

$$\begin{aligned} J_\epsilon^{(1b2)}(y) &:= \int_{\substack{|P-P(w)| \leq c_1\epsilon \\ |y^\perp - w^\perp| \leq c_1P}} K(y, y-w) [(g_\epsilon^{(\beta_0)}(w) - g_\epsilon^{(\beta_0)}(y)) \\ &\quad - (g^{(\beta_0)}(w) - g^{(\beta_0)}(y))] dw. \end{aligned} \quad (\text{F.14})$$

Lemma F.1 *There exist $c, c_1, \kappa_2 > 0$ so that*

$$\begin{aligned} |g^{(\beta_0)}(w) - g^{(\beta_0)}(y)| &\leq c|w - y|P(y)^{s_0-1-\beta_0}, \\ |g_\epsilon^{(\beta_0)}(w) - g_\epsilon^{(\beta_0)}(y)| &\leq c|w - y|P(y)^{s_0-1-\beta_0}, \\ \text{if } y, w \in \mathcal{V}, |P(y) - P(w)| &\leq c_1\epsilon, |y^\perp - w^\perp| \leq c_1P(y), P(y) > \kappa_2\epsilon. \end{aligned} \quad (\text{F.15})$$

Proof We begin by updating the choice of κ_2 . Select $\kappa_2 \geq 2\kappa_1$ so that $P \geq \kappa_2\epsilon$ implies

$$P(v) \geq cP \text{ for any } v \in \mathcal{V}, |v - y| \leq \epsilon d_\varphi, d_\varphi := \text{diam}(\text{supp } \varphi), \quad (\text{F.16})$$

for some $c > 0$.

Next we select c_1 . First, pick any c_1 so that $0 < c_1 \leq \kappa_1$. This ensures that $P(w) \geq P - |P - P(w)| \geq \kappa_1\epsilon$, and (7.11) can be used to estimate the derivatives of $\Delta g_\epsilon(w)$. Let $c_\psi := \max_{v \in \mathcal{V}} |\psi'(v)|$. Our assumptions imply

$$|y_1 - w_1| \leq |P - P(w)| + |\psi(y^\perp) - \psi(w^\perp)| \leq c_1(\epsilon + c_\psi P). \quad (\text{F.17})$$

Let v be any point on the line segment with the endpoints w and y , i.e. $v = y + \lambda(w - y)$, $0 \leq \lambda \leq 1$. Then

$$P(v) \geq P - (|y_1 - w_1| + |\psi(v^\perp) - \psi(y^\perp)|) \geq P - c_1(\epsilon + c_\psi P) - c_\psi c_1 P. \quad (\text{F.18})$$

Reducing $c_1 > 0$ even further, we can ensure that $P(v) \geq cP$ for some $c > 0$. This is the value of c_1 that is assumed starting from (F.10). In the rest of the proof we assume that $w, y \in \mathcal{V}$ satisfy the inequalities on the last line in (F.15) with the constants c_1 and κ_2 that we have just selected.

From (7.5) with $|m| = \beta_0 + 1$,

$$\begin{aligned} |g^{(\beta_0)}(w) - g^{(\beta_0)}(y)| &\leq |w - y| \max_{0 \leq \lambda \leq 1} |(\partial_y g^{(\beta_0)})(y + \lambda(w - y))| \\ &\leq c|w - y|P^{s_0-1-\beta_0} \end{aligned} \quad (\text{F.19})$$

for some $c > 0$.

To prove the second line in (F.15), find $c_{2,3} > 0$ such that

$$v \in \mathcal{V}, |v - y| \leq \epsilon(c_2 + d_\varphi) \text{ implies } P(v) \geq c_3P. \quad (\text{F.20})$$

By (F.16), $c_{2,3}$ with the required properties do exist.

Now, assume first that $|w - y| \geq c_2\epsilon$, where c_2 is the same as in (F.20). Clearly,

$$|g_\epsilon^{(\beta_0)}(w) - g_\epsilon^{(\beta_0)}(y)| \leq |\Delta g_\epsilon^{(\beta_0)}(w)| + |g^{(\beta_0)}(w) - g^{(\beta_0)}(y)| + |\Delta g_\epsilon^{(\beta_0)}(y)|. \quad (\text{F.21})$$

By construction, (7.11) applies to $\Delta g_\epsilon^{(\beta_0)}(w)$. Applying (7.11) to the first and third terms on the right in (F.21), and (F.19) – to the second term on the right, gives

$$\begin{aligned} |g_\epsilon^{(\beta_0)}(w) - g_\epsilon^{(\beta_0)}(y)| &\leq c\epsilon P(w)^{s_0-1-\beta_0} + c|w-y|P^{s_0-1-\beta_0} + c\epsilon P^{s_0-1-\beta_0} \\ &\leq c|w-y|P^{s_0-1-\beta_0}, \end{aligned} \quad (\text{F.22})$$

because $\epsilon \leq (1/c_2)|w-y|$ and

$$P(w) \geq P - |P - P(w)| \geq P(1 - c_1(\epsilon/P)) \geq P(1 - c_1/(2\kappa_1)) \geq P/2. \quad (\text{F.23})$$

If $|w-y| \leq c_2\epsilon$, we argue similarly to (C.22):

$$\begin{aligned} g_\epsilon^{(\beta_0)}(w) - g_\epsilon^{(\beta_0)}(y) &= \epsilon^{-\beta_0} \sum_j \left((\partial_{v_1}^{\beta_0} \varphi) \left(\frac{w - \hat{y}^j}{\epsilon} \right) - (\partial_{v_1}^{\beta_0} \varphi) \left(\frac{y - \hat{y}^j}{\epsilon} \right) \right) \\ &\quad \times \sum_{|m|=\beta_0+1} R_m(\hat{y}^j, y) (\hat{y}^j - y)^m \\ |R_m(\hat{y}^j, y)| &\leq \frac{1}{(\beta_0 + 1)!} \max_{|m'|=\beta_0+1, |y-v| \leq (c_2+d_\varphi)\epsilon} |\partial_v^{m'} g(v)|. \end{aligned} \quad (\text{F.24})$$

Here $(\partial_{v_1} \varphi)(\cdot)$ is the derivative of $\varphi(v)$ with respect to v_1 evaluated at the indicated point. By (F.20), (7.5) implies $|R_m(\hat{y}^j, y)| \leq cP^{s_0-1-\beta_0}$. The assertion follows because $\varphi \in C_0^{\lceil \beta_0+1 \rceil}(\mathbb{R}^n)$. $\square$

Applying (9.4) with $v = l = 0$ and (F.15) in (F.14) yields (cf. (9.6)–(9.8))

$$\begin{aligned} |J_\epsilon^{(1b2)}(y)| &\leq cP^{s_0-1-\beta_0} \int_{\substack{|P-P(w)| \leq c_1\epsilon \\ |y^\perp - w^\perp| \leq c_1P}} \frac{|w-y|}{|y-w|^n} dw \\ &\leq cP^{s_0-1-\beta_0} \int_{|w^\perp| \leq c_1P} \int_{|P-p| \leq c_1\epsilon} \frac{1}{(|P-p| + |w^\perp|)^{n-1}} dp dw^\perp \\ &\leq cP^{s_0-1-\beta_0} \int_{|w^\perp| \leq c_1P} \int_{|p| \leq c_1\epsilon} \frac{dp dw^\perp}{(|p| + |w^\perp|)^{n-1}} = c\epsilon P^{s_0-1-\beta_0} \ln\left(\frac{P}{\epsilon}\right). \end{aligned} \quad (\text{F.25})$$

The final major step is to estimate the integral in the definition of $J_\epsilon^{(1b3)}$.

$$\begin{aligned} I &= \int_{|y^\perp - w^\perp| \leq c_1P} \int_{|(y_1 - w_1) - (\psi(y^\perp) - \psi(w^\perp))| \leq c_1\epsilon} K(y, (y_1 - w_1, y^\perp - w^\perp)) dw_1 dw^\perp \\ &= \int_{|v^\perp| \leq c_1P} \int_{-c_1\epsilon}^{c_1\epsilon} K(y, (v_1 + h(v^\perp), v^\perp)) dv_1 dv^\perp, \quad h(v^\perp) := \psi(y^\perp) - \psi(y^\perp - v^\perp). \end{aligned} \quad (\text{F.26})$$

Let $\tilde{W}(y, \eta)$ be the amplitude of $\mathcal{W}_1 \in S^0(\mathcal{V} \times \mathbb{R}^n)$ in (F.1). Then

$$\begin{aligned} I &= c \int_{|v^\perp| \leq c_1 P} \int_{-c_1 \epsilon}^{c_1 \epsilon} \int_{\Omega} \tilde{W}(y, \eta) e^{-i(\eta_1(v_1 + h(v^\perp)) + \eta^\perp v^\perp)} d\eta dv_1 dv^\perp \\ &= c \int_{\Omega} \frac{\sin(c_1 \epsilon \eta_1)}{\eta_1} \int_{|v^\perp| \leq c_1 P} \tilde{W}(y, \eta) e^{-i(\eta_1 h(v^\perp) + \eta^\perp v^\perp)} dv^\perp d\eta. \end{aligned} \quad (\text{F.27})$$

Our goal is to show that I is uniformly bounded for all $\epsilon > 0$ sufficiently small and P that satisfy $P/\epsilon \geq \kappa_2 > 0$. We can select $\mathcal{W}_{1,2}$ in (F.1) so that the conic supports of their amplitudes are contained in that of $\mathcal{B}$. First, consider only the principal symbol of $\mathcal{W}_1$, which we denote $\tilde{W}_0(y, \eta)$. We can assume that $\tilde{W}_0(y, \eta) \equiv 0$ if $\eta \notin \Omega$, $\eta \neq 0$, where $\Omega \subset \mathbb{R}^n \setminus \{0\}$ is a small conic neighborhood of $\Theta_0 \cup (-\Theta_0)$. This set is used in (F.27). The corresponding value of I , which is obtained by replacing $\tilde{W}(y, \eta)$ with $\tilde{W}_0(y, \eta)$ in (F.27), is denoted I_0 .

As $\tilde{W}_0(y, \eta)$ is positively homogeneous of degree zero in η , set

$$\tilde{W}^\pm(y, u) := \tilde{W}(y, \eta_1(1, u)) = \tilde{W}_0(y, \pm(1, u)), \quad u = \eta^\perp/\eta_1 \in \Omega^\perp, \quad (\text{F.28})$$

where $\Omega^\perp$ is a small neighborhood of the origin in $\mathbb{R}^{n-1}$: $\Omega^\perp := \{u \in \mathbb{R}^{n-1} : u = \eta^\perp/\eta_1, \eta \in \Omega\}$. The sign $' + '$ is selected if $\eta_1 > 0$, and $' - '$ - otherwise. By the properties of $\mathcal{W}$, $\tilde{W}^\pm(y, \cdot) \in C_0^\infty(\Omega^\perp)$. Thus, (F.27) implies

$$\begin{aligned} I_0 &= c \int_{\mathbb{R}} \frac{\sin(c_1 \epsilon \eta_1)}{\eta_1} \int_{|v^\perp| \leq c_1 P} \int_{\Omega^\perp} \tilde{W}^\pm(y, u) e^{-i\eta_1 u v^\perp} du e^{-i\eta_1 h(v^\perp)} dv^\perp |\eta_1|^{n-1} d\eta_1 \\ &= c \int_{\mathbb{R}} \frac{\sin(c_1 \epsilon \eta_1)}{\eta_1} \int_{|v^\perp| \leq c_1 P} W^\pm(y, \eta_1 v^\perp) e^{-i\eta_1 h(v^\perp)} dv^\perp |\eta_1|^{n-1} d\eta_1 \\ &= c \int_{\mathbb{R}} \frac{\sin(c_1 \epsilon \eta_1)}{\eta_1} \int_{|w^\perp| \leq c_1 P |\eta_1|} W^\pm(y, w^\perp) e^{-i\eta_1 h(w^\perp/\eta_1)} dw^\perp d\eta_1 \\ &= c \int_{\mathbb{R}} \frac{\sin(\lambda)}{\lambda} \int_{|w^\perp| \leq \frac{P}{\epsilon} |\lambda|} W^\pm(y, w^\perp) \exp\left(-i\lambda \frac{h(c_1 \epsilon w^\perp/\lambda)}{c_1 \epsilon}\right) dw^\perp d\lambda, \end{aligned} \quad (\text{F.29})$$

where $W^\pm(y, w^\perp)$ is the inverse Fourier transform of $\tilde{W}^\pm(y, u)$ with respect to u . Since P/ϵ is bounded away from zero, $h(0) = 0$, and $W^\pm(y, w^\perp)$ is smooth and rapidly decreasing as a function of $w^\perp$, we have by the dominated convergence theorem

$$\begin{aligned} &\int_{|w^\perp| \leq \frac{P}{\epsilon} |\lambda|} W^\pm(y, w^\perp) \exp\left(-i\lambda \frac{h(c_1 \epsilon w^\perp/\lambda)}{c_1 \epsilon}\right) dw^\perp \\ &\rightarrow \int_{\mathbb{R}^{n-1}} W^\pm(y, w^\perp) e^{-ih'(0) \cdot w^\perp} dw^\perp = \tilde{W}^\pm(y, -\psi'(y^\perp)) = \tilde{W}_0(y, \pm(1, -\psi'(y^\perp))) \end{aligned} \quad (\text{F.30})$$

as $\lambda \rightarrow \pm\infty$, and convergence is uniform with respect to ϵ and P that satisfy $P/\epsilon \geq \kappa_2$. As is seen, $\begin{pmatrix} 1 \\ -\psi'(y^\perp) \end{pmatrix}$ is a vector normal to Γ at the point $\begin{pmatrix} \psi(y^\perp) \\ y^\perp \end{pmatrix}$.

The remainder term in (F.30) is bounded by the expression

$$\begin{aligned}
 & \int_{|w^\perp| \leq \frac{P}{\epsilon} |\lambda|} |W^\pm(y, w^\perp)| \left| \exp \left(-i \frac{\lambda h(c_1 \epsilon w^\perp / \lambda)}{c_1 \epsilon} + i h'(0) \cdot w^\perp \right) - 1 \right| dw^\perp \\
 & + \int_{|w^\perp| \geq \frac{P}{\epsilon} |\lambda|} |W^\pm(y, w^\perp)| dw^\perp \\
 & \leq c \frac{\epsilon}{|\lambda|} \int_{\mathbb{R}^{n-1}} |W^\pm(y, w^\perp)| |w^\perp|^2 dw^\perp \\
 & + \int_{|w^\perp| \geq \kappa_2 |\lambda|} |W^\pm(y, w^\perp)| dw^\perp = O(|\lambda|^{-1}).
 \end{aligned} \tag{F.31}$$

Due to $\tilde{W}^\pm(y, \cdot) \in C_0^\infty(\Omega^\perp)$, the big- O term on the right-hand side of (F.31) is uniform with respect to $y \in \mathcal{V}$ and $0 < \epsilon \leq 1$. Hence

$$\begin{aligned}
 I_0 &= c \int_{\mathbb{R}} \frac{\sin(\lambda)}{\lambda} \tilde{W}_0(y, \lambda(1, -\psi'(y^\perp))) d\lambda + O(1) \\
 &= c \frac{\pi}{2} \left[\tilde{W}_0(y, (1, -\psi'(y^\perp))) + \tilde{W}_0(y, -(1, -\psi'(y^\perp))) \right] + O(1),
 \end{aligned} \tag{F.32}$$

where $O(1)$ is uniform with respect to $y \in \mathcal{V}$ as well, which proves that I_0 is uniformly bounded.

The remaining term $\Delta I = I - I_0$ comes from the subprincipal terms of the amplitude $\Delta \tilde{W} = \tilde{W} - \tilde{W}_0$. The corresponding Ψ DO is in $S^{-\nu}(\mathcal{V} \times \mathbb{R}^n)$ for some $\nu > 0$, so its Schwartz kernel $\Delta K(y, w)$ is smooth as long as $w \neq 0$ and absolutely integrable at $w = 0$. It is now obvious that ΔI is bounded as well.

By Lemma 7.4 (use (7.11) with $l = k = \beta_0$), $|\Delta g_\epsilon^{(\beta_0)}(y)| \leq c\epsilon P^{s_0-1-\beta_0}$ if $P \geq \kappa_1 \epsilon$, combining with (F.12) proves that $|J_\epsilon^{(1b2)}(y)| \leq c\epsilon P^{s_0-1-\beta_0}$. By (F.12), (F.13), (F.25), we conclude $|J_\epsilon^{(1b)}(y)| \leq c\epsilon P^{s_0-1-\beta_0} \ln(P/\epsilon)$. Combining with (F.10) and (F.11) we finish the proof.

Appendix G. Proof of Lemma 11.1

We begin by proving (11.8). From (3.5) and (3.9), $|\mathrm{d}\Psi| \partial \Phi_1 / \partial y_1 = 1$, i.e. $\partial \Phi_1 / \partial y_1 > 0$.

Recall that $y = Y(y^{(2)}, x)$ is found by solving $x^{(1)} = \Phi^{(1)}(x^{(2)}, y)$ for $y^{(1)}$. Differentiating $x_1 \equiv \Phi_1(x^{(2)}, (Y^{(1)}(y^{(2)}, x), y^{(2)}))$ with respect to x_1 gives $1 = (\partial \Phi_1 / \partial y_1)(\partial Y_1 / \partial x_1)$. Since $\partial \Phi_1 / \partial x^{(2)} = 0$ and $\partial \Phi_1 / \partial y^\perp = 0$, differentiating the same identity with respect to $x^\perp$ gives $0 = (\partial \Phi_1 / \partial y_1)(\partial Y_1 / \partial x^\perp)$, and all the statements in (11.8) are proven.

By (3.13), (6.3), and (6.14),

$$|\det Q|^{1/2} = \frac{|\det M_{22}|}{|\det(\Psi \circ \Phi)_{x^{(2)}, x^{(2)}}|^{1/2}} = \frac{|\det(\partial^2 \Phi_1 / \partial x^{(2)} \partial y^{(2)})|}{|\mathrm{d}\Psi|^{N/2} |\det \Delta \Pi_S|^{1/2}}. \tag{G.1}$$

Using that $|\mathrm{d}\Psi| \partial \Phi_1 / \partial y_1 = 1$ completes the proof.

References

1. Abels, H.: Pseudodifferential and Singular Integral Operators: An Introduction with Applications. De Gruyter, Berlin/Boston (2012)
2. Airapetyan, R.G., Ramm, A.G.: Singularities of the Radon transform. *Appl. Anal.* **79**, 351–371 (2001)
3. Andersson, F., De Hoop, M.V., Wendt, H.: Multiscale discrete approximation of fourier integral operators. *Multiscale Model. Simul.* **10**(1), 111–145 (2012)
4. Candes, E., Demanet, L., Ying, L.: Fast computation of Fourier integral operators. *SIAM J. Sci. Comput.* **29**, 2464–2493 (2007)
5. Candes, E., Demanet, L., Ying, L.: A fast butterfly algorithm for the computation of Fourier integral operators. *SIAM Multiscale Model. Simul.* **7**, 1727–1750 (2009)
6. Duistermaat, J.J., Hormander, L.: Fourier integral operators. II. *Acta Math.* **128**, 183–269 (1972)
7. Faridani, A.: Sampling theory and parallel-beam tomography. In: Benedetto, J.J. (ed.) *Sampling, Wavelets, and Tomography, Applied and Numerical Harmonic Analysis*, vol. 63, pp. 225–254. Birkhauser, Boston (2004)
8. Faridani, A., Buglione, K., Huabsomboon, P., Iancu, O., McGrath, J.: Introduction to local tomography. In *Radon transforms and tomography*. *Contemp. Math.* **278**, 29–47 (2001)
9. Gelfand, I.M., Shilov, G.E.: *Generalized Functions Volume 1: Properties and Operations*. Academic Press, New York (1964)
10. Greenleaf, A., Seeger, A.: Oscillatory and fourier integral operators with degenerate canonical relations. *Publicacions Matemàtiques* **48**, 93–141 (2002)
11. Guillemin, V., Sternberg, S.: *Geometric Asymptotics*. Mathematical Surveys, vol. 14. American Mathematical Society, Providence (1977)
12. Hormander, L.: Fourier integral operators. I. *Acta Math.* **127**, 79–183 (1971)
13. Hormander, L.: *The Analysis of Linear Partial Differential Operators III. Pseudo-Differential Operators*. Springer-Verlag, Berlin (2007)
14. Hormander, L.: *The Analysis of Linear Partial Differential Operators IV. Fourier Integral Operators*. Springer-Verlag, Berlin (2009)
15. Kalender, W.A.: *Computed Tomography. Fundamentals, System Technology, Image Quality, Applications*, 3rd edn. Publicis, Erlangen (2011)
16. Katsevich, A.: Asymptotics of pseudodifferential operators acting on functions with corner singularities. *Appl. Anal.* **72**, 229–252 (1999)
17. Katsevich, A.: A local approach to resolution analysis of image reconstruction in tomography. *SIAM J. Appl. Math.* **77**(5), 1706–1732 (2017)
18. Katsevich, A.: Analysis of reconstruction from discrete Radon transform data in $\mathbb{R}^3$ when the function has jump discontinuities. *SIAM J. Appl. Math.* **79**, 1607–1626 (2019)
19. Katsevich, A.: Analysis of resolution of tomographic-type reconstruction from discrete data for a class of distributions. *Inverse Probl.* **36**(12), 124008 (2020)
20. Katsevich, A.: Resolution analysis of inverting the generalized Radon transform from discrete data in $\mathbb{R}^3$. *SIAM J. Math. Anal.* **52**, 3990–4021 (2020)
21. Kuipers, L., Niederreiter, H.: *Uniform Distribution of Sequences*. Dover Publications Inc, Mineola (2006)
22. Monard, F., Stefanov, P.: Sampling the X-ray transform on simple surfaces. <http://arxiv.org/abs/2110.05761> (2021)
23. Natterer, F.: Sampling in fan beam tomography. *SIAM J. Appl. Math.* **53**, 358–380 (1993)
24. Orhan, K. (ed.): *Micro-computed Tomography (micro-CT) in Medicine and Engineering*. Springer Nature, Switzerland (2020)
25. Palamodov, V.P.: Localization of harmonic decomposition of the Radon transform. *Inverse Probl.* **11**, 1025–1030 (1995)
26. Quinto, E.T.: The dependence of the generalized Radon transforms on defining measures. *Trans. Am. Math. Soc.* **257**, 331–346 (1980)
27. Ramm, A., Katsevich, A.: *The Radon Transform and Local Tomography*. CRC Press, Boca Raton (1996)

28. Ramm, A.G., Zaslavsky, A.I.: Singularities of the Radon transform. *Bull. Am. Math. Soc.* **25**, 109–115 (1993)
29. Salo, M.: Applications of microlocal analysis in inverse problems. *Mathematics* **8**, 1184 (2020)
30. Sawano, Y.: *Theory of Besov Spaces*. Springer, Singapore (2018)
31. Stefanov, P.: Semiclassical sampling and discretization of certain linear inverse problems. *SIAM J. Math. Anal.* **52**, 5554–5597 (2020)
32. Stefanov, P.: The Radon transform with finitely many angles. <http://arxiv.org/abs/2208.05936v1> pp. 1–30 (2022)
33. Treves, F.: *Introduction to Pseudodifferential and Fourier Integral Operators. Volume 2: Fourier Integral Operators*. The University Series in Mathematics, Plenum, New York (1980)
34. Tuy, H.K.: An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.* **43**, 546–552 (1983)
35. Yang, H.: *Oscillatory Data Analysis and Fast Algorithms for Integral Operators*. PhD thesis, Stanford University (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.