Propagation of singularities for subelliptic wave equations

Cyril Letrouit

Propagation of singularities for subelliptic wave equations

Cyril Letrouit*

May 4, 2022

Abstract

Hörmander's propagation of singularities theorem does not fully describe the propagation of singularities in subelliptic wave equations, due to the existence of doubly characteristic points. In the present work, building upon a visionary conference paper by R. Melrose [30], we prove that singularities of subelliptic wave equations only propagate along null-bicharacteristics and abnormal extremals, which are well-known curves in optimal control theory. As a consequence, we characterize the singular support of subelliptic wave kernels outside the diagonal. These results show that abnormal extremals play an important role in the classical-quantum correspondence between sub-Riemannian geometry and sub-Laplacians.

Contents

1 Introduction **1**

2 The cones Γ_m **6**

3 A positive commutator **12**

4 Proof of Theorem 1 **14**

5 Proof of Corollary 3 **22**

6 Proof of Theorem 4 and Corollary 5 **25**

A Appendix **28**

1 Introduction

1.1 Motivations

In microlocal analysis, the celebrated propagation of singularities theorem describes an invariance property for the singularities of the (distributional) solutions of a general class of PDEs. The singularities are encapsulated in the C^∞-wave-front set (see (5)), whose projection is the singular support of the solution. Precisely, if u is a distributional solution to a partial (or pseudo-) differential equation $Pu = f$ and p is the principal symbol of P, assumed to be real and homogeneous, this theorem asserts that $WF(u) \setminus WF(f) \subset p^{-1}(0)$, and $WF(u) \setminus WF(f)$ is invariant under the Hamiltonian flow induced by p.

This result was first proved in [11, Theorem 6.1.1] and [18, Proposition 3.5.1]. Its second part about invariance of the wave-front set is however not totally satisfactory: it does not

*DMA, École normale supérieure, CNRS, PSL Research University, 75005 Paris (cyril.letrouit@ens.psl.eu)

provide any information in the case where the characteristics of P are not simple, i.e., at points of the cotangent bundle outside the null section for which $p = 0$ and $dp = 0$ (since at these points, the Hamiltonian vector field of p vanishes). In a very short and impressive conference paper [30], Melrose sketched the proof of an analogous propagation of singularities result for the wave operator $P = D_t^2 - A$ when A is a self-adjoint non-negative real second-order differential operator which is only subelliptic. Such operators P are typical examples for which there exist double characteristic points.

Despite the potential scope of this result, we did not find in the literature any other paper mentioning it, although several papers make reference to other results contained in [30]. The proof provided in [30] is very sketchy, and we thought it would deserve to be written in full details. This is what we do in the first part of the present work (Sections 2, 3 and 4).

Then, pushing further the computations of [30] in the case where A is a sub-Laplacian (see Definition 2), we prove in Sections 5, 6 and Appendix A.3 that singularities of subelliptic wave equations driven by sub-Laplacians only propagate along null-bicharacteristics and abnormal extremals, which are curves arising as optimal trajectories in control theory.

In summary, the different objects that will be involved in the statements are the following:

- The *bicharacteristic flow* is the flow induced by the Hamiltonian vector field of p. The Hamiltonian curves of p are called *bicharacteristics*.
- The *null-bicharacteristics* are the bicharacteristics that are included in $p^{-1}(0)$.
- The *null-rays* are a larger set of trajectories (in the cotangent bundle) along which the singularities propagate for subelliptic wave equations. Null-bicharacteristics are special instances of null-rays. When $P = D_t^2 - A$ with A elliptic, all null-rays are null-bicharacteristics. However, this is not always the case for A subelliptic. These null-rays are defined in Definition 9.
- The *abnormal extremals*, defined when A is a sub-Laplacian, are the only null-rays that are not null-bicharacteristics. See Section 5.1 for a precise definition.

1.2 Statements

We now state our main results. For the sake of coherence, we borrow nearly all notations to [30]. A is a self-adjoint non-negative real second-order differential operator on a smooth compact manifold X without boundary:

$$\forall u \in C^\infty(X), \qquad (Au, u) = (u, Au) \geqslant 0 \tag{1}$$

with

$$(u, v) = \int_X u(x)\overline{v(x)}d\nu, \tag{2}$$

where ν is some positive C^∞ density. The associated norm is denoted by $\|\cdot\|$.

We also assume that A is subelliptic, in the following sense: there exist a (Riemannian) Laplacian Δ on X and $c, s > 0$ such that

$$\forall u \in C^\infty(X), \qquad \|(-\Delta)^{s/2}u\|^2 \leqslant c((Au, u) + \|u\|^2). \tag{3}$$

Finally, we assume that A has vanishing subprincipal symbol.

The assumption (1) implies that A has a self-adjoint extension with the domain

$$\mathscr{D}(A) = \{u \in \mathcal{D}'(X);\ Au \in L^2(X)\}.$$

By the spectral theorem, for any $t \in \mathbb{R}$, the self-adjoint operator

$$G(t) = A^{-1/2}\sin(tA^{1/2})$$

is a well-defined operator bounded on $L^2(X)$, in fact it maps $L^2(X)$ into $\mathscr{D}(A^{1/2})$. Together with the self-adjoint operator $G'(t) = \cos(tA^{1/2})$, this allows to solve the Cauchy problem for the wave operator $P = D_t^2 - A$ where $D_t = \frac{1}{i}\partial_t$:

$$\begin{cases} D_t^2 u - Au = 0 & \text{in } \mathbb{R} \times X, \\ (u_{|t=0}, \partial_t u_{|t=0}) = (u_0, u_1) & \text{in } X \end{cases} \tag{4}$$

by

$$u(t,x) = G'(t)u_0 + G(t)u_1.$$

For $(u_0, u_1) \in \mathscr{D}(A^{1/2}) \times L^2(X)$, we have $u \in C^0(\mathbb{R}; \mathscr{D}(A^{1/2})) \cap C^1(\mathbb{R}; L^2(X))$.

For $f \in \mathcal{D}'(Y)$ a distribution on a manifold Y (equal to X, $\mathbb{R} \times X$ or $\mathbb{R} \times X \times X$ in the sequel), we denote by $WF(f)$ the usual Hörmander wave-front set (see [17]):

$$WF(f) = \{(y,\eta) \in T^*Y \setminus 0,\ \exists A \in \Psi^0_{\text{phg}}(Y), A \text{ elliptic at } (y,\eta) \text{ and } Af \in C^\infty(Y)\}. \tag{5}$$

Here an in all the sequel $T^*Y \setminus 0$ denotes the cotangent bundle from which the null-section has been removed. The set $\Psi^0_{\text{phg}}(Y)$ is the set of 0-th order polyhomogeneous pseudodifferential operators (see Appendix A.2). We also recall that the projection through the canonical projection onto Y of $WF(f)$ is the singular support of f.

Section 6 in [30] is a sketch of proof for the following statement which characterizes the propagation of singularities in (4):

Theorem 1. *Let A be a self-adjoint non-negative real second-order differential operator which is subelliptic* (3) *and has vanishing subprincipal symbol. Let $t \mapsto u(t)$ be a solution of* (4)*. For any $t > 0$, if $(x,\xi) \in WF(u(0))$ then there exists $(y,\eta) \in WF(u(-t)) \cup WF(\partial_t u(-t))$ such that (y,η) and (x,ξ) can be joined by a null-ray of length t.*

As mentioned above, null-rays will be defined in Definition 9.

An important class of examples of operators A satisfying (1), (2), (3) and with vanishing sub-principal symbol is given by sub-Laplacians (or Hörmander's sums of squares, see [38] or [25]):

Definition 2. *A sub-Laplacian is an operator of the form*

$$A = \sum_{i=1}^{K} Y_i^* Y_i \tag{6}$$

for some smooth vector fields Y_i on X satisfying Hörmander's condition: the Lie algebra generated by $Y_1, \dots, Y_K$ is equal to the whole tangent bundle TX. In (6)*, Y_i^* denotes the adjoint of Y_i for the scalar product* (2)*. The vector fields Y_i are not assumed to be independent.*

If A is a sub-Laplacian, the null-rays of Theorem 1 have a particularly simple geometric interpretation. In this case, denoting by a the principal symbol of A and by H_a the associated Hamiltonian vector field, Theorem 1 can be reformulated as follows (the notions of abnormal extremal lift, singular curve, and associated length are introduced in Section 5.1):

Corollary 3. *Assume that A is a sub-Laplacian as in Definition 2. Let $t \mapsto u(t)$ be a solution of* (4)*. For any $t > 0$, if $(x,\xi) \in WF(u(0))$ then there exists $(y,\eta) \in WF(u(-t)) \cup WF(\partial_t u(-t))$ such that (y,η) and (x,ξ) can be joined*

- *either by an Hamiltonian curve: $(x,\xi) = e^{tH_a}(y,\eta)$;*
- *or by an abnormal extremal lift of a singular curve of length $\geqslant t$.*

A weakness of Corollary 3 is that it only describes "from which region of phase space a singularity possibly comes", but does not assert that singularities effectively propagate along abnormal extremal lifts of singular curves. In particular, the inequality $\geqslant t$ in the last part of the statement means that singularities could possibly propagate at any speed between 0 and 1 along singular curves, but does not prove that it is effectively the case. In a joint

work with Yves Colin de Verdière [9], we give explicit examples of initial data of a subelliptic wave equation whose singularities effectively propagate at any speed between 0 and 1 along a singular curve. The sub-Laplacian which we use in [9] is called the Martinet sub-Laplacian (see Example 26). The propagation at speeds between 0 and 1 along singular curves is in strong contrast with the propagation "at speed 1" along the integral curves of H_a (as in Hörmander's theorem recalled above).

Theorem 1.8 in [30] (given without proof in [30], since the sketch of proof in [30, Section 6] in fact corresponds to Theorem 1), which we provide here only in the context of sub-Laplacians, concerns the Schwartz kernel K_G of G, i.e., the distribution $K_G \in \mathcal{D}'(\mathbb{R}\times X\times X)$ defined by

$$\forall u \in C^\infty(X), \qquad G(t)u(x) = \int_X K_G(t,x,y)u(y)dy. \tag{7}$$

Theorem 4. *Assume that A is a sub-Laplacian as in Definition 2. Then $WF(K_G)$ is contained in the set of $(t,x,y,\tau,\xi,-\eta) \in T^*(\mathbb{R}\times X\times X)\setminus 0$ such that the following two conditions are satisfied:*

(i) $\tau^2 = a(x,\xi) = a(y,\eta)$;

(ii) (y,η) *and* (x,ξ) *can be joined*

- *either by an Hamiltonian curve:* $(x,\xi) = e^{tH_a}(y,\eta)$;
- *or by an abnormal extremal lift of a singular curve of length* $\geqslant t$.

Theorem 4 will be deduced from Corollary 3 by considering K_G itself as a solution of a subelliptic wave equation. The projections on M of integral curves of H_a are called *normal geodesics*. By an adequate projection, we obtain the following corollary in the spirit of the Duistermaat-Guillemin trace formula [10]:

Corollary 5. *We fix $x,y \in X$ with $x \neq y$. We denote by $\mathscr{L}$ the set of lengths of normal geodesics from x to y and by T_s the minimal length of a singular curve joining x to y. Then $\mathscr{G} : t \mapsto K_G(t,x,y)$ is well-defined as a distribution on $(-T_s,T_s)$, and*

$$\mathrm{SingSupp}(\mathscr{G}) \subset \mathscr{L}\cup -\mathscr{L}.$$

Note that this corollary does not say anything about times $|t| \geqslant T_s$.

1.3 Comments, related literature and open questions

Null-rays. The null-rays which appear in the statement of Theorem 1 are generalizations of the usual null-bicharacteristics, which are the integral curves of the Hamiltonian vector field H_p of the principal symbol $p = \tau^2 - a$ of P contained in the characteristic set $p^{-1}(0)$. Null-rays are introduced in Definition 9, they are paths tangent to a family of convex cones Γ_m defined in Section 2.1.

For $m \in T^*(\mathbb{R}\times X)$ which is not in the double characteristic set $\{p=0\}\cap\{dp=0\}$, Γ_m is simply the positive (or negative) cone generated by H_p taken at point m, i.e., $\Gamma_m = \mathbb{R}^+\cdot H_p(m)$ (or $\Gamma_m = \mathbb{R}^-\cdot H_p(m)$, depending on whether $\tau \geqslant 0$ or $\tau \leqslant 0$). In the double characteristic set $\Sigma_{(2)} = \{p=0\}\cap\{dp=0\} \subset M$, the definition of the cones Γ_m is more involved, and several formulas will be provided in Section 2. We can already say (see Appendix A.3) that Γ_m can be recovered as the convexification of the limits of all cones Γ_{m_j} for $m_j \to m$ and $m_j \notin \Sigma_{(2)}$.

Abnormal extremals. Roughly speaking, Corollary 3, Theorem 4 and Corollary 5 strengthen the idea that properties of general sub-Laplacians may be influenced not only by the geometry of null-bicharacteristics but also by the presence of abnormal extremal lifts of singular curves in the corresponding sub-Riemannian geometry; in other words, the latter curves play a role at the "quantum level" of general sub-Laplacians.

This role was already foreboded in a particular case in the work of Richard Montgomery [33] about zero loci of magnetic fields, and it is central in the Treves conjecture about

hypoelliptic analyticity (see [42] for the conjecture and [2] for recent results). To the author's knowledge, it is the first result which illustrates this fact for *general* sub-Laplacians.

Related literature. As a particular case of Theorem 1, if A is elliptic, then we recover Hörmander's result [18, Proposition 3.5.1] already mentioned above (see also [19, Theorem 8.3.1 and Theorem 23.2.9] and [27, Theorem 1.2.23]). In case A has only double characteristics on a symplectic submanifold it was obtained in [29] in codimension 2, and by B. and R. Lascar [23], [24] in the general case, using constructions of parametrices instead of positive commutator estimates as used in [30] (see also Remark 30). We also mention the paper [39] where a result about propagation of singularities in the case of "quasi-contact" sub-Laplacians is proved. The subelliptic wave propagator has also recently been studied in [28] to prove spectral multiplier estimates, but the construction is restricted to the "elliptic part" of the symbol where $a \neq 0$.

In [30], two other results are proved, namely the finite speed of propagation for P and an estimate on the heat kernel, but it is not our purpose to discuss here these other results, whose proofs are written in details in [30].

Open questions. Here are a few natural questions that our work could help to answer:

- Is it possible to find explicitly the form of the wave kernel K_G (and not only its singularities as in Theorem 4), even for simple sub-Laplacians? This would generalize the Hadamard parametrix to subelliptic wave equations. The only known cases are apparently the Heisenberg case (see [35], [40], [15]), the contact case ([23], [24], [29]) and the quasi-contact case [39].
- The answer to the above question could pave the way towards a better understanding of the *subelliptic heat kernel* in the presence of abnormal extremals, thanks to the "transmutation formula" sometimes attributed to Y. Kannai [22] (see also [6]). Indeed, as proved in [26], the subelliptic heat kernel $p_t(x,y)$ enjoys the asymptotics $2t\mathrm{Log}(p_t(x,y)) \to d(x,y)^2$ as $t \to 0$ where $d(x,y)$ is the sub-Riemannian distance (see also [21]), but the refinement of this limit into a full asymptotic development of $p_t(x,y)$ is known only in the absence of abnormal extremals minimizing the distance (see [5]). Note that subelliptic heat kernels have been studied a lot, see for instance [4] for one of the last major achievements in this field.
- Is it possible to prove an analogue of Egorov's theorem in the framework of sub-Riemannian geometry? The usual formulation of Egorov theorem is that the evolution of a quantum observable $\mathrm{Op}_\hbar(b)$ is analogous to the evolution of the corresponding classical observable:

$$U_H^{-t}\mathrm{Op}_\hbar(b)U_H^t = \mathrm{Op}_\hbar(b\circ\varphi_H^t) + O_t(\hbar), \quad \text{where} \quad U_H^t = e^{-\frac{it\hat{H}_h}{\hbar}}, \quad \hat{H}_h = -\frac{\hbar^2}{2m}\Delta$$

  and φ_H^t is the Hamiltonian flow associated to the principal symbol H of Δ (see for example [37, Theorem IV-10]). Analogues of Egorov's theorem in the subelliptic framework have been derived for instance in [8] and [14], but only in cases where there are no abnormal extremals. Would abnormal extremals play a role in a general subelliptic version of Egorov's theorem, as in the present work?
- Is it possible to design a physical experiment with electrons in a magnetic field which would illustrate the propagation along singular curves as in Corollary 3?[1] Indeed, certain singular curves appear naturally as zero loci of magnetic fields, and high-frequency quantum particles tend to concentrate on such curves, see [33]. R. Montgomery says in [33] that "singular curves persist upon quantization".

[1] Thanks to R. Montgomery for suggesting this question.

1.4 Organization of the paper

The goal of this work is to provide a fully detailed proof of Theorem 1, Corollary 3, Theorem 4 and Corollary 5 and to explain how these results are related to sub-Riemannian geometry.

In Section 2, we define the convex cones Γ_m generalizing bicharacteristics and give explicit formulas for them, then prove their semi-continuity with respect to m, and finally introduce null-rays and "time functions". These functions are by definition non-increasing along the cones Γ_m. In this section, there is no operator, we work at a purely "classical" level.

The proof of Theorem 1 is based on a positive commutator argument: the idea, which dates back at least to [18] (see also [20, Chapter I.2]), is to derive an *energy inequality* from the computation of a quantity of the form $\mathrm{Im}(Pu, Lu)$, where L is some well-chosen (pseudodifferential) operator. In Section 3, we compute this quantity for $L = \mathrm{Op}(\Phi)D_t$ where Φ is a time function, we write it under the form $\frac{1}{2}(Cu, u)$ for an explicit second-order operator C which, up to remainder terms, has non-positive symbol.

In Section 4, we derive from this computation the sought energy inequality, which in turn implies Theorem 1. This proof requires to construct specific time functions and to use the powerful Fefferman-Phong inequality [13].

In Section 5, we prove Corollary 3. This requires to explain basic concepts of sub-Riemannian geometry, notably we define normal geodesics, singular curves, and abnormal extremal lifts.

In Section 6, we prove Theorem 4: the main idea is to see K_G itself as the solution of a subelliptic wave equation. We also prove Corollary 5 in the same section.

The reader will find in Appendix A.1 the sign conventions for symplectic geometry that we use throughout this note, and a short reminder on pseudodifferential operators in Appendix A.2. Finally, in Appendix A.3, we explain how the cones Γ_m can be defined in a unified way as Clarke generalized gradients, thus making a bridge between our computations and Clarke's version of Pontryagin's maximum principle.

Acknowledgments. I am very grateful to Yves Colin de Verdière, for his help at all stages of this work. Several ideas, notably in Sections 6, are due to him. I also thank him for having first showed me R. Melrose's paper and for his constant support along this project, together with Emmanuel Trélat. I am also thankful to Andrei Agrachev, Richard Lascar and Nicolas Lerner for very interesting discussions related to this paper. Finally, many thanks are due to the referee whose suggestions improved the readability of the paper.

2 The cones Γ_m

At double characteristic points where in particular $dp = 0$, the Hamiltonian vector field H_p vanishes, and the usual propagation of singularities result [11, Theorem 6.1.1] recalled in Section 1.1 does not provide any information. In [30], R. Melrose defines convex cones Γ_m which replace the usual propagation cone $\mathbb{R}^+ \cdot H_p$ at these points, and which indicate the directions in which singularities of the subelliptic wave equation (4) may propagate. The cones Γ_m can be equivalently defined

- symplectically (Section 2.1) - this is the most synthetic definition;
- with analytic formulas (Section 2.2) - very useful for proofs;
- with the Clarke generalized gradient (Appendix A.3) - which gives a clear geometric insight, although it is not useful in our proofs.

The cones Γ_m have particularly simple expressions in case A is a sub-Laplacian as in Definition 2. These expressions are given along the proof of Corollary 3 in Section 5.2, and also linked with the "Clarke version" of the Pontryagin maximum principle in Appendix A.3.

2.1 First definition of the cones Γ_m

In this section, we introduce several notations, and we define the cones Γ_m.

We consider $a \in C^\infty(T^*X)$ satisfying

$$a(x,\xi) \geqslant 0, \quad a(x,r\xi) = r^2 a(x,\xi), \quad r > 0 \tag{8}$$

in canonical coordinates (x,ξ). Also we consider

$$p = \tau^2 - a \in C^\infty(M), \qquad \text{where} \quad M = T^*(\mathbb{R}\times X)\setminus 0.$$

In the end, a and p will be the principal symbols of the operators A and P introduced in Section 1, but for the moment we work at a purely classical level and forget about operators. In a nutshell, at points where $p = 0$ and $dp = 0$, the cones Γ_m are defined thanks to the Hessian of p.

We set

$$M_+ = \{m \in M,\ p(m) \geqslant 0, \tau \geqslant 0\}, \qquad M_- = \{m \in M,\ p(m) \geqslant 0, \tau \leqslant 0\};$$

in particular, $M_+ \cup M_- = \{p \geqslant 0\}$. Let

$$\Sigma = \{m \in M;\ p(m) = 0,\ \tau \geqslant 0\}.$$

The cones Γ_m for $m \in M_+$. For $m \in M_+$, we consider the set

$$\mathscr{H}_m = \mathbb{R}^+ \cdot H_p(m) \subset T_m M,$$

where H_p is the Hamiltonian vector field of p verifying $\omega(H_p, Z) = -dp(Z)$ for any smooth vector field Z. In this formula and in all the sequel, ω is the canonical symplectic form on the cotangent bundle M.

We note that

$$p(m) \geqslant 0,\ dp(m) = 0 \ \Rightarrow\ \mathscr{H}_m = \{0\}$$

where $dp(m)$ stands for the differential of p taken at point m. We therefore extend the notion of "bicharacteristic direction" at such m. This will be done first for $m \in M_+$, then also for $m \in M_-$, but never for $m \in \{p < 0\}$: the cones Γ_m are not defined for points $m \in \{p < 0\}$.

Let

$$\Sigma_{(2)} = \{m \in M, \tau = a = 0\} \subset \Sigma.$$

Note that

$$\Sigma_{(2)} = M_+ \cap M_- = \{m \in M;\ \tau = a = p = 0,\ da = dp = 0\}$$

due to the the positivity (8). Thus, for $m \in \Sigma_{(2)}$, the Hessian of a is well-defined: it is a quadratic form on $T_m M$. We denote by $a_m = \frac{1}{2}\mathrm{Hess}\, a$ *the half of this Hessian*, and by

$$p_m = (d\tau)^2 - a_m$$

the half of the Hessian of p. For $m \in \Sigma_{(2)}$, we set

$$\Lambda_m = \{w \in T_m M;\ d\tau(w) \geqslant 0,\ p_m(w) \geqslant 0\} \tag{9}$$

and, still for $m \in \Sigma_{(2)}$,

$$\Gamma_m := \{v \in T_m M;\ \omega(v,w) \leqslant 0 \ \ \forall w \in \Lambda_m\}. \tag{10}$$

If $m \in M_+ \setminus \Sigma_{(2)}$, we set

$$\Gamma_m = \mathscr{H}_m. \tag{11}$$

In particular, the cones Γ_m are defined also at points m outside Σ, i.e. for which $p(m) \neq 0$. Note also that the relation (11) says that the cones Γ_m are *positive* cones.

The cones Γ_m for $m \in M_-$. In order to extend the definition of the cones Γ_m to M_-, we want this extension to be consistent with the previous definition at points in $M_+ \cap M_- = \Sigma_{(2)}$. We observe that M_- is the image of M_+ under the involution sending τ to $-\tau$. For $(t, \tau, \alpha) \in M_-$, we set

$$\Gamma_m = \Gamma_{m'} \qquad \text{where} \qquad m' = (t, -\tau, \alpha) \in M_+.$$

It is clear that at points of $M_+ \cap M_- = \Sigma_{(2)}$, the two definitions of Γ_m coincide. With this definition in M_-, note that for $m \in M_- \setminus \Sigma_{(2)}$, there is a sign change:

$$\Gamma_m = -\mathscr{H}_m. \tag{12}$$

In summary, the formulas (10), (11) and (12) define Γ_m at any point $m \in M_+ \cup M_-$, with different definitions for $m \in \Sigma_{(2)}$, $m \in M_+ \setminus \Sigma_{(2)}$ and $m \in M_- \setminus \Sigma_{(2)}$. The cones Γ_m are not defined for $m \notin M_+ \cup M_-$. For any $m \in M_+ \cup M_-$, the cone Γ_m is closed and convex.

2.2 Formulas for the cones Γ_m

In this section, we derive a formula for the cones Γ_m when $m \in \Sigma_{(2)}$ which is more explicit than (10). It relies on the computation of the polar of a cone defined by a non-negative quadratic form:

Proposition 6. *Let Q be a non-negative quadratic form on a real vector space V, and let*

$$\Theta = (\ker(Q))^\perp \subset V^*$$

where $\perp$ is understood in the duality sense and V^* is the topological dual of V. Let

$$\Lambda = \left\{ \xi = (\xi_0, \eta) \in \mathbb{R} \times V;\ \xi_0 \geqslant Q(\eta)^{\frac{1}{2}} \right\}$$

and

$$\Lambda^0 = \{ \xi' \in (\mathbb{R} \times V)^*;\ \forall \xi \in \Lambda,\ \xi'(\xi) \leqslant 0 \}.$$

Then

$$\Lambda^0 = \left\{ \xi' = (\xi'_0, \eta') \in (\mathbb{R} \times V)^*;\ \eta' \in \Theta \text{ and } -\xi'_0 \geqslant (Q^*(\eta'))^{\frac{1}{2}} \right\} \tag{13}$$

where $\mathbb{R}^*$ is identified with $\mathbb{R}$ and

$$Q^*(\eta') = \sup_{\eta \notin \ker(Q)} \frac{\eta'(\eta)^2}{Q(\eta)}. \tag{14}$$

Proof. Let $\xi' = (\xi'_0, \eta') \in (\mathbb{R} \times V)^*$ such that $\eta' \in \Theta$ and $-\xi'_0 \geqslant (Q^*(\eta'))^{\frac{1}{2}}$, we seek to prove that $\xi' \in \Lambda^0$. Let $\xi = (\xi_0, \eta) \in \Lambda$. In particular, $\xi_0 \geqslant (Q(\eta))^{\frac{1}{2}}$. We have

$$\xi'(\xi) = \xi'_0(\xi_0) + \eta'(\eta) \leqslant -(Q^*(\eta'))^{\frac{1}{2}} (Q(\eta))^{\frac{1}{2}} + \eta'(\eta) \leqslant 0$$

hence $\xi' \in \Lambda^0$, which proves one inclusion.

Conversely, to prove that Λ^0 is included in the expression (13), we first note that if $\eta' \notin \Theta$, then $(\xi'_0, \eta') \notin \Lambda^0$ for any $\xi'_0 \in \mathbb{R}^*$. Indeed, if $\eta' \notin \Theta$, there exists $\eta \in V$ such that $Q(\eta) = 0$ and $\eta'(\eta) > 0$. Thus, considering $\xi = (0, \eta)$, which is in Λ by assumption, we get $\xi'(\xi) = \eta'(\eta) > 0$ for any $\xi'_0 \in \mathbb{R}^*$ and $\xi' = (\xi'_0, \eta')$, proving that $\xi' \notin \Lambda^0$. Now, if $\xi' = (\xi'_0, \eta') \in \Lambda^0$ with $\eta' \in \Theta$, we take $\xi_n = (\xi_{0n}, \eta_n)$ with $\eta_n \notin \ker(Q)$ so that $\eta'(\eta_n)^2 / Q(\eta_n) \to Q^*(\eta')$, and $\eta'(\eta_n) \geqslant 0$ and $\xi_{0n} = Q(\eta_n)^{\frac{1}{2}}$. Then $\xi_n \in \Lambda$. Therefore, $\xi'(\xi_n) \leqslant 0$, which implies that $-\xi'_0 \geqslant (Q^*(\eta'))^{\frac{1}{2}}$. This proves the result. □

Applying the previous proposition to $Q = a_m$ yields a different definition of the cones Γ_m. First, Λ_m, which has been defined in (9), can be written as

$$\Lambda_m = \left\{ w \in T_m M;\ d\tau(w) \geqslant (a_m(w))^{\frac{1}{2}} \right\}.$$

Since the definition of Λ_m does not involve dt, we have $v(\partial_t) = 0$ for any $v \in \Lambda_m^0$. Now, using the notation a_m^* to denote (14) when $Q = a_m$, Proposition 6 yields that

$$\begin{aligned} \Lambda_m^0 &= \mathbb{R}^+(-d\tau + B_0), \\ B_0 &= \left\{ b_0 \in (\ker(a_m))^\perp,\ a_m^*(b_0) \leqslant 1 \right\}. \end{aligned}$$

The duality $\perp$ is computed with respect to the space $\ker(a_m) \subset T(T^*X)$, i.e., $b_0 \in T^*(T^*X)$.

Here, we consider a_m as a quadratic form on $T(T^*X)$ instead of $M = T(T^*(\mathbb{R} \times X))$. This is also related to the fact that the Hessian a_m depends only on the projection $\pi_2(m)$, where $\pi_2 : M \to T^*X$ is the canonical projection on the second factor, and not on really on the other components of m. *These slight abuse of notations cause no problem, and will thus be repeated several times in the sequel.*

Comparing the definition of Λ_m^0 as the polar cone of Λ_m and the definition (10) of Γ_m, we see that Γ_m is exactly the image of Λ_m^0 through the canonical isomorphism $\omega(v, \cdot) \mapsto v$ between T_m^*M and $T_m M$. Thus,

$$\begin{aligned} \Gamma_m &= \mathbb{R}^+(\partial_t + B), \\ B &= \left\{ b \in \ker(a_m)^{\perp_{\omega_X}},\ a_m^*(\mathcal{I}(b)) \leqslant 1 \right\}. \end{aligned} \tag{15}$$

Here, $\perp_{\omega_X}$ designates the symplectic orthogonal with respect to the canonical symplectic form ω_X on T^*X and

$$\begin{aligned} \mathcal{I} : T(T^*X) &\to T^*(T^*X), \\ b &\mapsto \omega_X(b, \cdot) \end{aligned} \tag{16}$$

is the canonical isomorphism between $T(T^*X)$ and $T^*(T^*X)$. Formula (15) plays a key role in the sequel. An equivalent formula in terms of the so-called "fundamental matrix" associated to a_m is derived in Appendix A.3.

Important remark. In case $A = \sum_{i=1}^K Y_i^* Y_i$ is a sub-Laplacian as in Definition 2, the expression $a_m^*(\mathcal{I}(b))$ which appears in (15) is equal to $g(d\pi(b))$ where $\pi : T^*X \to X$ is the canonical projection, and g is the sub-Riemannian metric associated to the vector fields Y_i (see Lemma 28).

2.3 Inner semi-continuity of the cones Γ_m

Using the formula (15), we can prove a continuity property for the cones Γ_m.

Lemma 7. *Let $a \in C^\infty(T^*X)$ satisfying (8). The assignment $m \mapsto \Gamma_m$ is inner semi-continuous on $M_+ \cup M_- = \{p \geqslant 0\}$. In other words, if both conditions*

(i) $m_j \in M_+ \cup M_-$ for any $j \in \mathbb{N}$, and $m_j \to m$ as $j \to +\infty$;

(ii) $v_j \in \Gamma_{m_j}$ for any $j \in \mathbb{N}$, and $v_j \to v \in T_m M$ as $j \to +\infty$,

hold, then $v \in \Gamma_m$.

Before proving Lemma 7, let us explain the intuition behind this semi-continuity. Recall that the cones Γ_m generalize bicharacteristic directions at points where $\tau = a = p = 0$ and $da = dp = 0$. To define the cones Γ_m at these points, following formulas (9) and (10), we have first considered directions where p grows (since $p = 0$ and $dp = 0$, we consider the (half) Hessian p_m), yielding Λ_m, and then Γ_m has been defined as the (symplectic) polar cone of Λ_m. This is exactly parallel to a procedure which yields bicharacteristic directions in the non-degenerate case: the directions along which p grows, verifing $dp(v) \geqslant 0$, form a cone,

and it is not difficult to check that its (symplectic) polar consists of a single direction given by the Hamiltonian vector field of p. This is a unified vision of the cones Γ_m, in the sense that they are obtained in a unified way, no matter whether $m \in \Sigma_{(2)}$ or not. The proof of Lemma 7 we give below is however purely analytic, and does not use this geometric intuition. Note that Appendix A.3 provides still another unified vision of the cones Γ_m, thanks to the notion of Clarke generalized gradients.

Proof of Lemma 7. The assignments

$$\Sigma_{(2)} \ni m \mapsto \Gamma_m \quad \text{and} \quad M_+ \cup M_- \setminus \Sigma_{(2)} \ni m \mapsto \Gamma_m$$

are clearly continuous thanks to formula (10) (resp. (11) and (12)). Therefore, we restrict to the case where $m \in \Sigma_{(2)}$ and $m_j \in M_+ \cup M_- \setminus \Sigma_{(2)}$.

According to (11) and (12), the cone Γ_{m_j} at $m_j = (t_j, \tau_j, x_j, \xi_j)$ is given by:

$$\Gamma_{m_j} = \mathrm{sgn}(\tau_j)\mathbb{R}^+[2\tau_j\partial_t - H_a(m_j)] \tag{17}$$

where $H_a(m_j)$ is the Hamiltonian vector field of a at m_j. Dividing by $2\tau_j \neq 0$, we rewrite it as

$$\Gamma_{m_j} = \mathbb{R}^+\left(\partial_t - \frac{1}{2}\frac{H_a(m_j)}{\tau_j}\right) \tag{18}$$

If $a(m_j) = 0$, then $H_a(m_j) = 0$ since a is a quadratic form. Thus $\Gamma_{m_j} = \mathbb{R}^+\partial_t$, thus any limit of elements of Γ_{m_j} is contained in Γ_m according to (15) (take $b = 0$). We thus assume that $a(m_j) \neq 0$.

In the sequel, we work in a chart near m and $\|\cdot\|_M$ denotes a norm in this chart. We recall that a_m is half the Hessian of a at m, thus a bilinear form on T_mM. When its two arguments are $m_j - m$ (which we can view as an element of T_mM now that we are working in a chart), it is written $a_m(m_j - m)$. Also, in the sequel the notation o accounts for the $j \to +\infty$ limit.

We distinguish two cases.

Firstly, if $a_m(m_j - m) = o(\|m_j - m\|_M^2)$, then using a Taylor expansion and $a(m) = da(m) = 0$, we get $a(m_j)^{\frac{1}{2}} = o(\|m_j - m\|_M)$. Since a is a quadratic form, this implies that $d(a^{\frac{1}{2}})(m_j) = o(1)$, where the notation in the left-hand side stands for the differential of $a^{\frac{1}{2}}$ taken at point m_j. In turn, we obtain

$$\frac{1}{2}\frac{da(m_j)}{\tau_j} = \frac{a(m_j)^{\frac{1}{2}}}{\tau_j}d(a^{\frac{1}{2}})(m_j) = o(1)$$

since $|\tau_j| \geqslant a(m_j)^{\frac{1}{2}}$ due to $p(m_j) \geqslant 0$. This implies that $H_a(m_j)/\tau_j \to 0$ and plugging into (18), we conclude that the limiting directions of vectors v_j of Γ_{m_j} as $j \to +\infty$ belong to $\mathbb{R}^+\partial_t$ and thus to Γ_m.

Secondly, if $a_m(m_j - m)$ is not $o(\|m_j - m\|_M^2)$, then we use the following lemma.

Lemma 8. *If $a_m(m_j - m)$ is not $o(\|m_j - m\|_M^2)$, then for any $v \in T_mM$, there holds*

$$\frac{1}{2}\frac{da(m_j)(v)}{a(m_j)^{1/2}} = \frac{a_m(m_j - m, v)}{a_m(m_j - m)^{1/2}} + o(1).$$

The notation $a_m(m_j - m, v)$ means that we evaluate the bilinear form a_m at $(m_j - m, v)$.

Proof. In a chart, we combine the two expansions

$$da(m_j)(v) = 2a_m(m_j - m, v) + o(\|m_j - m\|_M)$$
$$a(m_j) = a_m(m_j - m) + o(\|m_j - m\|_M^2)$$

to get the result. □

In view of (18) and (15), the inner semi-continuity at m is equivalent to proving that

$$a_m^* \left(\frac{1}{2} \frac{a(m_j)^{\frac{1}{2}}}{\tau_j} \frac{da(m_j)}{a(m_j)^{\frac{1}{2}}} \right) \leqslant 1 + o(1). \tag{19}$$

Using $|\tau_j| \geqslant a(m_j)^{\frac{1}{2}}$ and Lemma 8, for any $v \in T_m M \setminus \ker(a_m)$, there holds

$$\frac{1}{a_m(v)} \left(\frac{1}{2} \frac{a(m_j)^{\frac{1}{2}}}{\tau_j} \frac{da(m_j)(v)}{a(m_j)^{\frac{1}{2}}} \right)^2 \leqslant \frac{a_m(m_j - m, v)^2}{a_m(v) a_m(m_j - m)} + o(1) \leqslant 1 + o(1)$$

by Cauchy-Schwarz. Hence, by definition of a_m^* (see (14)), (19) holds, which concludes the proof of Lemma 7.

2.4 Null-rays and time functions

We now define null-rays, which are the integral curves of the cone field Γ_m. They appear in the statement of Theorem 1, and they play an important role in the present work.

In this definition, we use the following notation: given a Lipschitz curve $\gamma : I \to M_+$ defined on some interval $I \subset \mathbb{R}$, the set-valued derivative $\gamma'(s)$ for $s \in I$ is the set of all tangent vectors $X \in T_{\gamma(s)} M$ such that there exists $s_n \to s$ with $s_n \neq s$ for any $n \in \mathbb{N}$, verifying that $\forall f \in C^\infty(M)$, $\frac{f(\gamma(s_n)) - f(\gamma(s))}{s_n - s} \to Xf$.

Definition 9. *A forward-pointing ray for p is a Lipschitz curve $\gamma : I \to M_+$ defined on some interval $I \subset \mathbb{R}$ with (set-valued) derivative $\gamma'(s) \subset \Gamma_{\gamma(s)}$ for all $s \in I$. Such a ray is forward-null if $\gamma(s) \in \Sigma$ for any $s \in I$. We define backward-pointing rays similarly, with γ valued in M_-, and backward-null rays, with γ valued in $\{m \in M;\ p(m) = 0,\ \tau \leqslant 0\}$.*

Under the terminology "*ray*", we mean either a forward-pointing or a backward-pointing ray; under the terminology "*null-ray*", we mean either a forward-null or a backward-null ray.

In particular null-rays live in $\{p = 0\}$.

Fixing a norm $|\cdot|$ on TM, the expression (15) implies that near any point $m \in M_+$, there is a (locally) uniform constant $c > 0$ such that

$$v \in \Gamma_m \Rightarrow v = T\partial_t + v', \qquad |v'| \leqslant cT \tag{20}$$

where v' is tangent to T^*X. Thus, if $\gamma : I \to M_+$ is a forward-pointing ray (thus a Lipschitz curve) defined for $s \in I$, (20) implies that $dt/ds \geqslant c' |d\gamma/ds|$, hence

$$\frac{d\gamma}{dt} = \frac{\frac{d\gamma}{ds}}{\frac{dt}{ds}}$$

is well-defined (possibly set-valued), i.e., γ can be parametrized by t.

We define the *length* of a ray $\gamma : s \in [s_0, s_1] \to M_+$ by

$$\ell(\gamma) := |t(s_1) - t(s_0)|.$$

Remark 10. *Thanks to the above parametrization and with a slight abuse in the terminology, we say that there is a null-ray of length $|T|$ from (y, η) to (x, ξ) if there exists a null-ray (in the sense of Definition 9) parametrized by t which joins $(0, \tau, y, \eta)$ to (T, τ, x, ξ), where τ verifies $\tau^2 = a(y, \eta) = a(x, \xi)$.*

Time functions, which we now introduce, are one of the key ingredients of the proof of Theorem 1.

Definition 11. *A C^∞ function ϕ near $\overline{m} \in \{p \geqslant 0\} \subset M$ is a time function near $\overline{m}$ if in some neighborhood N of $\overline{m}$,*

$$\phi \text{ is non-increasing along } \Gamma_m,\ m \in N \cap \{p \geqslant 0\}.$$

In particular, ϕ is non-increasing along the Hamiltonian vector field H_p in M_+ but non-decreasing along H_p in M_- (due to (12)).

Note that outside $\{p \geqslant 0\}$, there is no constraint on the values of ϕ. The proof of Theorem 1 relies on a positive commutator technique (Section 3) applied with a particular time function (Section 4).

3 A positive commutator

The proof of Theorem 1 is based on a "positive commutator" technique, also known as "multiplier" or "energy" method in the literature. The idea is to derive an inequality from the computation of a quantity of the form $\mathrm{Im}(Pu, Lu)$ where L is some well-chosen (pseudodifferential) operator. In the present work, the operator L is related to the time functions introduced in Definition 11.

In the sequel, we use polyhomogeneous symbols, denoted by S^m_{phg}, and the *Weyl quantization*, denoted by $\mathrm{Op} : S^m_{\mathrm{phg}} \to \Psi^m_{\mathrm{phg}}$ (see Appendix A.2). For example, we consider the operator $D_t = \frac{1}{i}\partial_t = \mathrm{Op}(\tau)$ (of order 1). The operator $A \in \Psi^2_{\mathrm{phg}}$ has principal symbol $a \in C^\infty(T^*X)$ satisfying (8), and $P = D_t^2 - A$ has principal symbol $p = \tau^2 - a$.

Also, $\Phi(t, x, \xi)$ designates a smooth *real-valued* function on M, homogeneous of degree $\alpha \in \mathbb{R}$ in ξ, compactly supported on the base $\mathbb{R} \times X$, and independent of τ. In Section 4, we will take Φ to be a time function. By the properties of the Weyl quantization, $\mathrm{Op}(\Phi)$ is a compactly supported selfadjoint (with respect to ν) pseudodifferential operator of order α.

Our goal in Section 3.1 will be to compute C defined by[2]

$$\mathrm{Im}(Pu, \mathrm{Op}(\Phi)D_t u) := \frac{1}{2}(Cu, u), \tag{21}$$

since this will allow us to derive the inequality (49) which is the main ingredient in the proof of Theorem 1.

3.1 The operator C

Our goal in this section is to compute C defined by (21).

Lemma 12. *We have*

$$C = D_t\mathrm{Op}(\Phi'_t)D_t - \frac{i}{2}([A, \mathrm{Op}(\Phi)]D_t + D_t[A, \mathrm{Op}(\Phi)]) + \frac{1}{2}(A\mathrm{Op}(\Phi'_t) + \mathrm{Op}(\Phi'_t)A). \tag{22}$$

where $\Phi'_t = \partial_t\Phi$.

Note that C is of order $2 + \alpha$, although we could have expected order $3 + \alpha$ by looking too quickly at (21).

Proof. We have

$$\mathrm{Im}(Pu, \mathrm{Op}(\Phi)D_t u) = I_1 - I_2 \tag{23}$$

with

$$I_1 = \mathrm{Im}(D_t^2 u, \mathrm{Op}(\Phi)D_t u) \qquad \text{and} \qquad I_2 = \mathrm{Im}(Au, \mathrm{Op}(\Phi)D_t u).$$

Noticing that

$$[D_t, \mathrm{Op}(\Phi)] = \mathrm{Op}(\frac{1}{i}\Phi'_t)$$

[2] In [30], C is explicitly defined as $\mathrm{Im}(\mathrm{Op}(\Phi)D_t u, Pu) := (Cu, u)$; however the formulas (6.1) and (6.2) in [30] are not coherent with this definition, but they are correct if we take the definition (21) for C.

(see [43, Theorem 4.6]), we have for I_1:

$$\begin{aligned} I_1 &= \frac{1}{2i}\left((D_t^2 u, \mathrm{Op}(\Phi) D_t u) - (\mathrm{Op}(\Phi) D_t u, D_t^2 u)\right) \\ &= \frac{1}{2i}\left((D_t \mathrm{Op}(\Phi) D_t^2 u, u) - (D_t^2 \mathrm{Op}(\Phi) D_t u, u)\right) \\ &= -\frac{1}{2i}(D_t [D_t, \mathrm{Op}(\Phi)] D_t u, u) \\ &= -\frac{1}{2i}(D_t \frac{1}{i} \mathrm{Op}(\Phi_t') D_t u, u) \\ &= \frac{1}{2}(D_t \mathrm{Op}(\Phi_t') D_t u, u) \end{aligned} \tag{24}$$

Then, we write $\mathrm{Op}(\Phi) D_t = S + iT$ where

$$\begin{aligned} S &= \frac{1}{2}(\mathrm{Op}(\Phi) D_t + D_t \mathrm{Op}(\Phi)) \\ T &= \frac{1}{2i}(\mathrm{Op}(\Phi) D_t - D_t \mathrm{Op}(\Phi)) = \frac{1}{2}\mathrm{Op}(\Phi_t'). \end{aligned} \tag{25}$$

Using that A, S and T are selfadjoint, we compute I_2:

$$\begin{aligned} I_2 &= \mathrm{Im}(Au, (S+iT)u) = \mathrm{Im}((S-iT)Au, u) = \frac{1}{2i}([S, A]u, u) - \mathrm{Re}((TAu, u)) \\ &= \frac{1}{2i}([S, A]u, u) - \frac{1}{2}((TA + AT)u, u). \end{aligned} \tag{26}$$

Furthermore,

$$[S, A] = \frac{1}{2}([\mathrm{Op}(\Phi), A] D_t + D_t [\mathrm{Op}(\Phi), A]). \tag{27}$$

All in all, combining (23), (24), (25), (26) and (27), we find that C in (21) is given by (22). □

3.2 The principal and subprincipal symbols of C

In this section, we compute the operator C modulo a remainder term in $\Psi^{\alpha}_{\mathrm{phg}}$. All symbols and pseudodifferential operators used in the computations are polyhomogeneous (see Appendix A.2); we denote by $\sigma_p(C), \sigma_{\mathrm{sub}}(C)$ the principal symbol of C and its sub-principal symbol. We use the Weyl quantization, denoted by Op, in the variables $y = (t, x)$, $\eta = (\tau, \xi)$, hence we have for any $b \in S^m_{\mathrm{phg}}$ and $c \in S^{m'}_{\mathrm{phg}}$:

$$\mathrm{Op}(b)\mathrm{Op}(c) - \mathrm{Op}(bc + \frac{1}{2i}\{b, c\}) \in \Psi^{m+m'-2}_{\mathrm{phg}} \tag{28}$$

and

$$[\mathrm{Op}(b), \mathrm{Op}(c)] - \mathrm{Op}(\frac{1}{i}\{b, c\}) \in \Psi^{m+m'-3}_{\mathrm{phg}}. \tag{29}$$

Note that in (29), the remainder is in $\Psi^{m+m'-3}_{\mathrm{phg}}$, and not only in $\Psi^{m+m'-2}_{\mathrm{phg}}$ (see [19, Theorem 18.5.4], [43, Theorem 4.12]). Finally, we recall that $\Phi(t, x, \xi)$ is homogeneous in ξ of degree α.

Lemma 13. *There holds*

$$\sigma_p(C) = \tau H_p \Phi - \Phi_t' p \tag{30}$$

and

$$\sigma_{sub}(C) = 0. \tag{31}$$

Proof. We compute each of the terms in (22) modulo $\Psi^{\alpha}_{\mathrm{phg}}$. We prove the following formulas:

$$\frac{1}{2}(A\mathrm{Op}(\Phi'_t) + \mathrm{Op}(\Phi'_t)A) = \mathrm{Op}(a\Phi'_t) \mod \Psi^{\alpha}_{\mathrm{phg}} \tag{32}$$

$$D_t\mathrm{Op}(\Phi'_t)D_t = \mathrm{Op}(\tau^2\Phi'_t) \mod \Psi^{\alpha}_{\mathrm{phg}} \tag{33}$$

$$\frac{i}{2}([A,\mathrm{Op}(\Phi)]D_t + D_t[A,\mathrm{Op}(\Phi)]) = \mathrm{Op}(\tau\{a,\Phi\}) \mod \Psi^{\alpha}_{\mathrm{phg}} \tag{34}$$

Firstly, (32) follows from the fact that $A = \mathrm{Op}(a) \bmod \Psi^{0}_{\mathrm{phg}}$ (since the subprincipal symbol of A vanishes) and from (28) applied once with $b = a$, $c = \Phi'_t$, and another time with $b = \Phi'_t$ and $c = a$.

Secondly,

$$\mathrm{Op}(\Phi'_t)D_t = \mathrm{Op}(\Phi'_t)\mathrm{Op}(\tau) = \mathrm{Op}(\Phi'_t\tau + \frac{1}{2i}\{\Phi'_t,\tau\}) + \Psi^{\alpha-1}_{\mathrm{phg}}$$

thanks to (28). Hence, using again (28), we get

$$\begin{aligned} D_t\mathrm{Op}(\Phi'_t)D_t &= \mathrm{Op}(\tau)\mathrm{Op}(\Phi'_t\tau + \frac{1}{2i}\{\Phi'_t,\tau\}) \mod \Psi^{\alpha}_{\mathrm{phg}} \\ &= \mathrm{Op}(\tau^2\Phi'_t + \frac{\tau}{2i}\{\Phi'_t,\tau\} + \frac{1}{2i}\{\tau,\Phi'_t\tau\}) \mod \Psi^{\alpha}_{\mathrm{phg}} \end{aligned}$$

which proves (33).

Thirdly, thanks to $A = \mathrm{Op}(a) \bmod \Psi^{0}_{\mathrm{phg}}$ and (29), we have

$$[A,\mathrm{Op}(\Phi)] = \mathrm{Op}\left(\frac{1}{i}\{a,\Phi\}\right) \mod \Psi^{\alpha-1}_{\mathrm{phg}}$$

(note that the remainder is in $\Psi^{\alpha-1}_{\mathrm{phg}}$, not in $\Psi^{\alpha}_{\mathrm{phg}}$). Using (28), we get

$$[A,\mathrm{Op}(\Phi)]D_t + D_t[A,\mathrm{Op}(\Phi)] = \mathrm{Op}\left(\frac{2\tau}{i}\{a,\Phi\}\right) \mod \Psi^{\alpha}_{\mathrm{phg}}$$

which proves (34).

In particular, we get the principal symbol

$$\sigma_p(C) = \tau^2\Phi'_t - \tau H_a\Phi + \Phi'_t a.$$

Using $p = \tau^2 - a$, we can write it differently:

$$\begin{aligned} \sigma_p(C) &= \tau^2\Phi'_t - \tau\{\tau^2 - p,\Phi\} + \Phi'_t a \\ &= \tau^2\Phi'_t - \tau\{\tau^2,\Phi\} + \tau H_p\Phi + \Phi'_t a \\ &= \tau^2\Phi'_t - 2\tau^2\Phi'_t + \tau H_p\Phi + \Phi'_t a \\ &= \tau H_p\Phi - \Phi'_t p. \end{aligned}$$

Moreover, the formulas (32), (33) and (34) imply that the subprincipal symbol of C vanishes, which concludes the proof. □

4 Proof of Theorem 1

The goal of this section is to prove Theorem 1. For $V \subset T^*X$ and $I \subset \mathbb{R}$, we set

$$\begin{aligned} \mathscr{S}(I;V) = \{&(s,y,\eta) \in I \times T^*X, \text{ there exist } (x,\xi) \in V,\ \tau \in \mathbb{R} \text{ and a ray} \\ &\text{from } (s,\tau,y,\eta) \text{ to } (0,\tau,x,\xi)\}. \end{aligned} \tag{35}$$

Most of the time, we will consider $I \subset (-\infty, 0)$. Also, when I or V is reduced to a single element, for example $I = \{t\}$ or $V = \{(x, \xi)\}$, we will simplify the notations by dropping out the brackets in the notation: for example, instead of $\mathscr{S}(\{t\}, V)$, we write $\mathscr{S}(t, V)$. Finally, take care that the above notation (35) refers to rays, and not null-rays (see Definition 9).

With the above notations, Theorem 1 can be reformulated as follows: for any $t > 0$ and any $(x_0, \xi_0) \in WF(u(0))$, there exists $(y_0, \eta_0) \in WF(u(-t)) \cup WF(\partial_t u(-t))$ such that $(-t, y_0, \eta_0) \in \mathscr{S}(-t; (x_0, \xi_0))$ and one of the rays from (y_0, η_0) to (x_0, ξ_0) is null.

First reduction of the problem. If $a(x_0, \xi_0) \neq 0$, then Theorem 1 follows from the usual propagation of singularities theorem [11, Theorem 6.1.1] and the fact that $\Gamma_m = \mathbb{R}^{\pm} \cdot H_p(m)$ for $m \notin \Sigma_{(2)}$. Therefore, in the sequel we assume that $a(x_0, \xi_0) = 0$.

Also, note that, to prove Theorem 1, it is sufficient to find $T > 0$ independent of (x, ξ) (and possibly small) such that the result holds for any $t \in (0, T)$.

Idea of the proof of Theorem 1. To show Theorem 1, we will prove for $T > 0$ sufficiently small an inequality of the form

$$\|\mathrm{Op}(\Psi_0)u\|^2_{H^s} \leqslant c(\|\mathrm{Op}(\Psi_0)u\|^2_{L^2} + \|\mathrm{Op}(\Psi_1)u\|^2_{L^2}) + \text{ Remainder terms} \tag{36}$$

where Ψ_0 and Ψ_1 are functions of t, x, ξ such that

- the function Ψ_0 is supported near $t \in [-T, 0]$ and the function Ψ_1 near $t = -T$;
- on their respective supports in t, the operators $\mathrm{Op}(\Psi_0)$ and $\mathrm{Op}(\Psi_1)$ microlocalize respectively near (x_0, ξ_0) and $\mathscr{S}(-T; (x_0, \xi_0))$.

Then, assuming that u is smooth on the support of Ψ_1, we deduce by applying (36) for different functions Ψ_0 with different degrees of homogeneity in ξ that u is smooth on the support of Ψ_0.

The inequality (36), written more precisely as (49) below, will be proved by constructing a time function $\Phi(t, x, \xi)$ such that the time derivative Φ'_t is equal to $\Phi'_t = \Psi_1^2 - \Psi_0^2$, and then by applying the Fefferman-Phong inequality to the operator C given by (22) (for this Φ).

Reduction to $\mathbb{R}^d$. Let us explain how to reduce Theorem 1 to a problem in $\mathbb{R}^d$. We first notice that it is sufficient to prove Theorem 1 locally, i.e., only for the restriction of u to some small open subset $U \subset X$. This follows from the two following facts:

- firstly, by the finite speed of propagation of subelliptic wave equations (proved in [30, Section 3]), for any small $t > 0$, the value of the solution of (4) at time 0 at $x \in X$ only depends on its values in a small neighborhood of x at time $-t$.
- secondly, null-rays stay close from their departure points for short times. This follows from (11), (12), (15).

Theorem 1 is thus a *local* statement for *short* times. As a consequence, a first reduction for its proof consists in fixing a small open subset $U \subset X$ and addressing the propagation of singularities problem in short time for the restriction of u to U.

Then, we consider a coordinate chart $\psi : U \to \Omega \subset \mathbb{R}^d$. The differential operator A is pushed forward by ψ into a differential operator $\widetilde{A}$ on $\mathbb{R}^d$ which is also real, second-order, self-adjoint, non-negative and subelliptic. Moreover, we can lift ψ to a symplectic mapping $\psi_{\mathrm{lift}} : (x, \xi) \mapsto (\psi(x), ((d_x\psi(x))^{-1})^T\xi)$. Through the differential of ψ_{lift}, the cones Γ_m (computed with $a = \sigma_P(A)$, in X) are sent to the same cones, computed this time with $\widetilde{a} = \sigma_P(\widetilde{A})$ in $\mathbb{R}^d$. This follows from the "symplectic" definition of the cones in Section 2.1 and the fact that $\sigma_P(\widetilde{A})$ is the pushforward of $\sigma_P(A)$. Hence, ψ_{lift} maps also null-rays to null-rays. As a consequence, if we prove Theorem 1 in Ω, then pulling back the situation to $U \subset X$ proves Theorem 1 in full generality.

In the sequel, we thus work in $\Omega \subset \mathbb{R}^d$.

4.1 Construction of the time function

As explained in the introduction of this section, we construct a time function $\Phi(t,x,\xi)$ which verifies several properties. A time function is also constructed in the classical proof of Hörmander's propagation of singularities theorem [18, Proposition 3.5.1], but in the present context of subelliptic wave equations, the construction is more involved since the cones Γ_m along which time functions should be non-increasing contain much more than a single direction (compare (11) with (15)). The following lemma summarizes the properties that the time functions we need thereafter should satisfy. The figures below, notably Figure 2, may help to understand the statement and its proof.

Lemma 14. *Let $(x_0,\xi_0) \in T^*\Omega \setminus 0$ and $V \subset V'$ be sufficiently small open conic (in ξ) neighborhoods of (x_0,ξ_0) such that $\overline{V} \subset V'$. There exists $T>0$ such that for any $0 \leqslant \delta_0 \leqslant T/10$ and any $\alpha \geqslant 0$, there exists a smooth function $\Phi(t,x,\xi)$ with the following properties:*

(1) it is compactly supported in t,x;

(2) it is homogeneous of degree α in ξ;

(3) it is independent of τ;

(4) there exists $\delta>0$ such that at any point of M where $p \geqslant -2\delta a$, there holds $\tau H_p\Phi \leqslant 0$.

(5) its derivative in t can be written $\Phi'_t = \Psi_1^2 - \Psi_0^2$ with Ψ_0 and Ψ_1 homogeneous of degree $\alpha/2$ in ξ;

(6) $\Psi_0 = 0$ outside $\mathscr{S}((-T,\delta_0);V')$ and $\Psi_1 = 0$ outside $\mathscr{S}((-T-\delta_0,-T+\delta_0);V')$;

(7) $\Psi_0 > 0$ on $\mathscr{S}((-T+\delta_0,0);V)$;

(8) Φ is a time function outside $\mathscr{S}((-T-\delta_0,-T+\delta_0);V')$.

All of the above properties of Φ will be used in Sections 4.2 and 4.4 to prove Theorem 1. The rest of Section 4.1 is devoted to the proof of Lemma 14.

We fix $(x_0,\xi_0) \in T^*\Omega \setminus 0$. As said in the introduction of Section 4, we assume that $a(x_0,\xi_0)=0$, and we set $\overline{m} = (0,0,x_0,\xi_0) \in \Sigma_{(2)}$ where the first two coordinates correspond to the variables t,τ. For m near $\overline{m}$, the cone $-\Gamma_m$ is the cone with base point m and containing the opposite of the directions of Γ_m.

We are looking for a τ-independent time function; since any ray lives in a slice $\tau=$ const. (see (11), (12) and (15)), we first construct Φ in the slice $\tau=0$, and then we extend Φ to any τ so that it does not depend on τ. If we start from a time function in $\{\tau=0\}$, then its extension is also a time function: indeed, the image of a ray contained in $\{\tau\neq 0, a=0\}$ under the map $\tau \mapsto 0$ is also a ray, this follows from the fact that $\mathbb{R}^+\partial_t \subset \Gamma_m$ for any $m \in \Sigma_{(2)}$ (see (15)). Thus, the property of being non-increasing along Γ_m is preserved under this extension process. Thus, in the sequel, we work in $\{\tau=0\}$ and do not care about (3).

We now explain why it is natural to impose condition (2) on time functions. Indeed, there is a global homogeneity in ξ of the cones Γ_m and consequently of the null-rays:

Homogeneity Property. If $[T_1,T_2] \ni t \mapsto \gamma(t) = (x(t),\xi(t)) \in \{a=0\}$ is a null-ray parametrized by t, then for any $\lambda>0$, $[T_1,T_2] \ni t \mapsto \gamma_\lambda(t) = (x(t),\lambda\xi(t))$ is a null-ray parametrized by t. Note that $\gamma(t)$ and $\gamma_\lambda(t)$ have the same projection on X for any $t \in [T_1,T_2]$.

This property, illustrated in Figure 1, follows from (10). It will be helpful to find Φ satisfying Point (2) in Lemma 14.

At this point we should say that since we are working in the slice $\{\tau=0\}$, we will use in the sequel the following convenient abuse of notations: for $m=(t,0,x,\xi) \in T^*\mathbb{R} \times T^*\Omega$, we still denote by m the projection of m on $\mathbb{R}\times T^*\Omega$ obtained by throwing away the coordinate $\tau=0$. The fact that the whole picture is now embedded in $\mathbb{R}^{2d+1}$ (see Figure 1) is very convenient: for example, after throwing away the coordinate $\tau=0$, we see the cones Γ_m as subcones of $\mathbb{R}^{2d+1}$ (and not of its tangent space).

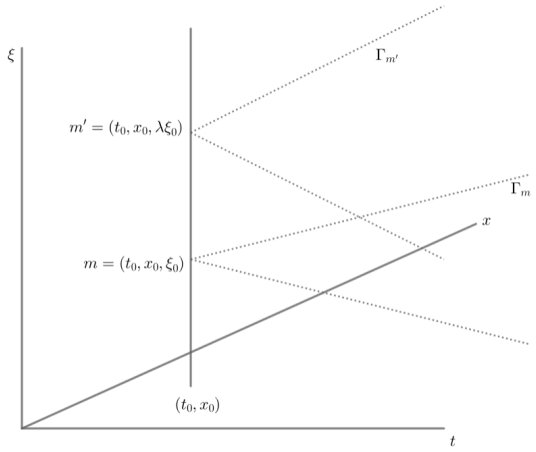

Figure 1: The coordinates and the cones Γ_m. On the picture, the cone $\Gamma_{m'}$ has an aperture which is equal to λ times the aperture of Γ_m (this is "homogeneity").

Also, in the sequel, we only consider points for which $t \geqslant -T$ for some (small) $T > 0$. We take $0 \leqslant \delta_0 \leqslant T/10$.

We will define Φ so that along any ray ending near $(t = 0, x_0, \xi_0)$, its profile looks like Figure 2, which is the standard picture for time functions (see [12, Figure E.2]):

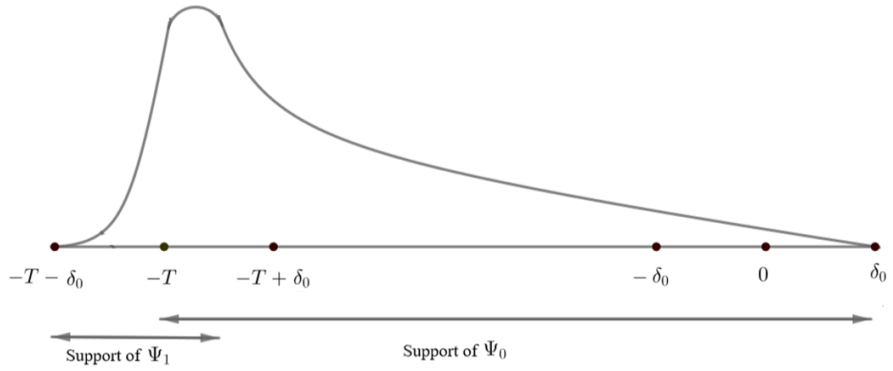

Figure 2: Profile of the function Φ along a ray ending near $(t = 0, x_0, \xi_0)$. The abscissa is this ray, parametrized by t.

The support of Φ is a conic set of points (due to homogeneity in ξ), namely $\mathscr{S}([-T - \delta_0, \delta_0]; V')$. To describe the construction of Φ, we will go backwards in time, from right to left in Figure 2, thus looking for Φ *increasing* along *backward* rays (at least up to time $-T + \delta_0$). Recall that backward rays are just integral curves of the field of cones $-\Gamma_m$. All the rays we consider in the sequel are parametrized by time.

The first important point is that rays enjoy closedness and continuity properties:

Lemma 15. *1. For any closed $V \subset T^*\Omega$ and any $T \geqslant 0$, the set $\mathscr{S}(-T; V)$ is closed.*

2. The mapping $(T, x, \xi) \mapsto \mathscr{S}(-T; (x, \xi))$ is inner semi-continuous, meaning that when $(T_n, x_n, \xi_n) \to (T, x, \xi)$, any point obtained as a limit, as $n \to +\infty$, of points of $\mathscr{S}(-T_n; (x_n, \xi_n))$ belongs to $\mathscr{S}(-T; (x, \xi))$.

Proof. Both properties follow from the locally uniform Lipschitz continuity (20) combined

with the extraction of Lipschitz rays as in the Arzelà-Ascoli theorem and the fact that the cones Γ_m are closed, convex and inner semi-continuous (Lemma 7) $\square$

This lemma implies that in the statement of Lemma 14, the set $\mathscr{S}((-T+\delta_0,0);V)$ is just slightly larger than $\mathscr{S}((-T+\delta_0,0);(x_0,\xi_0))$ (and similarly for other sets $\mathscr{S}(\cdot,\cdot)$ involved in the statement of Lemma 14).

The second important point is the following. We recall that we look for Φ homogeneous of degree $\alpha \geqslant 0$ in ξ; in particular it is increasing in the fibers. To guarantee simultaneously this homogeneity and the fact that Φ increases along backward rays, we have to consider the "worst rays", namely backward rays which are "diving" towards smaller $|\xi|$: we impose that even along the rays (parametrized by time) diving most quickly towards small $|\xi|$, Φ increases. Then, by homogeneity, Φ is also increasing along any other ray.

We apply this procedure for defining Φ to backward rays emanating from a fixed closed conic neighborhood V of (x_0,ξ_0). Point 2 of Lemma 15 implies that

$$\begin{gathered}\text{if } V \text{ is a sufficiently small neighborhood of } (x_0,\xi_0),\\ \Phi \text{ is strictly increasing from time 0 to time } -T+\delta_0\\ \text{along any backward-pointing ray starting from any point } (x,\xi)\in V.\end{gathered} \tag{37}$$

We choose V' a sufficiently small neighborhood of V, we can impose that Φ vanishes at any $(t,x,\xi)\notin \mathscr{S}((-T+\delta_0,\delta_0);V')$ for $t\in(-T+\delta_0,\delta_0)$.

Up to now, our construction defines Φ only for times satisfying $-T+\delta_0 \leqslant t \leqslant \delta_0$. To complete the constrution, we extend it arbitrarily (but smoothly) in $\mathscr{S}([-T-\delta_0,-T+\delta_0],V')$ so that it vanishes for $t\leqslant -T-\delta_0$ (refer again to Figure 2).

Note that with our construction, Φ is decreasing along any backward ray, even along backward rays coming from the region $\{\Phi=0\}$ and entering $\{\Phi\neq 0\}$, and not just along those emanating from $(0,x_0,\xi_0)$ or a nearby point. The only place where the behaviour of Φ is not controlled is $\mathscr{S}([-T-\delta_0,-T+\delta_0],V')$, and this is why Φ is a time function on $\mathscr{S}([-T+\delta_0,\delta_0],V')$ and not on the whole larger set $\mathscr{S}([-T-\delta_0,\delta_0],V')$ (see Property (8)).

For $t\geqslant -T+\delta_0$, we have $\Phi'_t\leqslant 0$ since $\partial_t\in\Gamma_m$, and thus we set $\Psi_0=\sqrt{-\Phi'_t}$. Then, following the rays backwards in time, we make Ψ_0 fall to 0 between times $-T+\delta_0$ and $-T$ (see Figure 2). Similarly, following the rays backward from time $-T+\delta_0$ to time $-T-\delta_0$, we extend Φ smoothly and homogeneously (in the fibers in ξ) in a way that Φ is compactly supported in the time-interval $(-T-\delta_0,\delta_0)$ and $\Phi'_t+\Psi_0^2\geqslant 0$. Finally, we set $\Psi_1=\sqrt{\Phi'_t+\Psi_0^2}$.

In Lemma 14, Properties (1), (2), (3), (5), (6), (8) follow from the construction. Property (7) follows from (37).

Finally, let us explain why Property (4) holds. It follows from our construction that for some $\delta>0$, Φ is decreasing along rays computed with respect to

$$\begin{aligned}\widetilde{\Gamma}_m &= \mathbb{R}^+(\partial_t+\widetilde{B}),\\ \widetilde{B} &= \left\{ b\in \ker(a_m)^{\perp_{\omega_X}},\ a_m^*(\mathcal{I}(b))\leqslant (1-2\delta)^{-1/2}\right\}.\end{aligned}$$

instead of Γ_m (i.e., integral curves of the field of cones $\widetilde{\Gamma}_m$); note that $\widetilde{\Gamma}_m$ is strictly larger than Γ_m due to (15). In fact, for $\delta>0$ sufficiently small (depending on V), these new rays are included in $\mathscr{S}((-T+\delta_0,\delta_0);V)$. Then, in view of the proof of Lemma 7, and notably (19), this gives Property (4).

4.2 A decomposition of C

When Φ satisfies (2), (3), (4) and (5) in Lemma 14, the operator C given by (22) can be expressed as follows:

Proposition 16. *If Φ satisfies (2), (3), (4) and (5) in Lemma 14, then writing $\Phi'_t = \Psi_1^2 - \Psi_0^2$, there holds*

$$C = C' + R + R'P + PR' - \delta(\mathrm{Op}(\Psi_0)A\mathrm{Op}(\Psi_0) + D_t\mathrm{Op}(\Psi_0)^2 D_t) \tag{38}$$

where $\delta > 0$ is the same as in (4),

$$R' = -\frac{\delta}{2}\mathrm{Op}(\Phi'_t) \in \Psi^{\alpha}_{\mathrm{phg}}, \qquad R = \delta\mathrm{Op}(\Psi_1)(D_t^2 + A)\mathrm{Op}(\Psi_1) \in \Psi^{2+\alpha}_{\mathrm{phg}},$$

and $C' \in \Psi^{2+\alpha}_{\mathrm{phg}}$ has non-positive principal symbol and vanishing subprincipal symbol.

We start the proof of this proposition with the following improved (and corrected) version of [30, Lemma 5.3]:

Lemma 17. *Let ϕ be a time function near $\overline{m} \in \Sigma_{(2)}$ which does not depend on τ and such that*

$$\tau H_p\phi \leqslant 0 \qquad \text{on} \quad \{p \geqslant -2\delta a\}. \tag{39}$$

Then there holds

$$\tau H_p\phi \leqslant \phi'_t(p + 2\delta a) \tag{40}$$

in a neighborhood of $\overline{m}$.

Note that for any time function, the inequality (39) holds on the smaller set $\{p \geqslant 0\}$. Assuming (39) is a stronger requirement.

Proof of Lemma 17. Since ϕ does not depend on τ, we know that $q = \tau\{p, \phi\}$ is a quadratic polynomial in τ, vanishing at $\tau = 0$:

$$q = b\tau^2 - c\tau, \quad p = \tau^2 - a, \quad a \geqslant 0.$$

More explicitly, $b = 2\phi'_t$ and $c = \{a, \phi\}$. From (39), we know that $b\tau^2 - c\tau \leqslant 0$ for τ sufficiently large, hence $b \leqslant 0$. Moreover, (39) also implies that if $b = 0$, then $c = 0$, hence $\phi'_t = H_p\phi = 0$, and (40) is automatically satisfied. Otherwise, $b < 0$. Since $q \leqslant 0$ on $\tau \notin [-((1-2\delta)a)^{1/2}, ((1-2\delta)a)^{1/2}]$ by (39), we get that the other zero of q, $\tau = c/b$, must lie in $[-((1-2\delta)a)^{1/2}, ((1-2\delta)a)^{1/2}]$. Thus, $c^2 \leqslant b^2 a(1-2\delta)$. Then,

$$\tau\{p, \phi\} - \phi'_t p = \frac{1}{2}b(\tau - c/b)^2 + (b^2 a - c^2)/2b \leqslant ba\delta = 2\phi'_t a\delta$$

where we used that $b < 0$. □

Proof of Proposition 16. Setting $r' = -\frac{\delta}{2}\Phi'_t$, we have according to Lemma 17 with $\phi = \Phi$:

$$\tau\{p, \Phi\} - \Phi'_t p - 2r'p \leqslant 2\Phi'_t a\delta + \Phi'_t p\delta = \Phi'_t\delta(\tau^2 + a) = \delta(\Psi_1^2 - \Psi_0^2)(\tau^2 + a). \tag{41}$$

We set $R = \delta\mathrm{Op}(\Psi_1)(D_t^2 + A)\mathrm{Op}(\Psi_1)$ and $R' = \mathrm{Op}(r')$. It follows from (41), (30), (31) and (28) that the operator

$$C' = C - R - (R'P + PR') + \delta(\mathrm{Op}(\Psi_0)A\mathrm{Op}(\Psi_0) + D_t\mathrm{Op}(\Psi_0)^2 D_t) \tag{42}$$

has non-positive principal symbol and vanishing subprincipal symbol. This proves Proposition 16. □

4.3 The Fefferman-Phong inequality

The Fefferman-Phong inequality [13] (see also [27, Section 2.5.3]) can be stated as follows: for any pseudodifferential operator C_1' of order $2+\alpha$ whose (Weyl) symbol is non-positive, there holds for any $u \in C_c^\infty(\mathbb{R}^n)$,

$$(C_1'u,u)_{L^2(\mathbb{R}^n)} \leqslant c((\mathrm{Id}-\Delta)^{\alpha/2}u,u)_{L^2(\mathbb{R}^n)} \tag{43}$$

where Δ is a Riemannian Laplacian on $\mathbb{R}^n$. The following lemma is a simple microlocalization of this inequality. The definition of the essential support, denoted by essupp, is recalled in Appendix A.2.

Lemma 18. *Let $\alpha \geqslant 0$, and let $W, W' \subset T^*(\mathbb{R}\times\Omega)$ be conic sets such that W' is a conic neighborhood of W. Let $C' \in \Psi_{\mathrm{phg}}^{2+\alpha}$ with $\mathrm{essupp}(C') \subset W$ such that $\sigma_p(C') \leqslant 0$ and $\sigma_{sub}(C') \leqslant 0$. Then there exists $C_\alpha \in \Psi_{\mathrm{phg}}^{\alpha/2}$ with $\mathrm{essupp}(C_\alpha) \subset W'$ such that*

$$\forall u \in C_c^\infty(\mathbb{R}\times\Omega), \qquad (C'u,u)_{L^2} \leqslant c(\|C_\alpha u\|_{L^2}^2 + \|u\|_{L^2}^2). \tag{44}$$

Proof. Taking a microlocal cut-off χ homogeneous of order 0, essentially supported in W' and equal to 1 on a conic neighborhood of W, we see that

$$\begin{aligned}(C'u,u) &= (C'(\mathrm{Op}(\chi)+\mathrm{Op}(1-\chi))u, (\mathrm{Op}(\chi)+\mathrm{Op}(1-\chi))u)\\ &= (\mathrm{Op}(\chi)C'\mathrm{Op}(\chi)u,u) + (Q'u,u)\end{aligned} \tag{45}$$

where $Q' \in \Psi^{-\infty}$ is explicit:

$$Q' = \mathrm{Op}(1-\chi)C'\mathrm{Op}(\chi) + \mathrm{Op}(\chi)C'\mathrm{Op}(1-\chi) + \mathrm{Op}(1-\chi)C'\mathrm{Op}(1-\chi).$$

Since $Q' \in \Psi^{-\infty}$, we have in particular

$$(Q'u,u) \leqslant c\|u\|_{L^2}^2. \tag{46}$$

Then, we write $C' = C_1' + C_2'$ where C_1' has non-positive full Weyl symbol, and $C_2' \in \Psi_{\mathrm{phg}}^{\alpha}$. First, we apply (43) with $\mathrm{Op}(\chi)u$ instead of u: we obtain

$$(\mathrm{Op}(\chi)C_1'\mathrm{Op}(\chi)u,u) \leqslant c\|C_\alpha u\|_{L^2}^2 \tag{47}$$

with $C_\alpha = (\mathrm{Id}-\Delta)^{\alpha/4}\mathrm{Op}(\chi)$. Secondly, writing $C_2' = (\mathrm{Id}-\Delta)^{\alpha/4}C_2''(\mathrm{Id}-\Delta)^{\alpha/4}$ with $C_2'' \in \Psi_{\mathrm{phg}}^0$, we see that

$$(\mathrm{Op}(\chi)C_2'\mathrm{Op}(\chi)u,u) \leqslant c\|C_\alpha u\|_{L^2}^2. \tag{48}$$

Combining (45), (46), (47) and (48), we get (44). □

4.4 End of the proof of Theorem 1

We come back to the proof of Theorem 1. We fix $(x_0,\xi_0) \in T^*\Omega \setminus 0$ and consider u a solution of (4). For the moment, we assume that u is *smooth*. We consider a time function Φ as constructed in Lemma 14.

Using (38), we have

$$\begin{aligned}0 &= 2\mathrm{Im}(Pu, \mathrm{Op}(\Phi)D_t u)\\ &= (Cu,u)\\ &= ((C' + R + R'P + PR' - \delta(\mathrm{Op}(\Psi_0)A\mathrm{Op}(\Psi_0) + D_t\mathrm{Op}(\Psi_0)^2D_t))u,u).\end{aligned}$$

Hence, using $Pu=0$ and applying Lemma 18 to C', we get:

$$\begin{aligned}(A\mathrm{Op}(\Psi_0)u, \mathrm{Op}(\Psi_0)u) + \|\mathrm{Op}(\Psi_0)D_t u\|_{L^2}^2 &\leqslant c((R_\alpha + R'P + PR' + C')u,u)\\ &\leqslant c_\alpha(\|C_\alpha u\|_{L^2}^2 + \|u\|_{L^2}^2 + (R_\alpha u,u)).\end{aligned}$$

with $c_\alpha \geqslant 1/\delta$ and $R_\alpha = R$, just to keep in mind in the forthcoming inequalities that it depends on α.

But $(A\mathrm{Op}(\Psi_0)u, \mathrm{Op}(\Psi_0)u) \geqslant \frac{1}{c}((-\Delta)^s \mathrm{Op}(\Psi_0)u, \mathrm{Op}(\Psi_0)u) - \|\mathrm{Op}(\Psi_0)u\|^2$ by subellipticity (3). Hence

$$\|(-\Delta)^{s/2}\mathrm{Op}(\Psi_0)u\|_{L^2}^2 + \|\mathrm{Op}(\Psi_0)D_t u\|_{L^2}^2 \leqslant c_\alpha(\|C_\alpha u\|_{L^2}^2 + \|u\|_{L^2}^2 + (R_\alpha u, u) + \|\mathrm{Op}(\Psi_0)u\|_{L^2}^2) \tag{49}$$

which we decompose into

$$\|(-\Delta)^{s/2}\mathrm{Op}(\Psi_0)u\|_{L^2}^2 \leqslant c_\alpha(\|C_\alpha u\|_{L^2}^2 + \|u\|_{L^2}^2 + (R_\alpha u, u) + \|\mathrm{Op}(\Psi_0)u\|_{L^2}^2) \tag{50}$$

and

$$\|\mathrm{Op}(\Psi_0)D_t u\|_{L^2}^2 \leqslant c_\alpha(\|C_\alpha u\|_{L^2}^2 + \|u\|_{L^2}^2 + (R_\alpha u, u) + \|\mathrm{Op}(\Psi_0)u\|_{L^2}^2). \tag{51}$$

Now, assume that u is a general solution of (4), not necessarily smooth. We have $u \in C^0(\mathbb{R}; \mathscr{D}(A^{1/2})) \cap C^1(\mathbb{R}; L^2(\Omega))$. We recall that Ω is a subset of $\mathbb{R}^d$. We have the following definition.

Definition 19. *Let $s_0 \in \mathbb{R}$ and $f \in \mathcal{D}'(\Omega)$. We shall say that f is H^{s_0} at $(x,\xi) \in T^*\Omega \setminus 0$ if there exists a conic neighborhood W of (x,ξ) such that for any 0-th order pseudodifferential operator B with* essupp$(B) \subset W$*, we have $Bf \in H^s_{loc}(\Omega)$.*
We shall say that f is smooth at (x,ξ) if it is H^{s_0} at (x,ξ) for any $s_0 \in \mathbb{R}$.
When we say that u is H^{s_0} at (t,y,η), we mean that $u(t)$ is H^{s_0} at $(y,\eta) \in T^*\Omega$.

Lemma 20. *Let V, V' be sufficiently small open conic neighborhoods of (x_0,ξ_0) such that $\overline{V} \subset V'$. Let u be a solution of (4). If u and $\partial_t u$ are smooth in $\mathscr{S}((-T-\delta_0, -T+\delta_0); V')$, then u is smooth in*

$$U = \mathscr{S}((-T+\delta_0, 0); V).$$

Proof of Lemma 20. We set $u_\varepsilon = \rho_\varepsilon * u$ where $\rho_\varepsilon = \varepsilon^{-(d+1)}\rho(\cdot/\varepsilon)$ and $\rho \in C_c^\infty(\mathbb{R}^{d+1})$ is of integral 1 (and depends on the variables t, x). Recall that d is the dimension of Ω.

Applying Lemma 14 for any $\alpha \geqslant 0$ yields a function Φ_α which is in particular homogeneous of degree α in ξ; its derivative in t can be written $\Phi'_\alpha = (\Psi_1^\alpha)^2 - (\Psi_0^\alpha)^2$ (the upper index being not an exponent). Then we apply (50) to u_ε and with $\alpha = 0$: we get

$$\|(-\Delta)^{s/2}\mathrm{Op}(\Psi_0^0)u_\varepsilon\|_{L^2}^2 \leqslant c_0(\|C_0 u_\varepsilon\|_{L^2}^2 + \|u_\varepsilon\|_{L^2}^2 + (R_0 u_\varepsilon, u_\varepsilon) + \|\mathrm{Op}(\Psi_0^0)u_\varepsilon\|_{L^2}^2) \tag{52}$$

where $R_0 = \delta\mathrm{Op}(\Psi_1^0)(D_t^2 + A)\mathrm{Op}(\Psi_1^0)$ (see Proposition 16) and $c > 0$ does not depend on ε. All quantities

$$\|C_0 u\|_{L^2}, \quad \|u\|_{L^2}, \quad (R_0 u, u), \quad \|\mathrm{Op}(\Psi_0^0)u\|_{L^2}^2$$

are finite. Therefore, taking the limit $\varepsilon \to 0$ in (52), we obtain $u \in H^{2s}$ in U. Using the family of inequalities (50), we can iterate this argument: first with $\alpha = 2s$, then with $\alpha = 4s, 6s$, etc, and each time we replace Ψ_0^0, R_0, C_0 by Ψ_0^α, R_α, C_α. At step k, we deduce thanks to (50) that $u \in H^{2ks}$. In particular, we use the fact that $\|C_\alpha u\|_{L^2}$ and $\|\mathrm{Op}(\Psi_0^\alpha)u\|_{L^2}$ are finite, which comes from the previous step of iteration since C_α is essentially supported close to the essential support of C' (which is contained in the essential support of Φ thanks to (42)). Thus, $u \in \bigcap_{k\in\mathbb{N}} H^{2ks} = C^\infty$ in U.

Then, using (51) for any $\alpha \in \mathbb{N}$ with Ψ_0^α in place of Ψ_0, we obtain that $D_t u$ is also H^α in U. Since this is true for any $\alpha \in \mathbb{N}$, $D_t u$ is C^∞ in U, which concludes the proof of Lemma 20. □

We conclude the proof of Theorem 1. We assume that

$$u \text{ is smooth in } W = \mathscr{S}((-T-\delta_0, -T+\delta_0); (x_0,\xi_0)). \tag{53}$$

Then, u is smooth in a slightly larger set W', i.e., such that $\overline{W} \subset W'$. By Lemma 15, there exists $V' \subset T^*\Omega \setminus 0$ an open neighborhood of (x_0,ξ_0) such that

$$W \subset \mathscr{S}((-T-\delta_0, -T+\delta_0); V') \subset W'.$$

Fix also an open set $V \subset T^*\Omega \setminus 0$ such that

$$(x_0, \xi_0) \in V \subset \overline{V} \subset V'.$$

Lemma 20 implies that u is smooth in $\mathscr{S}((-T + \delta_0, 0); V)$. In particular,

$$u \text{ is smooth in } \mathscr{S}((-T + \delta_0, 0); (x_0, \xi_0)). \tag{54}$$

The fact that (53) implies (54) proves that singularities of (4) propagate only along *rays*. Using that singularities of P are contained in $\{p = 0\}$, we obtain finally Theorem 1.

5 Proof of Corollary 3

In all the sequel, that is, in Sections 5 and 6, we assume that A is a sub-Laplacian. As mentioned in Definition 2 (see also [16]), it means that we assume that A has the form

$$A = \sum_{i=1}^{K} Y_i^* Y_i \tag{55}$$

where the global smooth vector fields Y_i satisfy Hörmander's condition.

The principal symbol of A, which is also the natural Hamiltonian, is

$$a = \sum_{i=1}^{K} h_{Y_i}^2.$$

Here, for Y a vector field on X, we denoted by h_Y the momentum map given in canonical coordinates (x, ξ) by $h_Y(x, \xi) = \xi(Y(x))$.

In this Section, we prove Corollary 3. For that purpose, we introduce in Section 5.1 notations and concepts from sub-Riemannian geometry, which is a natural framework to study geometric properties of the vector fields $Y_1, \dots, Y_K$. Our presentation is inspired by [34, Chapter 5 and Appendix D] (see also [1]).

5.1 Sub-Riemannian geometry and horizontal curves

We consider the sub-Riemannian distribution

$$\mathcal{D} = \mathrm{Span}(Y_1, \dots, Y_K) \subset TX.$$

There is a metric g, namely

$$g_x(v) = \inf \left\{ \sum_{i=1}^{K} u_i^2 \mid v = \sum_{i=1}^{K} u_i Y_i(x) \right\}, \tag{56}$$

which is Riemannian on $\mathcal{D}$ and equal to $+\infty$ outside $\mathcal{D}$. The triple $(X, \mathcal{D}, g)$ is called a sub-Riemannian structure (see [34]).

Fix an interval $I = [b, c]$ and a point $x_0 \in X$. We denote by $\Omega(I, x_0; \mathcal{D})$ the space of all absolutely continuous curves $\gamma : I \to X$ that start at $\gamma(b) = x_0$ and whose derivative is square integrable with respect to g, implying that the length

$$\int_I \sqrt{g_{\gamma(t)}(\dot{\gamma}(t))} dt$$

of γ is finite. Such a curve γ is called *horizontal*. The *endpoint map* is the map

$$\mathrm{End} : \Omega(I, x_0; \mathcal{D}) \to X, \quad \gamma \mapsto \gamma(c).$$

The metric (56) induces a distance d on X, and $d(x,y) < +\infty$ for any $x, y \in X$ thanks to Hörmander's condition (this is the Chow-Rashevskii theorem).

Two types of curves in $\Omega(I, x_0; \mathcal{D})$ will be of particular interest: the critical points of the endpoint map, and the curves which are projections of the Hamiltonian vector field H_a associated to a.

Projections of integral curves of H_a are geodesics:

Theorem 21. *[34, Theorem 1.14] Let $\gamma(s)$ be the projection on X of an integral curve (in T^*X) of the Hamiltonian vector field H_a. Then γ is a horizontal curve and every sufficiently short arc of γ is a minimizing sub-Riemannian geodesic (i.e., a minimizing path between its endpoints in the metric space (X, d)).*

Such horizontal curves γ are called *normal* geodesics, and they are smooth.
The differentiable structure on $\Omega(I, x_0; \mathcal{D})$ described in [34, Chapter 5 and Appendix D] allows to give a sense to the following notion:

Definition 22. *A singular curve is a critical point for the endpoint map.*

Note that in Riemannian geometry (i.e., for a elliptic), there exist no singular curves.
In the next definition, we use the notation $\mathcal{D}^\perp$ for the annihilator of $\mathcal{D}$ (thus a subset of the cotangent bundle T^*X), and $\overline{\omega}_X$ denotes the restriction to $\mathcal{D}^\perp$ of the canonical symplectic form ω_X on T^*X.

Definition 23. *A characteristic for $\mathcal{D}^\perp$ is an absolutely continuous curve $\lambda(t) \in \mathcal{D}^\perp$ that never intersects the zero section of $\mathcal{D}^\perp$ and that satisfies $\dot{\lambda}(t) \in \ker(\overline{\omega}_X(\lambda(t)))$ at every point t for which the derivative $\dot{\lambda}(t)$ exists.*

Theorem 24. *[34, Theorem 5.3] A curve $\gamma \in \Omega$ is singular if and only if it is the projection of a characteristic λ for $\mathcal{D}^\perp$ with square-integrable derivative. λ is then called an abnormal extremal lift of the singular curve γ.*

Normal geodesics and singular curves are particularly important in sub-Riemannian geometry because of the following fact (Pontryagin's maximum principle):

any minimizing geodesic in (X, d) is either a singular curve or a normal geodesic.

The existence of minimizing geodesics which are singular curves but not normal geodesics was proved in [32].

Let us mention three examples which are well-known in sub-Riemannian geometry (see [34] and [1]) where the singular curves and their abnormal extremal lifts can be explicitly computed. These examples are presented in $\mathbb{R}^n$ $(n = 3, 4)$ for simplicity, but they could also have been written on adequate compact manifolds in order to fit better with the framework of the present work.

Example 25. *When the vector fields are $Y_1 = \partial_x - \frac{y}{2}\partial_z$ and $Y_2 = \partial_y + \frac{x}{2}\partial_z$ in $\mathbb{R}^3$, and the measure is $\nu = dxdydz$, then*

$$A = -\left[\left(\partial_x - \frac{y}{2}\partial_z\right)^2 + \left(\partial_y + \frac{x}{2}\partial_z\right)^2\right]$$

is the Heisenberg sub-Laplacian. In this case, the only singular curves are the trivial constant ones $t \mapsto q_0$ for any $q_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$, and the abnormal extremal lifts are given by $(q_0, p_0) \in T^*\mathbb{R}^3$ where $p_0 \neq 0$ annihilates $Y_1(q_0)$ and $Y_2(q_0)$ and is thus proportional to $dz + \frac{y_0}{2}dx - \frac{x_0}{2}dy$.

Example 26. *When the vector fields are $Y_1 = \partial_x$ and $Y_2 = \partial_y + x^2\partial_z$ in $\mathbb{R}^3$, and the measure is $\nu = dxdydz$, then*

$$A = -\left[\partial_x^2 + \left(\partial_y + x^2\partial_z\right)^2\right]$$

is the Martinet sub-Laplacian. Apart from the trivial constant ones, the singular curves are of the form $t \mapsto q(t) = (0, t, z_0)$ for any $z_0 \in \mathbb{R}$ and the abnormal extremal lifts are given by $(q(t), p(t)) \in T^*\mathbb{R}^3$ where $p(t) \neq 0$ annihilates $Y_1(q(t)) = \partial_x$ and $Y_2(q(t)) = \partial_y$. Note that many other parametrizations of these curves are possible.

Example 27. *When the vector fields are* $Y_1 = \partial_x - \frac{y}{2}\partial_z$, $Y_2 = \partial_y + \frac{x}{2}\partial_z$ *and* $Y_3 = \partial_w$ *in* $\mathbb{R}^4$, *and the measure is* $\nu = dxdydzdw$, *then*

$$A = -\left[\left(\partial_x - \frac{y}{2}\partial_z\right)^2 + \left(\partial_y + \frac{x}{2}\partial_z\right)^2 + \partial_w^2\right]$$

is the quasi-contact (or even contact) sub-Laplacian. Apart from the trivial constant ones, the singular curves are of the form $t \mapsto q(t) = (x_0, y_0, z_0, t)$ *for any* $x_0, y_0, z_0 \in \mathbb{R}$, *and their abnormal extremal lifts are given by* $(q(t), p(t)) \in T^*\mathbb{R}^4$ *where* $p(t) \neq 0$ *annihilates* $Y_1(q(t)), Y_2(q(t))$ *and* $Y_3(q(t))$, *and is thus proportional to* $dz + \frac{y_0}{2}dx - \frac{x_0}{2}dy$. *Again, many other parametrizations of these curves are possible.*

5.2 End of the proof of Corollary 3

The basic facts of sub-Riemannian geometry recalled in Section 5.1 now allow us to deduce Corollary 3 from Theorem 1.

We recall a few notations already used: π denotes the canonical projection $\pi : T^*X \to X$, and $\mathcal{I}$ is the canonical isomorphism between $T(T^*X)$ and $T^*(T^*X)$ introduced in (16). The notation a_m stands for half the Hessian of the principal symbol of A at m, and a_m^* is then defined as in (14).

The expression (15) of the cones Γ_m can be simplified thanks to the following lemma.

Lemma 28. *When* A *is a sub-Laplacian, there holds* $a_m^*(\mathcal{I}(b)) = g(d\pi(b))$ *for any* $b \in (\ker(a_m))^{\perp_{\omega_X}} \subset T(T^*X)$.

Proof. We consider a frame $Z_1, \dots, Z_N$ which is g-orthonormal at the (canonical) projection of m on X. In particular, the Z_j are independent, and the $H_{h_{Z_j}}$ are also independent. We have $a_m = \sum_{j=1}^N (dh_{Z_j})^2$. Hence, $H_{h_{Z_1}}, \dots, H_{h_{Z_N}}$ span $(\ker(a_m))^{\perp_{\omega_X}}$ since

$$\begin{aligned}\ker(a_m) &= \bigcap_{j=1}^N \ker(dh_{Z_j}) = \{\xi \in T(T^*X),\ dh_{Z_j}(\xi) = 0,\ \forall 1 \leqslant j \leqslant N\}\\ &= \{\xi \in T(T^*X),\ \omega(\xi, H_{h_{Y_N}}) = 0,\ \forall 1 \leqslant j \leqslant N\}\\ &= \mathrm{span}(H_{h_{Y_1}}, \dots, H_{h_{Y_N}})^{\perp_{\omega_X}}.\end{aligned}$$

We fix $b \in (\ker(a_m))^{\perp_{\omega_X}}$ and we write $b = \sum_{j=1}^N u_j H_{h_{Z_j}}$. By definition, $\mathcal{I}(H_{h_{Z_j}}) = -dh_{Z_j}$ and $d\pi(H_{h_{Z_j}}) = Z_j$ for any j, so there holds

$$\begin{aligned}a_m^*\left(\mathcal{I}\left(\sum_{j=1}^N u_j H_{h_{Z_j}}\right)\right) &= a_m^*\left(\sum_{j=1}^N u_j dh_{Z_j}\right) = \sup_{\eta \notin \ker(a_m)} \frac{\left(\sum_{j=1}^N u_j dh_{Z_j}(\eta)\right)^2}{\sum_{j=1}^N dh_{Z_j}(\eta)^2}\\ &= \sup_{(\theta_j)\in\mathbb{R}^N} \frac{\left(\sum_{j=1}^N u_j\theta_j\right)^2}{\sum_{j=1}^N \theta_j^2} = \sum_{j=1}^N u_j^2 = g\left(\sum_{j=1}^N u_j Z_j\right)\\ &= g\left(d\pi\left(\sum_{j=1}^N u_j H_{h_{Z_j}}\right)\right)\end{aligned}$$

where, to go from line 1 to line 2, we used that the dh_{Z_j} are independent. □

From Lemma 28, we deduce the following expression for Γ_m:

$$\begin{aligned}\Gamma_m &= \mathbb{R}^+(\partial_t + B),\\ B &= \left\{b \in \ker(a_m)^{\perp_{\omega_X}},\ g(d\pi(b)) \leqslant 1\right\}.\end{aligned} \tag{57}$$

Recall that $M = T^*(\mathbb{R} \times X) \setminus 0$, and let $\pi_2 : M \to T^*X$ be the canonical projection on the second factor.

Proposition 29. *Let $\gamma : I \to M$ be a null-ray, as introduced in Definition 9. Then, τ is constant along this null-ray, and necessarily:*

(i) if $\tau \equiv c \neq 0$, then γ is a null-bicharacteristic;

(ii) if $\tau \equiv 0$, then γ is contained in $\Sigma_{(2)}$ and tangent to the cones Γ_m given by (57). Moreover, $\pi_2(\gamma) \subset T^*X$ is a characteristic curve, and its projection on X (a singular curve) is traveled at speed $\leqslant 1$.

Proof. First, we note that τ is constant along null-rays since τ is preserved along integral curves of H_p, and $d\tau(v) = 0$ for any $v \in \Gamma_m$ when Γ_m is given by (57).

If $\tau \equiv c \neq 0$, then $\Gamma_m = \mathbb{R}^{\pm} \cdot H_p(m)$ for $m \in \gamma(I)$. Thus γ is a null-bicharacteristic, which proves (i).

We finally prove (ii); we assume that $\tau \equiv 0$ along γ. Let $s \in I$. We set $n(s) = \pi_2(\gamma(s))$. According to (57), we can write $\dot\gamma(s) = c(s)(\partial_t + b(s))$ with $b(s) \in T_{n(s)}\mathcal{D}^\perp$ since $a \equiv 0$ along the path. There holds $\ker(a_{n(s)}) = T_{n(s)}\mathcal{D}^\perp$ where $a_{n(s)}$ is half the Hessian of a at point $n(s)$. Plugging into the above formula, we also get $b(s) \in (T_{n(s)}\mathcal{D}^\perp)^{\perp_{\omega_X}}$. It follows that

$$b(s) \in T_{n(s)}\mathcal{D}^\perp \cap (T_{n(s)}\mathcal{D}^\perp)^{\perp_{\omega_X}} = \ker(\overline{\omega}_X(n(s))).$$

This implies that $\pi_2(\gamma)$ is a characteristic curve. Its projection on X is a singular curve, by definition. Moreover, the inequality $g(d\pi(b)) \leqslant 1$ in (57) exactly means that this projection is traveled at speed $\leqslant 1$. □

Proposition 29 directly implies Corollary 3. To sum up, singularities of the wave equation (4) when A is a sub-Laplacian propagate only along integral curves of H_a and characteristics for $\mathcal{D}^\perp$ (at speed $\leqslant 1$).

Comments. The propagation of singularities at speeds < 1 along singular curves is not excluded by Corollary 3. If such a slow propagation effectively exists (which is not proved by Corollary 3), it is in strong contrast with the usual propagation "at speed 1" along the integral curves of H_a (as in Hörmander's theorem). As already mentioned in the introduction, we proved this surprising fact in a joint work with Yves Colin de Verdière [9]: we gave explicit examples of initial data of a subelliptic wave equation whose singularities effectively propagate at any speed between 0 and 1 along a singular curve.

Note that a similar phenomenon occurs for "partial sub-Laplacians" (typically, ∂_x^2 in $\mathbb{R}^2_{x,y}$), as proved in [31] using a parametrix construction. The result of [31] is presented as a model for the "conical refraction", which is the splitting of a ray into a cone of rays by a biaxial crystal.

6 Proof of Theorem 4 and Corollary 5

We now turn to the study of the wave kernel K_G. Section 6 is devoted to the proof of Theorem 4 and Corollary 5, i.e., we deduce the wave-front set of the Schwartz kernel K_G from the "geometric" propagation of singularities given by Corollary 3. The idea is to consider K_G itself as the solution of a subelliptic wave equation to which we can apply Corollary 3.

6.1 K_G as the solution of a wave equation

We consider the product manifold $X \times X$, with coordinate x on its first copy, and coordinate y on its second copy. We set

$$A^{\otimes} = \frac{1}{2}(A_x \otimes \mathrm{Id}_y + \mathrm{Id}_x \otimes A_y)$$

and we consider the operator

$$P^{\otimes} = \partial_{tt}^2 - A^{\otimes}$$

acting on functions of $\mathbb{R} \times X_x \times X_y$. Using (7), we can check that the Schwartz kernel K_G is a solution of

$$K_{G|t=0} = 0, \qquad \partial_t K_{G|t=0} = \delta_{x-y}, \qquad P^{\otimes} K_G = 0.$$

The operator $A^{\otimes}$ is a self-adjoint non-negative real second-order differential operator on $X \times X$. Moreover it is subelliptic: it is immediate that the vector fields $Y_1 \otimes \mathrm{Id}_y, \ldots, Y_K \otimes \mathrm{Id}_y, \mathrm{Id}_x \otimes Y_1, \ldots, \mathrm{Id}_x \otimes Y_K$ verify Hörmander's Lie bracket condition, since it is satisfied by $Y_1, \ldots, Y_K$. Hence, Theorem 1 applies to P, with the cones (and the null-rays) being computed with $A^{\otimes}$ in $T^*(X \times X)$ instead of A (see (60)). We denote by $\sim_t$ the relation of existence of a null-ray of length $|t|$ joining two given points of $T^*(X \times X) \setminus 0$ (see Remark 10 for the omission of the variables t and τ in the null-rays).

Since $WF(K_G(0)) = \emptyset$ and

$$WF(\partial_t K_G(0)) = \{(z, z, \zeta, -\zeta) \in T^*(X \times X) \setminus 0\},$$

(see [36, p. 93]) we have, according to Theorem 1,

$$\begin{aligned} WF(K_G(t)) \subset \{(x, y, \xi, -\eta) \in T^*(X \times X) \setminus 0,\ \exists (z, \zeta) \in T^*X \setminus 0, \\ (z, z, \zeta, -\zeta) \sim_t (x, y, \xi, -\eta)\}. \end{aligned} \tag{58}$$

Our goal is now to give a simpler expression for the right-hand side of (58).

Let us denote by g^1 the sub-Riemannian metric on X_x and by g^2 the sub-Riemannian metric on X_y. The sub-Riemannian metric on $X_x \times X_y$ is $g^{\otimes} = \frac{1}{2}(g^1 \oplus g^2)$. In other words, if $q = (q_1, q_2) \in X \times X$ and $v = (v_1, v_2) \in T_q(X \times X) \approx T_{q_1}X \times T_{q_2}X$, we have

$$g^{\otimes}_q(v) = \frac{1}{2}(g^1_{q_1}(v_1) + g^2_{q_2}(v_2)). \tag{59}$$

Now, according to (57), the cones $\Gamma^{\otimes}_m$ associated to $A^{\otimes}$ are given by

$$\begin{aligned} \Gamma^{\otimes}_m &= \mathbb{R}^+(\partial_t + B), \\ B &= \{b \in \ker(a^{\otimes}_m)^{\perp_{\omega^{\otimes}}},\ g^{\otimes}(d\pi^{\otimes}(b)) \leqslant 1\}. \end{aligned} \tag{60}$$

Here, $\perp_{\omega^{\otimes}}$ designates the symplectic orthogonal with respect to the canonical symplectic form $\omega^{\otimes}$ on $T^*(X \times X)$, and $\pi^{\otimes} : T^*(X \times X) \to X \times X$ is the canonical projection.

To evaluate the right-hand side of (58), we denote by $\approx_t$ the relation of existence of a null-ray of length $|t|$ joining two given points of $T^*X \setminus 0$ (the cones Γ_m are subsets of $T(T^*(\mathbb{R} \times X))$ as defined in Section 2). Let us prove that

$$\begin{aligned} \{(x, y, \xi, -\eta) \in T^*(X \times X) \setminus 0,\ \exists (z, \zeta) \in T^*X \setminus 0,\ (z, z, \zeta, -\zeta) \sim_t (x, y, \xi, -\eta)\} \\ \subset \{(x, y, \xi, -\eta) \in T^*(X \times X) \setminus 0,\ (x, \xi) \approx_t (y, \eta)\}. \end{aligned} \tag{61}$$

Combining with (58), it will immediately follow that

$$WF(K_G(t)) \subset \{(x, y, \xi, -\eta) \in T^*(X \times X) \setminus 0,\ (x, \xi) \approx_t (y, \eta)\}. \tag{62}$$

6.2 Proof of (61).

We denote by $\gamma : [0, t] \to T^*(X \times X) \setminus 0$ a null-ray from $(z, z, \zeta, -\zeta)$ to $(x, y, \xi, -\eta)$, parametrized by time. Our goal is to construct a null-ray of length $|t|$ in $T^*X \setminus 0$, from (y, η) to (x, ξ). It is obtained by concatenating a null-ray from (y, η) to (z, ζ) with another one, from (z, ζ) to (x, ξ). However, there are some subtleties hidden in the parametrization of this concatenated null-ray.

We write $\gamma(s) = (\alpha_1(s), \alpha_2(s), \beta_1(s), \beta_2(s))$, and for $i = 1, 2$ and $0 \leqslant s \leqslant t$, we set $\gamma_i(s) = (\alpha_i(s), \beta_i(s)) \in T^*X$. We also set $\delta_i(s) = g^i(d\pi_i(\dot{\gamma}_i(s)))$, where $\pi_i : T^*X \to X$ (here X is the i-th copy of X). The upper dot denotes here and in the sequel the derivative with

respect to the time variable. Since $g^{\otimes}(d\pi^{\otimes}(\dot\gamma(s))) \leqslant 1$ for any $s \in [0,t]$, we deduce from (59) that

$$\frac{1}{2}(\delta_1(s) + \delta_2(s)) \leqslant 1.$$

Note that it is possible that $\delta^1(s) > 1$ or $\delta^2(s) > 1$.

We are going to construct a null-ray $\varepsilon : [0,t] \to T^*X$ of the form

$$\begin{aligned} \varepsilon(s) &= (\alpha_2(\theta(s)), -\beta_2(\theta(s))), \qquad 0 \leqslant s \leqslant s_0 \\ \varepsilon(s) &= (\alpha_1(\theta(s)), \beta_1(\theta(s))), \qquad s_0 \leqslant s \leqslant t. \end{aligned} \tag{63}$$

The parameter s_0 and the parametrization θ will be chosen so that the first part of ε joins (y,η) to (z,ζ) and the second part joins (z,ζ) to (x,ξ). We choose $\theta(0) = t$, hence $\varepsilon(0) = (y,\eta)$. Then, for $0 \leqslant s \leqslant s_0$, we choose $\theta(s) \leqslant t$ in a way to guarantee that $g^1(d\pi_1(\dot\varepsilon(s))) = 1$. This defines s_0 in a unique way as the minimal time for which $\varepsilon(s_0) = (z,\zeta)$. In particular, $\theta(s_0) = 0$. A priori, we do not know that $s_0 \leqslant t$, but we will prove it below. Then, for $s \geqslant s_0$, we choose $\theta(s) \geqslant 0$ in order to guarantee that $g^2(d\pi_2(\dot\varepsilon(s))) = 1$. This defines a time s_1 in a unique way as the minimal time for which $\varepsilon(s_1) = (x,\xi)$. Finally, if $s_1 \leqslant t$, we extend ε by $\varepsilon(s) \equiv (x,\xi)$ for $s_1 \leqslant s \leqslant t$.

We check that ε is a null-ray in T^*X. We come back to the definition of null-rays as tangent to the cones Γ_m. It is clear that

$$\ker(a_m^{\otimes})^{\perp_{\omega^{\otimes}}} = \ker(a_m)^{\perp_{\omega_1}} \times \ker(a_m)^{\perp_{\omega_2}}$$

where ω_i is the canonical symplectic form on T^*X_i. Therefore, $\dot\varepsilon(s) \in \ker(a_m)^{\perp_{\omega_i}}$ for $i = 1$ when $0 \leqslant s \leqslant s_0$ and for $i = 2$ when $s_0 \leqslant s \leqslant t$. Thanks to Lemma 28, the inequality in (15) (but for the cones in X_1 and X_2) is verified by $\dot\varepsilon(s)$ for any $0 \leqslant s \leqslant t$ by definition. There is a "time-reversion" (or "path reversion") in the first line of (63); the property of being a null-ray is preserved under time reversion together with momentum reversion. Hence ε is a null-ray in T^*X.

The fact that $s_0, s_1 \leqslant t$ follows from the following computation:

$$\begin{aligned} t \geqslant \int_0^t g^{\otimes}(d\pi^{\otimes}(\dot\gamma(s)))ds &= \frac{1}{2}\int_0^t g^1(d\pi_1(\dot\gamma_1(s)))ds + \frac{1}{2}\int_0^t g^2(d\pi_2(\dot\gamma_2(s)))ds \\ &= \frac{1}{2}\int_0^{s_0} g^1(d\pi_1(\dot\varepsilon(s)))ds + \frac{1}{2}\int_{s_0}^{s_1} g^2(d\pi_2(\dot\varepsilon(s)))ds \\ &= s_0 + (s_1 - s_0) = s_1. \end{aligned}$$

where the second equality follows from the fact that ε is a reparametrization of γ_1 (resp. γ_2) for $s \in [0, s_0]$ (resp. $[s_0, s_1]$). This concludes the proof of (61).

6.3 Conclusion of the proof of Theorem 4

Let us finish the proof of Theorem 4. We fix (x_0,ξ_0), (y_0,η_0) and t_0 such that there is no null-ray from $(y_0,\eta_0) \in T^*X$ to $(x_0,\xi_0) \in T^*X$ in time t_0.

Claim. There exist a conic neighborhood V of $(x_0, y_0, \xi_0, -\eta_0)$ in $T^*(X \times X)$ and a neighborhood V_0 of t_0 in $\mathbb{R}$ such that for any $N \in \mathbb{N}$ and any $t \in V_0$, $\partial_t^{2N} K_G(t)$ is smooth in V.

Proof. We choose V so that for $(x,y,\xi,-\eta) \in V$ and $t \in V_0$, there is no null-ray from (y,η) to (x,ξ) in time t. Such a V exists, since otherwise by extraction of null-rays (which are Lipschitz with a locally uniform constant, see (20)), there would exist a null-ray from (y_0,η_0) to (x_0,ξ_0) in time t_0. Then, we can check that for any $N \in \mathbb{N}$, $K_G^{(2N)} = \partial_t^{2N} K_G$ is a solution of

$$K_G^{(2N)}{}_{|t=0} = 0, \qquad \partial_t K_G^{(2N)}{}_{|t=0} = (A^{\otimes})^N \delta_{x-y}, \qquad P^{\otimes} K_G^{(2N)} = 0.$$

Repeating the above argument leading to (62) with $K_G^{(2N)}$ instead of K_G, we obtain

$$WF(K_G^{(2N)}(t)) \subset \{(x,y,\xi,-\eta) \in T^*(X\times X)\setminus 0,\ (x,\xi) \approx_t (y,\eta)\},$$

which proves the claim. □

We deduce from the claim that if there is no null-ray from $(y_0,\eta_0) \in T^*X$ to $(x_0,\xi_0) \in T^*X$ in time t_0, then $(t_0,\tau_0,x_0,y_0,\xi_0,-\eta_0) \notin WF(K_G)$ for any $\tau_0 \in \mathbb{R}$.

Finally, if there is a null-ray from (y_0,η_0) to (x_0,ξ_0) in time t_0, then $a(x_0,\xi_0) = a(y_0,\eta_0)$, and due to the fact that $WF(K_G)$ is included in the characteristic set of $\partial_{tt}^2 - A^{\otimes}$, the only τ_0's for which $(t_0,\tau_0,x_0,y_0,\xi_0,-\eta_0) \in WF(K_G)$ is possible are the ones satisfying $\tau_0^2 = a(x_0,\xi_0) = a(y_0,\eta_0)$. This concludes the proof of Theorem 4.

Remark 30. *Theorem 4 allows to recover some results already known in the literature.*

In the situations studied in [23], [24] and [30], $\Sigma_{(2)}$ is a symplectic manifold (a typical example is given by Example 25). In this case, thanks to (15)*, we see that the only null-rays starting from points in $\Sigma_{(2)}$ are lines in t. Therefore Theorem 4 implies:*

- *the "wave-front part" of the main results of [23] and [24] (but not the effective construction of parametrices handled in these papers);*
- *Theorem 1.8 in [29].*

6.4 Proof of Corollary 5

We finally prove Corollary 5. We fix $x,y \in X$ with $x \neq y$ and we denote by T_s the minimal length of a singular curve joining x to y.

We consider $\varphi : \mathbb{R} \to \mathbb{R}\times X\times X$, $t \mapsto (t,x,y)$ which has conormal set $N_\varphi = \{(t,x,y,0,\xi,\eta)\}$ (in other words N_φ corresponds to $\tau = 0$). Using Theorem 4 and Proposition 29, we see that $WF(\mathscr{G})$ does not intersect the conormal set of $\varphi_{|(-T_s,T_s)}$. Then, [17, Theorem 2.5.11'] ensures that $\mathscr{G}$, which is the pull-back of K_G by $\varphi_{|(-T_s,T_s)}$, is well-defined as a distribution over $(-T_s,T_s)$. Of course, SingSupp($\mathscr{G}$) is the projection of $WF(\mathscr{G})$ (for $|t| < T_s$).

By definition of T_s, for $|t| < T_s$, null-rays between x and y are contained in $\{\tau \neq 0\}$, thus they are null-bicharacteristics (see Proposition 29). Hence, the singularities of $\mathscr{G}$ occur at times belonging to the set $\mathscr{L}$ of lengths of normal geodesics (for $\tau > 0$, we obtain normal geodesics from y to x, and for $\tau < 0$, normal geodesics from x to y).

Remark 31. *If $x = y$, the same reasoning as in the proof of Corollary 5 says nothing more than* SingSupp$(K_G(\cdot,x,x)) \subset \mathbb{R}$ *since for any point $(x,\xi) \in \mathcal{D}^\perp$ and any $t \in \mathbb{R}$, the constant path joining (x,ξ) to (x,ξ) in time t is a null-ray (with $\tau \equiv 0$).*

A Appendix

A.1 Sign conventions in symplectic geometry

In the present work, we take the following conventions (the same as [19], see Chapter 21.1): on a symplectic manifold with canonical coordinates (x,ξ), the symplectic form is $\omega = d\xi \wedge dx$, and the Hamiltonian vector field H_f of a smooth function f is defined by the relation $\omega(H_f,\cdot) = -df(\cdot)$. In coordinates, it reads

$$H_f = \sum_j (\partial_{\xi_j} f)\partial_{x_j} - (\partial_{x_j} f)\partial_{\xi_j}.$$

In these coordinates, the Poisson bracket is

$$\{f,g\} = \omega(H_f,H_g) = \sum_j (\partial_{\xi_j} f)(\partial_{x_j} g) - (\partial_{x_j} f)(\partial_{\xi_j} g),$$

which is also equal to $H_f g$ and $-H_g f$.

A.2 Pseudodifferential operators

This appendix is a short reminder on basic properties of pseudodifferential operators. Most proofs can be found in [19]. In this paper, we work with the class of polyhomogeneous symbols (defined below), which is slightly smaller than the usual class of symbols but has the advantage that the subprincipal symbol can be read easily when using the Weyl quantization (see [19], the paragraph before Section 18.6).

We consider Ω an open set of a d-dimensional manifold, and μ a smooth volume on Ω. The variable in Ω is denoted by q. Let $\pi : T^*\Omega \to \Omega$ be the canonical projection.

$S^n_{\mathrm{hom}}(T^*\Omega)$ stands for the set of homogeneous symbols of degree n with compact support in Ω. We also denote by $S^n_{\mathrm{phg}}(T^*\Omega)$ the set of polyhomogeneous symbols of degree n with compact support in Ω. Hence, $a \in S^n_{\mathrm{phg}}(T^*\Omega)$ if $a \in C^\infty(T^*\Omega)$, the projection $\pi(\mathrm{supp}(a))$ is a compact of Ω, and there exist $a_j \in S^{n-j}_{\mathrm{hom}}(T^*\Omega)$ such that for any $N \in \mathbb{N}$, $a - \sum_{j=0}^{N} a_j \in S^{n-N-1}_{\mathrm{phg}}(T^*\Omega)$. We denote by $\Psi^n_{\mathrm{phg}}(\Omega)$ the space of polyhomogeneous pseudodifferential operators of order n on Ω, with a compactly supported kernel in $\Omega \times \Omega$.

We use the Weyl quantization denoted by $\mathrm{Op} : S^n_{\mathrm{phg}}(T^*\Omega) \to \Psi^n_{\mathrm{phg}}(\Omega)$. It is obtained by using partitions of unity and the formula in local coordinates

$$\mathrm{Op}(a)f(q) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d_{q'} \times \mathbb{R}^d_p} e^{i\langle q-q', p\rangle} a\left(\frac{q+q'}{2}, p\right) f(q')dq'dp.$$

If a is real-valued, then $\mathrm{Op}(a)^* = \mathrm{Op}(a)$. Moreover, with this quantization, the principal and subprincipal symbols of $A = \mathrm{Op}(a)$ with $a \sim \sum_{j \leqslant n} a_j$ are simply $\sigma_p(A) = a_n$ and $\sigma_{\mathrm{sub}}(A) = a_{n-1}$ (usually, the subprincipal symbol is defined for operators acting on half-densities, but we make here the identification $f \leftrightarrow f d\nu^{1/2}$).

We also have the following properties:

1. If $A \in \Psi^l_{\mathrm{phg}}(\Omega)$ and $B \in \Psi^n_{\mathrm{phg}}(\Omega)$, then $[A, B] \in \Psi^{l+n-1}_{\mathrm{phg}}(\Omega)$. Moreover, $\sigma_p([A, B]) = \frac{1}{i}\{\sigma_p(a), \sigma_p(b)\}$ where the Poisson bracket is taken with respect to the canonical symplectic structure of $T^*\Omega$.
2. If X is a vector field on Ω and X^* is its formal adjoint in $L^2(\Omega, \mu)$, then X^*X is a second order pseudodifferential operator, with $\sigma_p(X^*X) = h_X^2$ and $\sigma_{\mathrm{sub}}(X^*X) = 0$. Here, for X a vector field, we denoted by h_X the momentum map given in canonical coordinates (x, ξ) by $h_X(x, \xi) = \xi(X(x))$.
3. If $A \in \Psi^n_{\mathrm{phg}}(\Omega)$, then A maps continuously the space $H^s(\Omega)$ to the space $H^{s-n}(\Omega)$.

Finally, we define the essential support of A, denoted by $\mathrm{essupp}(A)$, as the complement in $T^*\Omega$ of the points (q, p) which have a conic-neighborhood W so that A is of order $-\infty$ in W.

A.3 The cones Γ_m as generalized Hamiltonians

In this section, we interpret the set $B = B(m)$ which appears in the formula (15), namely

$$B(m) = \left\{ b \in \ker(a_m)^{\perp_{\omega_X}},\ a_m^*(\mathcal{I}(b)) \leqslant 1 \right\},$$

as a generalized Hamiltonian, just adapting the notion of Clarke generalized gradient (see [7, Chapter 1.2]) to the "Hamiltonian" framework.

Definition 32. *Let f be an almost everywhere differentiable function on T^*X and let Ω_f be the set of points where it is not differentiable. Its generalized Clarke Hamiltonian $\mathcal{H}f(x)$ at $x \in \Omega_f$ is the set*

$$\mathcal{H}f(x) = \mathrm{cxhl}\left\{ \lim_{j \to +\infty} H_f(x_j),\ x_j \to x,\ x_j \notin \Omega_f \right\} \subset T_x(T^*X)$$

where cxhl *denotes the convex hull.*

The main result of this section is the following:

Proposition 33. *For any $m \in \Sigma_{(2)}$, $B(m) = \mathcal{H}\sqrt{a}(m)$.*

This proposition, beside giving an alternative proof of Lemma 7, draws a link between our computations and the Pontryagin maximum principle in the Clarke formulation, which asserts that any sub-Riemannian geodesic (see Section 5.1) is a solution of the differential inclusion

$$\dot\gamma(s) \in \mathcal{H}\sqrt{a}(\gamma(s)).$$

The projection of a null-ray in T^*X is also by Definition 9 a solution of this differential inclusion, and this "explains" why abnormal extremals appear naturally in Corollary 3.

Before proving Proposition 33, we introduce the "fundamental matrix" F (see [19, Section 21.5]) defined as follows:

$$\forall Y, Z \in T_m(T^*X), \qquad \omega_X(Y, FZ) = a_m(Y, Z). \tag{64}$$

Here $a_m(Y,Z) = \frac{1}{2}(\mathrm{Hess}\, a)(m)(Y,Z)$. Then, $\omega_X(FY, Z) = -\omega_X(Y, FZ)$. As already explained in Section 2.2, there is here a slight abuse of notations since $T_m(T^*X)$ stands for $T_{\pi_2(m)}(T^*X)$ where $\pi_2 : M \to T^*X$ is the canonical projection on the second factor.

Lemma 34. *The fundamental matrix induces an isomorphism*

$$F : T_m(T^*X)/\ker(a_m) \to \ker(a_m)^{\perp_{\omega_X}}$$

Proof. F clearly passes to the quotient by $\ker(a_m)$ by (64). Let $b \in \ker(a_m)^{\perp_{\omega_X}}$. We set $b_0 = -\mathcal{I}(b) \in \ker(a_m)^{\perp}$. The bilinear form a_m is continuous and coercive on $T_m(T^*X)/\ker(a_m)$, and b_0 is a linear form on this space, thus by Lax-Milgram's lemma we get the existence of $Z \in T_m(T^*X)/\ker(a_m)$ such that $b_0 = a_m(Z, \cdot)$. Finally, we have

$$-\mathcal{I}(b) = b_0 = a_m(\cdot, Z) = \omega(\cdot, FZ) = -\mathcal{I}(FZ)$$

according to (16), which means that $b = FZ$. □

Now we derive a formula for $B(m)$ in terms of the fundamental matrix (see formula (2.6) in [30]):

Lemma 35. *There holds*

$$B(m) = \mathrm{cxhl}\left\{ \frac{FZ}{a_m(Z)^{\frac{1}{2}}},\ Z \in T_m(T^*X)/\ker(a_m) \right\}. \tag{65}$$

Proof. We have to compare (15) with (65).

First, let $b \in \ker(a_m)^{\perp_{\omega_X}}$ with $a_m^*(\mathcal{I}(b)) \leqslant 1$. By the proof of Lemma 34, there exists $Z \in T_m(T^*X)/\ker(a_m)$ such that $-\mathcal{I}(b) = b_0 = a_m(Z, \cdot)$. Using that $a_m^*(b_0) \leqslant 1$, we obtain $a_m(Z) \leqslant 1$, hence $b_0 = \lambda a_m(Z,\cdot)/a_m(Z)^{\frac{1}{2}}$ where $|\lambda| \leqslant 1$. It follows that $b = -\mathcal{I}^{-1}(b_0) = \lambda FZ/a_m(Z)^{\frac{1}{2}}$. This proves that the cones given by (15) are included in those given by (65).

For the converse, we first notice that $FZ/a_m(Z)^{\frac{1}{2}} \in \ker(a_m)^{\perp_{\omega_X}}$, and thus it is also the case for any convex combination. Also, it follows from the definitions of $\mathcal{I}$, F, a_m^* and the Cauchy-Schwarz inequality that

$$\forall Z \in T_m(T^*X)/\ker(a_m), \quad a_m^*(\mathcal{I}(FZ)/a_m(Z)^{\frac{1}{2}}) \leqslant 1.$$

By convexity of a_m^*, we obtain that any convex combination b of elements of the form $FZ/a_m(Z)^{\frac{1}{2}}$ satisfies $a_m^*(\mathcal{I}(b)) \leqslant 1$. This concludes the proof. □

Proof of Proposition 33. As in Section 2.3, we work in a chart near m. Following the computations of Lemma 8, we have for any sequence of points $(m_j)_{j\in\mathbb{N}}$ such that $m_j - m \notin \ker(a_m)$,

$$\begin{aligned}\frac{1}{2}\omega_X(H_a(m_j),w) &= -\frac{1}{2}da(m_j)(w) = -a_m(m_j-m,w) + o(m_j-m)\\ &= \omega_X(F(m_j-m),w) + o(m_j-m),\end{aligned}$$

which implies

$$H_{\sqrt{a}}(m_j) = \frac{1}{2}\frac{H_a(m_j)}{a(m_j)^{\frac{1}{2}}} = \frac{F(m_j-m)}{a_m(m_j-m)^{\frac{1}{2}}} + o(1). \tag{66}$$

Choosing $m_j = m + \varepsilon_j Z$ with $\varepsilon_j \to 0$, we obtain

$$H_{\sqrt{a}}(m_j) \underset{j\to+\infty}{\longrightarrow} \frac{FZ}{a_m(Z)^{\frac{1}{2}}}$$

which proves that $B(m) \subset \mathcal{H}\sqrt{a}(m)$ according to Lemma 35.

Conversely, since F is a linear isomorphism (see Lemma 34), it is not difficult to see that any limit of $\frac{F(m_j-m)}{a_m(m_j-m)^{\frac{1}{2}}}$ is of the form $\frac{FZ}{a_m(Z)^{\frac{1}{2}}}$. Using (66) and taking convex hulls, this proves that $\mathcal{H}\sqrt{a}(m) \subset B(m)$. □

References

[1] Andrei Agrachev, Davide Barilari and Ugo Boscain. *A comprehensive introduction to sub-Riemannian geometry.* Cambridge University Press, 2019.

[2] Paolo Albano, Antonio Bove, and Marco Mughetti. *Analytic hypoellipticity for sums of squares and the Treves conjecture.* Journal of Functional Analysis, vol. 274, no 10, p. 2725-2753, 2018.

[3] Augustin Banyaga and David E. Hurtubise. *A proof of the Morse-Bott lemma.* Expositiones Mathematicae, vol. 22, no 4, p. 365-373, 2004.

[4] Davide Barilari, Ugo Boscain and Robert W. Neel. *Small-time heat kernel asymptotics at the sub-Riemannian cut locus.* Journal of Differential Geometry, vol. 92, no 3, p. 373-416, 2012.

[5] Gérard Ben Arous. *Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus.* Annales Scientifiques de l'Ecole normale supérieure, vol. 21, no 3, p. 307-331, 1988.

[6] Jeff Cheeger, Mikhail Gromov and Michael Taylor. *Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds.* Journal of Differential Geometry, vol. 17, no 1, p. 15-53, 1982.

[7] Frank H. Clarke. *Optimization and nonsmooth analysis.* Society for Industrial and Applied Mathematics, 1990.

[8] Yves Colin de Verdière, Luc Hillairet and Emmanuel Trélat. *Spectral asymptotics for sub-Riemannian Laplacians, I: Quantum ergodicity and quantum limits in the 3-dimensional contact case.* Duke Mathematical Journal, vol. 167, no 1, p. 109-174, 2018.

[9] Yves Colin de Verdière and Cyril Letrouit. *Propagation of well-prepared states along Martinet singular geodesics.* To appear in Journal of Spectral Theory. Arxiv preprint arXiv:2105.03305.

[10] Johannes J. Duistermaat and Victor W. Guillemin. *The spectrum of positive elliptic operators and periodic bicharacteristics.* Inventiones Mathematicae, vol. 29, p. 39-80, 1975.

[11] Johannes J. Duistermaat and Lars Hörmander. *Fourier integral operators, II.* Acta Mathematica, vol. 128, no 1, p. 183-269, 1972.

[12] Semyon Dyatlov and Maciej Zworski. *Mathematical theory of scattering resonances.* Graduate Studies in Mathematics, vol. 200. American Mathematical Society, 2019.

[13] Charles L. Fefferman and Duong Hong Phong. *On positivity of pseudo-differential operators.* Proceedings of the National Academy of Sciences of the United States of America, vol. 75, no 10, p. 4673-4674, 1978.

[14] Clotilde Fermanian Kammerer and Véronique Fischer. *Quantum evolution and sub-Laplacian operators on groups of Heisenberg type.* Journal of Spectral Theory, vol. 11, no 3, p. 1313–1367, 2021.

[15] Peter Greiner, David Holcman, and Yakar Kannai. *Wave kernels related to second-order operators.* Duke Mathematical Journal, vol. 114, no 2, p. 329-386, 2002.

[16] Lars Hörmander. *Hypoelliptic second order differential equations.* Acta Mathematica, vol. 119, no 1, p. 147-171, 1967.

[17] Lars Hörmander. *Fourier integral operators, I.* Acta Mathematica, vol. 127, no 1, p. 79-183, 1971.

[18] Lars Hörmander *On the existence and the regularity of solutions of linear pseudodifferential equations.* L'enseignement Mathématique, XVII, p. 99-163, 1971.

[19] Lars Hörmander. *The analysis of linear partial differential operators, vol. I-IV.* Springer Science & Business Media, 2007.

[20] Victor Ivrii. *Microlocal Analysis, Sharp Spectral Asymptotics and Applications I: Semiclassical Microlocal Analysis and Local and Microlocal Semiclassical Asymptotics.* Springer Nature, 2019.

[21] David S. Jerison and Antonio Sánchez-Calle. *Estimates for the heat kernel for a sum of squares of vector fields.* Indiana University mathematics journal, vol. 35, no 4, p. 835-854, 1986.

[22] Yakar Kannai. *Off diagonal short time asymptotics for fundamental solution of diffusion equation.* Communications in Partial Differential Equations, vol. 2, no 8, p. 781-830, 1977.

[23] Bernard Lascar. *Propagation des singularités pour des équations hyperboliques à caractéristiques de multiplicité au plus double et singularités masloviennes.* American Journal of Mathematics, vol. 104, no 2, p. 227-285, 1982.

[24] Bernard Lascar and Richard Lascar. *Propagation des singularités pour des équations hyperboliques à caractéristiques de multiplicité au plus double et singularités masloviennes II.* Journal d'Analyse Mathématique, vol. 41, no 1, p. 1-38, 1982.

[25] Camille Laurent and Matthieu Léautaud. *Tunneling estimates and approximate controllability for hypoelliptic equations.* To appear in Memoirs of the American Mathematical Society.

[26] Rémi Léandre. Minoration en temps petit de la densité d'une diffusion dégénérée. Journal of Functional Analysis, vol. 74, no 2, p. 399-414, 1987.

[27] Nicolas Lerner. *Metrics on the phase space and non-selfadjoint pseudo-differential operators.* Springer Science & Business Media, 2011.

[28] Alessio Martini, Detlef Müller and Sebastiano Nicolussi Golo. *Spectral multipliers and wave equation for sub-Laplacians: lower regularity bounds of Euclidean type.* To appear in Journal of the European Mathematical Society. ArXiv preprint arXiv:1812.02671.

[29] Richard B. Melrose. *The wave equation for a hypoelliptic operator with symplectic characteristics of codimension two.* Journal d'Analyse Mathématique, vol. 44, no 1, p. 134-182, 1984.

[30] Richard B. Melrose. *Propagation for the wave group of a positive subelliptic second-order differential operator.* In : Hyperbolic equations and related topics. Academic Press, p. 181-192, 1986.

[31] Richard B. Melrose and Gunther A. Uhlmann. *Microlocal structure of involutive conical refraction.* Duke Mathematical Journal, vol. 46, no 3, p. 571-582, 1979.

[32] Richard Montgomery. *Abnormal minimizers.* SIAM Journal on Control and Optimization, vol. 32, no 6, p. 1605-1620, 1994.

[33] Richard Montgomery. *Hearing the zero locus of a magnetic field.* Communications in Mathematical Physics, vol. 168, no 3, p. 651-675, 1995.

[34] Richard Montgomery. *A tour of subriemannian geometries, their geodesics and applications.* American Mathematical Soc., 2002.

[35] Adrian Nachman. *The wave equation on the Heisenberg group.* Communications in Partial Differential Equations, vol. 7, no 6, p. 675-714, 1982.

[36] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics. II Fourier Analysis, Self-adjointness* New York, Academic Press, 1975.

[37] Didier Robert. *Autour de l'approximation semi-classique.* Progress in Mathematics, Vol. 68. Birkhäuser, 1987.

[38] Linda P. Rothschild and Elias M. Stein. *Hypoelliptic differential operators and nilpotent groups.* Acta Mathematica, vol. 137, no 1, p. 247-320, 1976.

[39] Nikhil Savale. *Spectrum and abnormals in sub-Riemannian geometry: the 4D quasi-contact case.* ArXiv preprint arXiv:1909.00409, 2019.

[40] Michael E. Taylor *Noncommutative harmonic analysis* Math. Surveys Monogr. 22, Amer. Math. Soc., Providence, 1986.

[41] Emmanuel Trélat. *Some properties of the value function and its level sets for affine control systems with quadratic cost.* Journal of Dynamical and Control Systems, vol. 6, no 4, p. 511-541, 2000.

[42] François Treves. *Symplectic geometry and analytic hypo-ellipticity.* In Proceedings of Symposia in Pure Mathematics, volume 65, p. 201-219. American Mathematical Society, 1999.

[43] Maciej Zworski. *Semiclassical analysis.* American Mathematical Soc., 2012.