REGULARITY PROPERTIES, REPRESENTATION OF SOLUTIONS AND SPECTRAL ASYMPTOTICS OF SYSTEMS WITH MULTIPLECTIES

ILIA KAMOTSKI AND MICHAEL RUZHANSKY

Abstract. Properties of solutions of generic hyperbolic systems with multiple characteristics with microlocally diagonalizable principal part are investigated. Solutions are represented as a Picard series with terms in the form of iterated Fourier integral operators. It is shown that this series is an asymptotic expansion with respect to smoothness under quite general geometric conditions on characteristics. Both constant and variable multiplicities are allowed. Propagation of singularities is described and sharp regularity properties of solutions are obtained. Results are applied to establish regularity estimates for scalar weakly hyperbolic equations with involutive characteristics. They are also applied to derive the first and second terms of the spectral asymptotics for the corresponding self-adjoint elliptic systems.

Keywords: Hyperbolic systems; Systems with multiplicities; Fourier integral operators; Regularity of solutions; Elliptic systems; Spectral asymptotics;

MSC 2000 classification: 35S30, 35L45, 35L30, 35C20, 58J40

1. Introduction

Let $X$ be a smooth manifold without boundary of dimension $n \geq 3$. Let $P$ be a pseudo-differential operator of order one acting on half-densities on $m$-dimensional sections of vector bundles on $X$. We consider the following Cauchy problem for $u = u(t, x)$:

\[
\begin{cases}
  i\partial_t u - Pu = 0, & (t, x) \in [0, T] \times X, \\
  u|_{t=0} = u_0.
\end{cases}
\]

It is well known that if equation (1.1) is strictly hyperbolic, the system can be diagonalized and its solution can be given as a sum of Fourier integral operators applied to the Cauchy data (e.g. [6]). An important question that has been studied over many years is what happens when $P$ has multiple characteristics.

Since we will be mostly interested in local properties of solutions, we may already assume that $P$ acts on functions, and can think of it as an $m \times m$ matrix of pseudo-differential operators of order one and we think of $u_0$ as of an $m$-vector.

The analysis of the paper in the cases of lower dimensions $n = 1$ and $n = 2$ can be dealt with in a similar manner. We will point out the necessary modifications later in the introduction.

Let $A(x, \xi)$ denote the principal symbol of $P$. If $A$ is a diagonal real matrix, properties of system (1.1) have been studied by many authors. For example, in [19] and [18] Kumano-go and coauthors used the calculus of Fourier integral operators
with multi-phases to show that the Cauchy problem (1.1) is well-posed in $L^2$, Sobolev spaces $H^s$, and to study its propagation of singularities. Systems with symmetric principal part $A$ have been extensively studied as well (e.g. [17], [13], etc.) In a generic situation, such systems have double characteristics, and their normal forms have been found by Braam and Duistermaat [2]. Recently, Colin de Verdière [4] used these representations to derive some asymptotic properties of such systems. Polarization properties of similar systems have been studied by Dencker in [5].

More elaborate analysis of system (1.1) becomes possible if one assumes that the principal symbol matrix $A(x, \xi)$ is symmetric and smoothly microlocally diagonalizable. This means that matrix $A(x, \xi)$ can be smoothly diagonalized microlocally in cones in $T^*X$, with smooth real eigenvalues $a_j(x, \xi)$ and smooth eigenspaces. In this situation Rozenblum showed that the Picard series for this problem gives an expansion with respect to smoothness in the case of non-involutive characteristics. In other words, this assumption means that if $a_j(x, \xi) = a_k(x, \xi)$ for $j \neq k$, then the Poisson bracket $\{a_j, a_k\}(x, \xi) \neq 0$. This also means that at all points of multiplicity, bicharacteristics intersect characteristics surfaces transversally. However, this condition is non-generic even for diagonal systems. For example, it is clear that if one of characteristics has a fold, there may be a point where this transversality condition fails, and it is not possible to remove it by small perturbations.

One purpose of this paper is to present results removing the transversal intersection condition. Also, we will not assume that the principal part is symmetric but that the eigenvalues (characteristics) are real, i.e. equation (1.1) is weakly hyperbolic. Obviously, this includes the case of systems with symmetric principal part. We will allow characteristics to be involutive of finite type and some characteristics to be identically equal to each other. Operators satisfying our Condition $C$ introduced below will be generic in the class of smoothly microlocally diagonalizable systems. Below we will explain that the microlocal diagonalizability condition is quite natural when considering weakly hyperbolic scalar equations with Levi conditions (Examples 2.1 and 2.2). This is also the case for differential systems and Maxwell equations (e.g. [1]).

We will also investigate regularity properties of solutions to system (1.1) in this generic setting. $L^p$–properties of solutions of the strictly hyperbolic equations have been studied for many years. They have several important applications for nonlinear equations and some problems in harmonic analysis. In the strictly hyperbolic case the question may be reduced to the corresponding question for Fourier integral operators. Regularity properties of non-degenerate Fourier integral operators have been established by Seeger, Sogge and Stein in [27]. They showed that a Fourier integral operator $T$ of order zero satisfying the local graph condition, is locally bounded from $(L^p_{\text{comp}})^{\alpha}$ to $L^{\alpha}_{\text{loc}}$ for $1 < p < \infty$ and $\alpha = (n - 1)|1/p - 1/2|$. As a consequence one readily deduces that if system (1.1) is strictly hyperbolic, there is a loss of $\alpha$ derivatives in $L^p$, i.e. $u_0 \in (L^p_{\text{comp}})^{\alpha}$ implies $u(t, \cdot) \in L^{\alpha}_{\text{loc}}$. Moreover, if at least one of characteristic roots $a_j$ is elliptic, the loss of $\alpha$ derivatives is sharp. If none of $a_j$’s is elliptic, this result can be improved and is related to certain geometric properties of the corresponding canonical relation (see [25]).

Our Theorem 2.2 will establish a similar property for systems (1.1) with multiplicities. Moreover, this will imply $L^p$ estimates for scalar weakly hyperbolic equations
with involutive characteristics. It is known that in general weakly hyperbolic cases one often loses regularity even in $L^2$. However, in the case of involutive characteristics the equation can be reduced to a matrix form with the diagonal principal part and in Theorem 2.3 we will present $L^p$ estimates for such equations. This will, on one hand, extend the $L^p$ result of Seeger, Sogge and Stein to weakly hyperbolic equations and systems with multiplicities, while on the other hand will establish $L^p$ estimates for systems considered by Kumano-go, Rozenblum, and others. The result will be general for scalar weakly hyperbolic equations satisfying Levi conditions with characteristics satisfying Condition C below.

Note that if a scalar operator of order $m$ is strictly hyperbolic and we rewrite it in the form of the first order system (1.1), we can diagonalize the corresponding operator $P$ together with lower order terms (e.g. Kumano-go [18]) and split it into $m$ first order scalar equations, for which many things are known. However, in the case of multiple characteristics this is impossible, so a more elaborate analysis is needed. To carry out such analysis is one of the aims of the present paper.

Now we will formulate our main assumption. Let us define operator $H_{a_j} f = \{a_j, f\}$, $j = 1, \ldots, m,$

where

$$H_g(f) = \{g, f\} = \sum_{k=1}^{n} \left( \frac{\partial g}{\partial \xi_k} \frac{\partial f}{\partial x_k} - \frac{\partial g}{\partial x_k} \frac{\partial f}{\partial \xi_k} \right)$$

is the usual Poisson bracket. Our assumption is that at points of multiplicity $a_j = a_k$ of non-identical characteristics $a_j$ and $a_k$, bicharacteristics of $a_j$ intersect level sets $\{a_k = 1 \}$ with finite order, i.e. $H^M_{a_j} a_k \neq 0$ for some $M$, at points where $a_j = a_k$. In other words, we do allow involutive characteristics of finite type. Let us now reformulate this condition is a slightly different form.

**Condition C:**

Suppose that the principal part $A(x, \xi)$ of operator $P$ is microlocally smoothly diagonalizable with smooth real eigenvalues $a_1(x, \xi), \ldots, a_m(x, \xi)$. Suppose that there exists $M \in \mathbb{N}$ such that if for some $j$ and $k$, we have $a_j(x, \xi) = a_k(x, \xi)$, and $a_j$ and $a_k$ are not identically equal near $(x, \xi)$, then

$$H^\lambda_{a_j} a_k(x, \xi) = \{ a_j, \{ a_j, \ldots \{ a_j, a_k \} \ldots \} (x, \xi) \neq 0,$$

for some $\lambda \leq M$.

We note that in this paper we do not assume the manifold $X$ to be compact, with the exception of Theorem 2.5 on spectral asymptotics. The number $\lambda = \lambda(x, \xi)$ in (1.2) is a function of $(x, \xi)$ and **Condition C** says that it is bounded. However, if the function $\lambda(x, \xi)$ is locally bounded, all the results of the paper remain valid even if we allow $\lambda(x, \xi)$ to grow at infinity. Indeed, since all our results are local, we may reduce this situation to **Condition C** by taking $M$ to be the maximum of $\lambda(x, \xi)$ over a compact set under consideration.
We also note here that the non-involutive case considered by Rozenblum [24] requires (1.2) to hold with \( M = 1 \). Strictly hyperbolic case is also covered by this condition (in which case we set \( M = 0 \)). The case of \( a_j \) and \( a_k \) defining glancing hypersurfaces (as in Melrose [20]) corresponds to \( M = 2 \).

A couple of other remarks are in order. First, since most of the analysis of the paper is local, operators in regions of small frequencies are smoothing. Thus, for small frequencies we may somewhat relax Condition \( C \) including assumptions on the smoothness of \( a_j \)'s. Second, most results of this paper extend to the case of operator \( P \) dependent on \( t \) as well. The difference with the present paper is that assumption (1.2) changes and does not seem to have a clear geometric meaning. In particular, it should contain some terms involving time derivatives of characteristics. This matter is technical and will be addressed elsewhere.

In Section 2 we will give several examples of characteristics satisfying Condition \( C \), in particular those arising from weakly hyperbolic scalar equations with involutive characteristics. Such equations and propagation of their singularities have been analysed in [3], [21], [22], [11], [14], etc. We will also establish estimates in \( L^p \) and other spaces for the weakly hyperbolic equations or systems satisfying Condition \( C \).

It is interesting to note that conditions similar to Condition \( C \) appeared in the study of subelliptic operators (e.g. Hörmander [10, Chapter 27]). For instance, in the case of a \( 2 \times 2 \) system \( P \) with characteristics \( a_1 \) and \( a_2 \), we can consider operator \( Q \) with principal symbol \( q = a_1 + ia_2 \). Then microlocal subellipticity of \( Q \) implies Condition \( C \), with some \( M \) dependent on the loss of regularity for \( Q \), which, therefore, implies the Weyl formula for \( P \) (Theorem 2.5), regularity estimates for solutions to (1.1) and all other results of this paper.

Now we will give an informal explanation of the strategy of our analysis. First, let us follow [24] to explain that microlocal diagonalizability implies a local one on some cover \( \tilde{X} \) of \( X \) with finitely many sheets. For this argument we assume that \( X \) is compact. Since all the analysis of this paper will be local, if \( X \) is not compact, we can always assume that the amplitude of \( P(x,D) \) is compactly supported in \( x \).

Let a pseudo-differential operator \( P(x,D) \) of order one act on sections of an \( m \)-dimensional vector bundle \( E \). Let \( L^2(E) \) be the space of sections of half-densities on \( E \). Let \( E' \) be the lifting of \( E \) to \( T^*X \).

Let us assume everywhere that the principal symbol \( A(x,\xi) \) is microlocally (smoothly) diagonalizable. This means that microlocally in \( \Lambda \subset T^*X \) such that \( E'\Lambda \cong \Lambda \times \mathbb{C}^m \), principal symbol \( A(x,\xi)\Lambda \) has \( m \) smooth real eigenvalues \( a_j(x,\xi) \) and one dimensional eigenspaces \( V_j(x,\xi) \), and such diagonalizations are compatible in intersecting cones. In this situation Lemma 5.1 insures that there is a cover \( \tilde{X} \) of \( X \) with finitely many sheets such that the principal symbol of the lifting of \( P(x,D) \) to \( T^*\tilde{X} \) can be globally diagonalized modulo lower order terms. Note that since dimensions of \( X \) and \( \tilde{X} \) are the same and because of formula (5.1) all our results on \( \tilde{X} \) will imply the corresponding results on \( X \).

Therefore, without loss of generality we may freely assume that the principal symbol \( A \) of operator \( P \) may be smoothly diagonalized over compact subsets of \( X \), that is

\[
P = A + B, A = \text{diag}\{A_1, \ldots, A_m\},
\]
where \( A_j \in \Psi^1 \) are scalar pseudo-differential operators with principal symbols \( a_j(x, \xi) \) (eigenvalues of \( A(x, \xi) \)). Here \( a_j \)'s may be identically equal to each other or may intersect with any finite order, according to our Condition C. Thus, we allow characteristics to be of variable multiplicities while they may include some blocks of constant multiplicity.

Here \( B \) is an \( m \times m \) matrix of pseudo-differential operators or order zero. We can make \( B_{jj} = 0 \) for \( 1 \leq j \leq m \) by adding these terms to the diagonal of \( A \). For the moment we will assume that none of \( a_j \)'s are identical. Otherwise, if some of \( a_j \)'s are identically equal to each other locally at some points, the construction is slightly different, but the statements of Theorems 2.2–2.5 remain valid. Moreover, if the system has constant multiplicity the arguments are much simpler if we use the standard representation of solutions as Fourier integral operators from, for example, [29, VIII]. In the case of variable multiplicities we use Theorem 2.1. This will be carried out in detail in Section 3 in the proof of Theorem 2.2.

We also note that this argument of microlocal diagonalization implying the local one on a covering space is the reason for us to restrict the considerations to manifolds \( X \) of dimension \( n \geq 3 \). For \( n < 3 \) we may either assume that the principal part of \( P \) is locally smoothly diagonalizable with all the same conclusions of the paper or make all the arguments in the microlocal setting. We omit details of the latter since they would make the presentation much more technical.

Now, substitution \( U = e^{-iAt}V \) transforms equation

\[
\begin{align*}
\begin{cases}
iU' = PU, \\
U|_{t=0} = I,
\end{cases}
\end{align*}
\]

into equation

\[
\begin{align*}
\begin{cases}
V' = Z(t)V, \\
V|_{t=0} = I,
\end{cases}
\end{align*}
\]

with \( Z(t) = -ie^{iAt}Be^{-iAt} \). Writing the Picard series for problem (1.3), we obtain the expansion

\[
V(t) = I + \int_0^t Z(t_1)dt_1 + \int_0^t \int_0^{t_1} Z(t_1)Z(t_2)dt_2dt_1 + \cdots
\]

A general term of this series is

\[
(1.5) \quad Q_l = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{l-1}} e^{iA_{j_1}t_1}b_{j_1j_2}e^{iA_{j_2}(t_2-t_1)} \cdots dt_1 \ldots dt_l.
\]

It is easy to see that \( ||Q_l||_{L^2 \to L^2} \leq C_l/l! \) for some \( C > 0 \), and that series (1.4) converges in \( L^2 \) and in \( H^s \). Using the notion of a multi-phase for Fourier integral operators, Kumano-go et al. ([19], [18]) studied propagation of singularities of \( Q_1 \).

Instead of introducing multi-phases for \( Q_1 \), we will analyse the structure of operators \( Q_l \) in more detail and will show their regularising properties in Sobolev spaces under Condition C. Here, contrary to the non-involutive case (when \( M = 1 \)), we do not have good control on the singular supports of integral kernels of operators \( Q_l \), so more elaborate geometric analysis is required.

In fact, our Theorem 2.1 asserts that \( Q_l(t) \) maps \( L^2 \) to some Sobolev space \( H^{p(l)} \), and that \( p(l) \to \infty \) as \( l \to \infty \). Then we will show that this allows to treat the Picard
expansion almost as a series with finitely many terms, with many conclusions, such as $L^p$ estimates for solutions and spectral asymptotics of $P$.

In particular, $L^p$ estimates will follow from a general principle which we will prove in Theorem 3.1 for the equation

$$\begin{cases}
u' - Z(t) \nu = f, \\
u(0) = u_0.
\end{cases}$$

In Corollaries 3.2 and 3.3 we will show regularity of solutions of this equation for pseudo-differential operators of order zero $Z(t) \in \Psi^0$ or for Fourier integral operators of negative orders $Z(t) \in I^{-\epsilon}$. However, in problem (1.3) in general $Z(t) \in I^0$ is a family of Fourier integral operators of order zero, so we need to use their structural properties to show that assumptions of Theorem 3.1 are satisfied. This part is an extension of the method presented by the authors in [16].

Everywhere in this paper $\Psi^\mu = \Psi^\mu_{1,0}(X)$ will denote the space of classical pseudo-differential operators of order $\mu$ of type $(1,0)$. Locally these operators have amplitudes $c = c(x, y, \xi)$ satisfying

$$|\partial_\alpha^\mu \partial_\beta^\mu \partial_\gamma^\mu c(x, y, \xi)| \leq C_{\alpha, \beta, \gamma} (1 + |\xi|)^{\mu - |\gamma|}$$

for all $\alpha, \beta, \gamma$ and all $\xi$, locally uniformly in $x$ and $y$. In this case we will also write $c \in S^\mu$.

The space of Fourier integral operators of order $\mu$ with amplitudes of type $(1,0)$ will be denoted by $I^\mu$. All Fourier integral operators in this paper will be non-degenerate, which means that its canonical relation satisfies the local graph condition, i.e. it is a graph of a symplectic diffeomorphism from $T^*X$ to itself. Constants $C$ may be different in different formulae throughout the paper. We will use the following notation for norms and spaces. By $L^s_a$ we will denote the Sobolev space of functions $f$ such that $(I - \Delta)^{s/2} f \in L^p$. For a function $f$ we will denote its $L^p$-norm by $\|f\|_{L^p}$ and its Sobolev $H^s = L^2_s$ norm by $\|f\|_{H^s}$ or simply by $\|f\|_s$ when the meaning is clear. By $\|f\|_{L^p_{loc}}$ we will denote any localisation of the $L^p$-norm. So, an estimate of the form $\|f\|_{L^p_{loc}} \leq C$ will mean that we have estimates $\|\chi f\|_{L^p} \leq C$ for all $\chi \in C_0^\infty$, with constants $C$ possibly dependent on $\chi$. If $T$ is an operator, by $\|T\|_s$ we will denote its operator norm from $L^2$ to $H^s$. As usual, we denote $D_x = -i\partial_x$ and for $u = u(t, x)$ by $u'$ we always denote the time derivative $\partial_t u(t, x) = \partial_u \frac{u}{t}(t, x)$. By $\nabla_x$ or $\partial_x$ we denote the gradient.

Results of this paper can be established also in the case of operators $P$ dependent on $t$. This will be the subject of a separate paper.

We would like to thank Ari Laptev for drawing our attention to Rozenblum’s paper [24], Grigori Rozenblum for several valuable references and Yuri Safarov for discussions. We would also like to thank the referees for many valuable remarks leading to an improvement of the manuscript.

The research was supported by EPSRC grant GR/R67583/01.

2. Main results

Now we will formulate our results concerning terms of the Picard series (1.4) and solutions to systems (1.1) and (1.3). We will also give a Weyl formula for the spectral function for elliptic self-adjoint operators $P$. 

6
Our first main result will be Theorem 2.1 on the smoothing properties of terms $Q_l$. The other two important results will be Theorem 2.2 on the $L^p$-regularity of solutions to system (1.1) and Theorem 2.5 on the spectral asymptotics for elliptic self-adjoint operator $P$ satisfying Condition C. To obtain $L^p$-estimates, we use Theorem 3.1, which we regard as a general principle behind regularity estimates for general Cauchy problems based on several natural properties of the right hand side operators $Z(t)$. We will illustrate its use in several situations in Corollaries 3.2 and 3.3. Theorem 2.4 is a statement on the propagation of singularities for operators $Q_l$ or, more generally, of Fourier integral operators in which the frequency integration is performed over a cone rather than over the whole space. Note that these results will hold for microlocally smoothly block-diagonalizable operators with any (finite) geometry of characteristics, i.e. characteristics satisfying our Condition C. As another consequence of this analysis, we can obtain the first and the second term of the spectral asymptotics of operator $P$. This result was announced by the authors in [15] and is proved here in Theorem 2.5.

The $L^2$-boundedness of non-degenerate Fourier integral operators of order zero immediately implies that operators $Q_l$ are continuous from $L^2_{\text{comp}}$ to $L^2_{\text{loc}}$. The following theorem establishes a smoothing property of operators $Q_l$ under Condition C.

**Theorem 2.1.** Assume that not all of $a_j$’s are identically equal to each other. Let Condition C be satisfied, that is assume that there is $M$ such that for any $(x, \xi) \in T^*X$ and any $1 \leq j, k \leq m$, $j \neq k$, with $a_j$ and $a_k$ not identically the same near $(x, \xi)$ and $a_j(x, \xi) = a_k(x, \xi)$, we have

$$\{a_j, \{a_j, \cdots \{a_j, a_k\}\}\} (x, \xi) \neq 0$$

for some number $\lambda \leq M$. Then operators $Q_l$ in (1.5) are continuous from $L^2_{\text{comp}}$ to $H^{N(l)}_{\text{loc}}$, where $N(l) \to +\infty$ as $l \to +\infty$.

The proof of the theorem will yield the expression

$$(2.1) \quad N(l) = \frac{(-3n/2 - 2)(3l/2 - 1 - n) + ([l/2] - n - 1)(l/(2M) - n - 1)}{(3l/2 - 2n - 2 + l/(2M))}$$

for large $l$. It can be readily seen that

$$(2.2) \quad N(l) \sim \frac{l}{6M + 2} \quad \text{as} \quad l \to \infty,$$

while we also have $N(l) \geq 0$ for all $l$ because $Q_l$ are locally continuous in $L^2$. However, the exact order $N(l)$ can be improved. We leave the statement of Theorem 2.1 in such asymptotic form because for the applications in this paper it is only important that $N(l)$ increases to infinity as $l \to \infty$. The exact formula for $N(l)$ for all $l$ will be treated in a subsequent paper.

Theorem 2.1 implies, in particular, that the series (1.4), i.e. the series

$$V(t) = I + Q_1(t) + Q_2(t) + \cdots$$

is a series with respect to smoothness. This fact allows one to refine the study of propagation of singularities and regularity properties of solutions to systems (1.1) and (1.3).
It turns out that smoothing properties of $Q_l$ in Theorem 2.1 can be used to establish local $L^p$ properties of solutions to systems with multiplicities (1.1). We may assume that not all of $a_j$’s are identically equal to each other since otherwise the system has constant multiplicity and the results in the remaining part of this section follow from representation in [29, VIII] and properties of Fourier integral operators.

Our next result concerns regularity of solutions to Cauchy problem (1.1). In this theorem we also allow the lower order term $B$ to depend on $t$.

**Theorem 2.2.** Let $1 < p < \infty$ and $\alpha = (n - 1)\lfloor 1/p - 1/2 \rfloor$. Let $P = P(t, x, D_x)$ be an $m \times m$ matrix of classical pseudo-differential operators of order one. Let

$$P(t, x, D_x) = A(x, D_x) + B(t, x, D_x),$$

where $A$ is a matrix of pseudo-differential operators of order one and $B$ is a matrix of operators of order zero. Assume that the matrix $A$ is smoothly microlocally diagonalizable, with smooth eigenspaces and real eigenvalues $a_j(x, \xi)$, satisfying Condition C. Then for any compactly supported $u_0 \in L^p_\alpha \cap L^2_{\text{comp}}$, the solution $u = u(t, x)$ of the Cauchy problem

$$(2.3) \quad i\partial_t u - P(t, x, D_x)u = 0, \quad u(0) = u_0,$$

satisfies $u(t, \cdot) \in L^p_{\text{loc}}$, for all $0 \leq t \leq T$. Moreover, there is a constant $C_T > 0$ such that

$$(2.4) \quad \sup_{0 \leq t \leq T} \|u(t, \cdot)\|_{L^p_{\text{loc}}} \leq C_T \|u_0\|_{L^p_\alpha}.$$ 

We note that it is sufficient to only assume that $A$ is microlocally diagonalizable. Then by Lemma 5.1 we can first have the statement of Theorem 2.2 on the cover $\bar{X}$. Then, using formula (5.1) and the fact that $\dim \bar{X} = \dim X$, we get the same conclusion on $X$. In general, the loss of $\alpha$ derivatives in Theorem 2.2 is sharp. However, under certain further geometric conditions on characteristics this order may be improved (see [25] for details).

As a consequence, if the Cauchy data $u_0$ is compactly supported, we obtain local estimates in other spaces as well:

- $u_0 \in L^p_{s+\alpha}$ implies $u(t, \cdot) \in L^p_s$, $s \in \mathbb{R}$.
- $u_0 \in \text{Lip}(s + (n - 1)/2)$ implies $u(t, \cdot) \in \text{Lip}(s)$.
- Let $1 < p \leq q \leq 2$. Then $u_0 \in L^p_{s-1/q+n/p-(n-1)/2}$ implies $u(t, \cdot) \in L^q_s$. The dual result holds for $2 \leq p \leq q < \infty$.

The proof of Theorem 2.2 is based on the other main result Theorem 3.1 which we will discuss in the next section. Estimates in other spaces follow by standard methods of harmonic analysis ([28]). The first one is a consequence of the composition formulae for Fourier integral operators, while the second one follows from a suitable representation of the Lipschitz space with Riesz multipliers. The last statement follows by interpolation from Theorem 2.2 and the $H^1$ to $L^2$ continuity of operators of order $-n/2$. Here $H^1$ stands for the Hardy space and the $H^1$ to $L^2$ continuity is a consequence of composition formulae and the Hardy space version of the Hardy-Littlewood-Sobolev theorem for fractional integrals (see also [25] for a detailed discussion).

Now we will give some examples where Condition C holds while the transversality condition ($M = 1$) fails. This is, for example, the case when pairs $a_j, a_k$ define
glancing hypersurfaces or when we consider Maxwell systems with variable coefficients. Below we will concentrate on systems arising from scalar weakly hyperbolic equations with Levi conditions.

**Example 2.1.** In scalar equations with Levi conditions studied by Chazarain [3], Mizohata-Ohya [21], Zeman [31], one assumed that

\[ \{a_j, a_k\} = C_{jk}(a_j - a_k). \]

It is clear that in this situation \(a_j(x, \xi) = a_k(x, \xi)\) implies \(\{a_j, a_k\}(x, \xi) = 0\). However, in a general case when \(C_{jk}(x, \xi)\) is non-constant, Condition \(C\) is satisfied generically.

**Example 2.2.** Let \(L\) be a scalar operator with involutive characteristics. More precisely, let us denote \(\partial_j = D_t + \lambda_j(t, x, Dx)\) and let

\[
L = \partial_1 \cdots \partial_m + \sum_{k<m} b_{j_1 \cdots j_k} \partial_{j_1} \cdots \partial_{j_k} + c,
\]

where \(b_{j_1 \cdots j_k}(t, x, Dx), c(t, x, Dx) \in \Psi^0\) are pseudo-differential operators of order zero for all \(t \in [0, T]\). We will assume that symbols of all operators are infinitely differentiable with respect to \(t\) in the topology of symbols of the corresponding order. Let us assume that operator \(L\) has involutive characteristics, i.e. that

\[
[\partial_j, \partial_k] = \partial_j \partial_k - \partial_k \partial_j = \alpha_{jk} \partial_j + \beta_{jk} \partial_k + \gamma_{jk},
\]

where \(\alpha_{jk}, \beta_{jk}, \gamma_{jk} \in \Psi^0\) are pseudo-differential operators of order zero. Then it was shown by Morimoto in [22] that the Cauchy problem for the equation \(Lu = f\) is diagonalizable (with \(1 + \sum_{j=1}^{m-1} m! / j! \) components). Even in the simplest case of characteristics not depending on \(t\), we have

\[ \{\lambda_j, \lambda_k\} = \alpha_{jk}(x, \xi)(\lambda_j - \lambda_k) + \gamma_{jk}, \]

similar to Example 2.1.

Propagation of singularities of systems with vanishing Poisson brackets has been also studied in these situations. For example, Iwasaki and Morimoto [14] studied propagation of singularities of \(3 \times 3\) systems, where the second Poisson bracket vanish. Also, Ichinose [11] studied \(2 \times 2\) systems with vanishing second Poisson brackets. Theorem 2.2 implies a precise statement on \(L^p\) estimates for solutions to such Cauchy problems.

**Theorem 2.3.** Let \(1 < p < \infty\), \(\alpha = (n - 1)|1/p - 1/2|\) and \(s \in \mathbb{R}\). Let \(L\) be as in (2.5) and suppose that the principal symbols \(a_j(x, \xi)\) of \(\lambda_j\) satisfy Condition \(C\) and do not depend on \(t\). Let \(u\) be a solution to the Cauchy problem

\[
\begin{aligned}
Lu &= 0, \\
\partial^j u(0, x) &= g_j(x), \quad 0 \leq j \leq m - 1,
\end{aligned}
\]

and let Cauchy data \(g_j \in L^p_{\alpha-j+s} \cap L^2_{\text{comp}}\) be compactly supported. Then \(u(t, \cdot) \in (L^p_t)^{m-1}\) for all \(t \in [0, T]\) and

\[
\sup_{t \in [0, T]} \|\partial^j u(t, \cdot)\|_{(L^p_t)^{m-1}} \leq C \sum_{j=0}^{m-1} \|g_j\|_{L^p_{\alpha-j+s+m-1}},
\]

where
where $\partial^j = \partial_{j_1} \ldots \partial_{j_k}$, $k \leq m - 1$, and $(j_1, \ldots, j_k)$ are permutations of some elements of $\{1, \ldots, m\}$.

Note that in the strictly hyperbolic case as well as in some very special cases of operator $L$ in (2.5) (e.g. when all $b_{j_1, \ldots, j_k}$ and $c$ are zero), following the method described by Treves in [30] and estimates for Fourier integral operators, it is possible to obtain the estimate for the Sobolev norm $\|u\|_{L^p_{s+m-1}}$ in the left hand side of (2.7).

Let us now give a final example of $L^p$ estimates for second order equations, which we will prove in the next section.

**Example 2.3.** Let us consider the second order equation

$$u'' + b(x, D_x)u' + c(x, D_x)u = 0,$$

where $b \in \Psi^1$, $c \in \Psi^2$. Let us denote $\langle x \rangle = (1 + |x|^2)^{1/2}$. Introducing $v = \left( \begin{array}{c} \langle D_x \rangle u \\ u' \end{array} \right)$, the matrix form of this equation is given by

$$v' = \left( \begin{array}{cc} 0 & \langle D_x \rangle \\ -\langle D_x \rangle^{-1}c & -b \end{array} \right) v.$$

Let $b_1$ and $c_2$ be principal symbols of $b$ and $c$. The equation is hyperbolic if $b_1^2 \geq 4c_2$, with multiple roots at $b_1^2 = 4c_2$. Assume that $b_1^2 - 4c_2 = \mu^2$, with $\mu \in S^1$ being a symbol of order one. Then characteristics $a_1, a_2$ satisfy $\{a_1, a_2\} = \frac{1}{2}\{b_1, \mu\}$, which may vanish, dependent on properties of $\mu$.

Let the Cauchy data be $u(0) = f_0$, $u'(0) = f_1$. Let $\alpha = (n - 1)|1/p - 1/2|$ and $1 < p < \infty$. If $\mu$ is elliptic, it is known that if $f_j \in L^p_{\alpha-j}$, $j = 0, 1$, are compactly supported, then $u(t, \cdot) \in L^p_{loc}$. Our result of Theorem 2.2 will imply that the same is true for any $\mu \in S^1$.

Note that in this example we may also require only microlocal diagonalization with the same conclusion.

Let us now discuss the propagation of singularities for operators $Q_t$. This result is essentially a reformulation of Rozenblum’s result to the case of finite geometry under Condition C. It is clear (also from multi-phase analysis) that singularities propagate along broken Hamiltonian flows. Let

$$J = \{j_1, \ldots, j_{l+1}\}, 1 \leq j_k \leq m, j_k \neq j_{k+1}.$$

Let $\Phi_J(t, x, \xi)$ be the corresponding broken Hamiltonian flow. It means that points follow bicharacteristics of $a_{j_1}$ until meeting the characteristic of $a_{j_2}$, and then continue along the bicharacteristics of $a_{j_2}$, etc. Note that singularities may accumulate if wave front sets for different broken trajectories project to the same point of $X$.

We can write

$$Q_t = \int_\Delta I(\bar{t}) d\bar{t},$$

where $\bar{t} = (t_1, \ldots, t_l)$ and

$$\Delta = \{0 \leq t_l \leq t_{l-1} \leq \cdots \leq t_1 \leq t\}$$

is a simplex in $\mathbb{R}^l$ and

$$I(\bar{t}) = Z(t_1) \circ \cdots \circ Z(t_l).$$
It is possible to treat it as a standard Fourier integral operator with the change of variables $\tilde{t} = \zeta |\xi|^{-1}$. Let $K$ be a cone in $\mathbb{R}^N = \mathbb{R}^{n+1}$. Let

$$Iu(x) = \int_K \int_Y e^{i\varphi(x,y,\theta)} a(x,y,\theta) u(y) dy d\theta$$

be a Fourier integral operator with integration over the cone $K$ with respect to $\theta$. Let $K_j$ be $K$ or a face of $K$. Let $\varphi_j(x,y,\theta_j) = \varphi|_{K_j}$, $\theta_j \in K_j$. Let $\Lambda_j \subset T^*X \times T^*X$ be a Lagrangian manifold with boundary:

$$\Lambda_j = \left\{ \left( x, \frac{\partial \varphi_j}{\partial x}, y, -\frac{\partial \varphi_j}{\partial y} \right) : \frac{\partial \varphi_j}{\partial \theta_j} = 0 \right\}.$$

For $G \subset T^*Y$, let

$$\Lambda_j(G) = \{ z \in T^*X : \exists \zeta \in G : (z, \zeta) \in \Lambda_j \}.$$

Then we have the following statement on the propagation of singularities.

**Theorem 2.4.** Let $u \in \mathcal{D}'(Y)$. Then $WF(Iu) \subset \bigcup_j \Lambda_j(WF(u))$.

The proof follows Hörmander [9] with minor differences.

From this, we can deduce first and second terms of the asymptotics of the spectral counting function for an elliptic self-adjoint operator $P$. Let us call $T$ a period of symbol $A(x,\xi)$ if there exists $J$ such that $j_1 = j_{l+1}$, and the trajectory of $\Phi_J$ is closed: $\Phi_J(T,x,\xi) = (x,\xi)$. Then we have the following extension of well-known results of Hörmander [8], Duistermaat–Guillemin [7], Safarov–Vassiliev [26], and Rozenblum [24].

**Theorem 2.5.** Assume that $X$ is compact. Let $P$ be an elliptic self-adjoint pseudodifferential operator of order one acting on half-densities on $m$-dimensional sections of vector bundles on $X$. Assume that Condition C is satisfied. Let $D$ be the set of $(x,\xi) \in T^*X$ such that there exist $T$ and $J$ such that $\Phi_J(T,x,\xi) = (x,\xi)$. Assume that the measure of $D$ is zero. Then for the spectrum of $P$ the following Weyl formula holds:

$$N(\lambda) = \# \{ j : \lambda_j < \lambda \} = c_n \lambda^n + c'_n \lambda^{n-1} + o(\lambda^{n-1}),$$

where $\lambda_j$ are eigenvalues of $P$.

We note that some results about the second term of spectral asymptotics have been obtained by Ivrii in [12] without the assumption of microlocal smooth diagonalizability, but under additional assumptions on the propagation of singularities for solutions of system (1.1) and on the flow from the set of multiplicities having measure zero. We do not impose such assumptions in this paper.

The proof of the $L^p$ estimates will be based on Theorem 3.1, which we regard as an independent result on regularity of solutions of partial differential equations. Proposition 5.3 concerns the measure of the set where a function is small given some information on the multiplicity of its roots. It will play a crucial role in the proof of the smoothing property in Theorem 2.1.
3. Regularity of solutions to Cauchy problems

In this section we will present a principle governing solutions of first order systems. Let $Z(t) \in \mathcal{L}(C_0^\infty(X), \mathcal{D}'(X))$, $t \in [0, T]$, be a time dependent family of operators. Let $W_0, W_1$ and $W$ be linear subspaces of $\mathcal{D}'(X)$ such that $W_0, W_1 \hookrightarrow W$. We will make different choices of these spaces in the future, dependent on the properties of operators $Z(t)$. Let us consider the Cauchy problem for $u = u(t, x)$:

\begin{equation}
\begin{cases}
u' - Z(t)\nu = r(t), & r(t) \in W_0, \\
u(0) \in W_1. & 
\end{cases}
\end{equation}

One is often interested in the following question:

\textit{If the right hand side and Cauchy data satisfy $r(t) \in W_0$ and $u(0, \cdot) \in W_1$, when do fixed time solutions $u(t, \cdot)$ of (3.1) belong to $W$?}

In general, some loss of regularity often happens in problem (3.1). So we will think of $W_0$ being the smallest, $W_1$ an intermediate, and $W$ the largest among these spaces.

Since we are mainly interested in the regularity properties of solutions rather than in their existence and uniqueness, we will assume the well-posedness of the Cauchy problem (3.1) in (the larger) space $L^2$. In fact, formula (3.6) and other properties of $Z(t)$ will insure that it is sufficient to consider (3.1) with zero Cauchy data. Thus, in the rest of this section we will always assume that $Z(t)$ and $r(t)$ are such that

(WP) the Cauchy problem (3.1) with $u(0) \equiv 0$ has the unique solution $u = u(t, x)$ such that $u(t) = u(t, \cdot) \in L^2$ for all $t \in [0, T]$.

The following theorem says that if operators $Z(t)$ have some structure, and solutions of Cauchy problem (3.1) with zero Cauchy data are in $W$, so will be solutions with Cauchy data from some sufficiently large space $W_1$.

**Theorem 3.1.** Let $W_0, W_1 \hookrightarrow W$ be linear subspaces of $\mathcal{D}'(X)$. Let $Z(t) \in \mathcal{L}(C_0^\infty(X), \mathcal{D}'(X))$, $t \in [0, T]$. Assume that

1. (Boundedness) $Z$ extends to an operator in $L^\infty([0, T], \mathcal{L}(L^2(X), L^2(X)))$, and $Z(t)$ extends to a continuous linear operator from $W_1$ to $W$, for all $t \in [0, T]$.
2. (Calculus) $Z(t_1) \circ \cdots \circ Z(t_l)$ extends to a continuous operator from $W_1$ to $W$, for all $l$ and for all $t_1, \ldots, t_l \in [0, T]$.
3. (Smoothing) There exists $l$ such that $Z(t) \int_0^t \int_0^{t_1} \cdots \int_0^{t_{l-1}} Z(t_1) \circ \cdots \circ Z(t_l) dt_1 \cdots dt_l$ extends to a continuous operator from $W_1$ to $W_0$, for all $t \in [0, T]$.
4. (Zero Cauchy data) Solutions $v = v(t, x)$ of the Cauchy problem

\begin{equation}
\begin{cases}
v' - Z(t)v = r(t), & r(t) \in W_0, \\
v(0) = 0, & 
\end{cases}
\end{equation}

satisfy (WP) and $v(t, \cdot) \in W$ for $t \in [0, T]$.

\textit{12}
Then the solution $u = u(t, x)$ of the Cauchy problem

\begin{equation}
\begin{cases}
u' - Z(t)u = r(t), & r(t) \in W_0, \\
u(0) \in W_1,
\end{cases}
\end{equation}

satisfies $u(t, \cdot) \in W$ for all $t \in [0, T]$.

Moreover, if $W_0, W_1, W$ are normed spaces and if solutions $v(t, \cdot)$ to (3.2) in (iv) satisfy $\|v(t, \cdot)\|_W \leq C\|r(t)\|_{W_0}$ for all $t \in [0, T]$, then also

$$\|u(t, \cdot)\|_W \leq C(\|u(0)\|_{W_1} + \|r(t)\|_{W_0}),$$

for all $t \in [0, T]$.

Conditions (i) and (ii) ensure that operators $Z(t)$ have some structure. Indeed, if $W_1 \subset W$ is different from $W$, (ii) does not follow from (i). In our typical applications, $Z(t)$ will be time dependent pseudo-differential or Fourier integral operators, and compositions in (ii) are essentially of the similar type as operator $Z(t)$. Condition (iii) is natural from the point of view of harmonic analysis, since integration with respect to a parameter often brings additional regularity. Condition (iv) ensures that solutions with zero Cauchy data and regular right hand side are also sufficiently regular.

Proof of Theorem 3.1. Let us consider the operator Cauchy problem

\begin{equation}
\begin{cases}
u'' - Z(t)U = R(t), & R(t) \in \mathcal{L}(W_1, W_0), \\
U(0) = I,
\end{cases}
\end{equation}

with a continuous linear operator $R(t) : W_1 \to W_0$. We claim that under assumptions of the theorem its solution operator $U(t)$ satisfies $U(t)w \in W$ for all $w \in W_1$. Moreover, if $W_0, W_1, W$ are normed spaces and the estimate for solutions in part (iv) of Theorem 3.1 holds, then there is a constant $C > 0$ such that

\begin{equation}
\|U(t)w\|_W \leq C(\|w\|_{W_1} + \|R(t)w\|_{W_0}),
\end{equation}

for all $w \in W_1$.

Before we prove this claim we note that the application of the claim with $R(t) = 0$ and assumption (iv) imply the statement of the theorem. Therefore, it is sufficient to prove the claim.

Let $U_0(t)$ be some partial solution to the problem

\begin{equation}
\begin{cases}
u''_0 - Z(t)U_0(t) = R(t), & R(t) \in \mathcal{L}(W_1, W_0), \\
U_0(0) = 0.
\end{cases}
\end{equation}

Then the solution $U(t)$ of the Cauchy problem (3.4) satisfies

\begin{equation}
U(t) = U_0(t) + I + \int_0^t Z(t_1)dt_1 + \int_0^t \int_0^{t_1} Z(t_1)Z(t_2)dt_2dt_1 + \ldots
\end{equation}

The convergence of this series can be understood in $L^2$. Indeed, because of assumption (i), the term of this series with $k$ integrals can be estimated by $t^k \sup_t \|Z(t)\|_{L^2 \to L^2}^k / k!$. From this and condition (WP) it also follows that $U(t)$ is a solution of (3.4) from $L^2$.
to $L^2$. Let us now define

$$S_N(t) = I + \int_0^t Z(t_1)dt_1 + \int_0^t \int_0^{t_1} Z(t_1)Z(t_2)dt_2dt_1 + \ldots + \int_0^t \int_0^{t_1} \cdots \int_0^{t_{N-1}} Z(t_1)Z(t_2)\ldots Z(t_N)dt_N \ldots dt_2dt_1.$$ 

Let $V(t) = U(t) - S_N(t)$, it is equal to $U_0(t)$ plus the remainder of the series (3.6). Then we have

$$V'(t) - Z(t)V(t) =$$

(3.7)  $$R(t) - Z(t) \int_0^t \int_0^{t_2} \cdots \int_0^{t_{N-1}} Z(t_2)\ldots Z(t_N)dt_2 \ldots dt_N.$$ 

Choosing $N = l$, and renumbering $t_l$'s, it follows from assumption (iii) of the theorem that the second term is continuous from $W_1$ to $W_0$. Since also $R(t) \in L(W_1,W_0)$, it follows that the right hand side is a continuous linear operator from $W_1$ to $W_0$.

Let $w = u(0) \in W_1$ be the Cauchy data for (3.3). If we denote by $\rho(t)$ the value of the operator in the last line of (3.7) at $w$, we will have $\rho(t) \in W_0$. The value of $V(0)$ is

$$V(0) = U(0) - S_N(0) = 0.$$ 

It follows now that $V(t)w$ solves Cauchy problem (3.2), so it belongs to $W$ by assumption (iv). Since $S_N(t)$ is continuous from $W_1$ to $W$ by assumption (ii), and $V(t)w = U(t)w - S_N(t)w$ is in $W$, we obtain $u(t,\cdot) = U(t)w \in W$.

Moreover, suppose that we also have the estimate

$$\|u(t,\cdot)\|_W \leq C\|\rho(t)\|_{W_0}$$

in (iv). Then we also have

$$\|u(t,\cdot)\|_W \leq \|V(t)w\|_W + \|S_N(t)w\|_W \leq C\|\rho(t)\|_{W_0} + C\|w\|_{W_1} \leq C(\|w\|_{W_1} + \|R(t)w\|_{W_0}).$$

□

Later we will need this in the case of $Z(t)$ being Fourier integral operators of order zero. However, let us point out several applications to other cases of pseudodifferential and Fourier integral operators. In these cases we will make different appropriate choices of spaces $W_0, W_1, W$. Moreover, in Corollaries 3.2 and 3.3 we will assume condition (WP) so that the corresponding non-homogeneous Cauchy problems with zero Cauchy data have unique solutions. This is often a natural assumption if $Z(t)$ behave sufficiently well with respect to $t$ since we are working in subspaces of $L^2$.

**Corollary 3.2.** Let $1 < p < \infty$. Let $Z(t) \in \Psi^0, t \in [0,T]$, be a family of pseudodifferential operators of order zero with amplitudes compactly supported in $x,y$, uniformly in $t$. Suppose that there exists $C > 0$ such that

(3.8)  $$\|Z(t)\|_{L^2 \rightarrow L^2} \leq C, \|Z(t)\|_{L^p \rightarrow L^p} \leq C, \forall t \in [0,T].$$
Then the solution \( u = u(t, x) \) of the Cauchy problem

\[
\begin{aligned}
\left\{ \begin{array}{l}
u' - Z(t)u = r(t), \quad r(t) \in L^p, \ t > 0, \\
u(0) \in L^p \cap L^2_{\text{comp}},
\end{array} \right.
\end{aligned}
\]

satisfies \( u(t, \cdot), u'(t, \cdot) \in L^p \), for all \( t \in (0, T] \). Moreover, for each \( t \in [0, T] \) we have the estimate

\[
\|u(t, \cdot)\|_{L^p} \leq C\|u(0)\|_{L^p} + \|r(t)\|_{L^p}.
\]

If amplitudes of \( Z(t) \) are not compactly supported with respect to \( x, y \), we have a similar local statement for compactly supported Cauchy data. Note also that pseudo-differential operators of order zero are locally bounded in a similar local statement for compactly supported Cauchy data. Note also that pseudo-differential operators of order zero and our assumptions. Property (iii) conditions of Theorem 3.1. Properties (i) and (ii) follow from regularity properties Duhamel’s principle and is similar to the one in Corollary 3.3. The Picard series is convergent in \( L^p \) provided that operators norms \( \|Z(t)\|_{L^p \to L^p} \) are uniformly bounded for \( t \in [0, T] \). The estimate for the norm follows from this as well.

We can see that \( u'(t, \cdot) \in W \) from \( u' = Z(t)u + r \) and from the continuity of \( Z(t) \) in \( W \). □

We will now apply Theorem 3.1 in the case of \( Z(t) \) being Fourier integral operators. While our case (1.3) corresponds to \( Z(t) \) being operators of order zero, the crucial smoothing property (iii) will follow from the fact that operators \( Z(t) \) have a special structure. For general Fourier integral operators \( Z(t) \) without structure, we need to assume that they are of negative orders. This is for example the case when the zero order term \( B \) in Theorem 2.2 is actually a pseudo-differential operator of some negative order.

**Corollary 3.3.** Let \( 1 < p < \infty, \epsilon > 0, \) and \( \alpha = (n-1)(1/p - 1/2) \). Let \( Z(t) \in \mathcal{I}^{-\epsilon}, \ t \in [0, T] \), be a family of non-degenerate Fourier integral operator of order \( -\epsilon \) with amplitudes compactly supported in \( x, y, \) uniformly in \( t \). Suppose that the composition of two operators \( Z(t) \) and \( Z(s) \) is again a non-degenerate Fourier integral operator, for any \( t, s \in [0, T] \). Suppose also that

\[
\|Z(t)\|_{H^s \to H^s} \leq C \quad \text{for all} \quad t \in [0, T],
\]

for some \( s > (pn - 2n)/2p \) when \( p > 2 \) and \( s = 0 \) when \( p \leq 2 \). Then the solution \( u = u(t, x) \) of the Cauchy problem

\[
\begin{aligned}
\left\{ \begin{array}{l}
u' - Z(t)u = r(t), \quad r(t) \in H^s, \ t > 0, \\
u(0) \in (L^p_\alpha) \cap L^2_{\text{comp}},
\end{array} \right.
\end{aligned}
\]

satisfies \( u(t, \cdot) \in L^p, \) for all \( t \in (0, T] \). Moreover, for all \( t \in [0, T] \) we have the estimate

\[
\|u(t, \cdot)\|_{L^p} \leq C\|u(0)\|_{L^p_\alpha} + \|r(t)\|_{H^s}.
\]

Note that operators \( Z(t) \) are locally bounded in \( H^s \), so assumption (3.10) simply means that their operator norms are uniformly bounded on \( [0, T] \).
Proof of Corollary 3.3. Let \( W = L^p \cap L^2_{\text{comp}} \), \( W_1 = L^p_\alpha \cap L^2_{\text{comp}} \), and \( W_0 = H^s_{\text{comp}} \subset W \). Let us check conditions of Theorem 3.1. Condition (i) follows from the fact that non-degenerate Fourier integral operators of order 0 are bounded from \( (L^p_\alpha)\text{comp} \) to \( L^1_{\text{loc}} \). Condition (ii) follows from the calculus of non-degenerate Fourier integral operators, since we assumed that compositions of \( Z(t) \) are again non-degenerate Fourier integral operators. Smoothing condition (iii) for large \( l \) follows again from the calculus, since operators \( Z(t) \) are of order \( -\epsilon \).

Finally, let us show that solutions of \( \nu' = Z(t)v = r(t) \), \( r(t) \in H^s_{\text{comp}} \), with zero Cauchy data \( v(0) = 0 \), satisfy \( v(t, \cdot) \in L^2 \cap L^p \). In fact, we will show that \( v(t, \cdot) \in H^s \subset L^p \cap L^2 \).

From the uniqueness of the solution of this problem it follows that we can use Duhamel’s principle to write

\[
(3.12) \quad v(t, x) = \int_0^t E(t, s)r(s, x)ds,
\]

where \( E(t, s) \) is the propagator of

\[
\begin{cases}
(\partial_t - Z(t))E(t, s) = 0, \\
E(t, s)|_{t=s} = I.
\end{cases}
\]

The Picard series for this problem gives the asymptotic expansion of \( E(t, s) \), in particular implying that \( E(t, s) \) is bounded in \( L^2 \) and in \( H^s \) provided operator norms \( \|Z(t)\|_{H^s \rightarrow H^s} \) are uniformly bounded for \( t \in [0, T] \). From (3.12) it follows that \( v(t, \cdot) \in H^s \). Moreover, since \( r(s, \cdot) \in H^s \), we also get an estimate

\[
\|v(t, \cdot)\|_{H^s} \leq C \sup_{\tau \in [0, T]} \|r(\tau, \cdot)\|_{H^s},
\]

implying the estimate in Corollary 3.3. \( \square \)

Proof of Theorem 2.2. As we have already mentioned, by Lemma 5.1 we can assume that characteristics of \( A \) are correctly defined on \( T^*X \). As explained in the introduction, this means that we can assume without loss of generality that \( A \) is smoothly diagonalizable, and thus we can write

\[
(3.13) \quad P(t, x, D_x) = \bigoplus a_j(x, D_x) + B(t, x, D_x),
\]

\[
B(t, x, D_x) = (B_{jk}(t, x, D_x))_{1 \leq j, k \leq m}, \quad B_{jk} \in C^\infty([0, T], \Psi^0).
\]

Some of \( a_j \)'s may be identically equal to each other. We can renumber \( a_j \)'s into \( r \) groups (possibly of size one) of equal characteristics. These are the eigenvalues of the matrix \( A(x, \xi) \) counted with multiplicity. Thus, we have \( 1 = k_1 < k_2 < \ldots < k_r = n + 1 \), and \( a_{k_1} \equiv \ldots \equiv a_{k_{i+1}-1} \neq a_k \), for \( k < k_i \) or \( k \geq k_{i+1} \). This means that we have a group of the same roots \( a_1, \ldots, a_{k_2-1} \), etc., while roots from different groups are not identically the same. Therefore, this is a decomposition of the first order principal part into a block-diagonal form with the same roots in each block, with possible equality of roots in different blocks at some points. So we can write

\[
(3.14) \quad P(t, x, D_x) = \text{diag}(\tilde{a}_1, \ldots, \tilde{a}_r) + B(t, x, D_x),
\]

where
where \( \tilde{a}_i = \text{diag}(a_{k_i}, \ldots, a_{k_{i+1}-1}) \) are diagonal matrices with equal roots at the diagonal. Let us set
\[
\tilde{A}_i = \tilde{a}_i + (B_{\mu \nu})_{k_i \leq \mu, \nu \leq k_{i+1}-1},
\]
so that
\[
P = \tilde{A} + B = \text{diag} (\tilde{A}_1, \ldots, \tilde{A}_r) + B.
\]
Note that in the last equality we can make
\[
B_{\mu \nu} = 0 \quad \text{for} \quad k_i \leq \mu, \nu \leq k_{i+1} - 1
\]
if we add these components to the corresponding components of \( \tilde{A} \). Let \( U(t) = \exp(-i \tilde{A} t) V(t) \). This is well defined in view of, for example, [29, VIII]. If \( r = 1 \) (i.e. in the constant multiplicities case) this implies the statement of the theorem in view of \( L^p \) estimates for Fourier integral operators. So we may assume \( r \geq 2 \) and will use Theorem 2.1. Then
\[
V' = Z(t)V = -ie^{iA t} Be^{-iA t} V, \quad V(0) = I.
\]
Now we will apply Theorem 3.1 with \( Z(t) = -ie^{iA t} Be^{-iA t} \). Let us choose \( W = L^p \cap L^2_{\text{comp}}, W_1 = L^p_a \cap L^2_{\text{comp}}, \) and \( W_0 = H^s_{\text{comp}} \) with \( s > (pm-2n)/2p \) for \( p > 2 \) and \( s = 0 \) for \( 1 < p \leq 2 \). Conditions (i) and (ii) follow from the calculus and regularity properties of non-degenerate Fourier integral operators of order zero. Smoothing condition (iii) follows from Theorem 2.1. For condition (iv) we can use Duhamel’s principle similar to the proof of Corollary 3.3. Thus, Theorem 3.1 implies that propagator \( V(t) \) is locally continuous from \( L^p_0 \) to \( L^p \). Operator \( U(t) \) is a composition of \( V(t) \) with a non-degenerate Fourier integral operator \( \exp(-i \tilde{A} t) \), so \( U(t) \) is given as a sum of a series obtained by the multiplication of the Picard series for \( V(t) \) with \( \exp(-i \tilde{A} t) \). Using the calculus of Fourier integral operators in each term of the series and its smoothing property we can repeat the argument of Theorem 3.1 in this case to see that \( u(t, \cdot) \in L^p_{\text{loc}} \) with estimate (2.4) for its norm. \( \square \)

Note that if \( B \) in Theorem 2.2 is a pseudo-differential operator of negative order, \( B \in \Psi^\mu \), for some \( \mu < 0 \), the proof is much simpler. In this case operator \( Z(t) \) in (3.16) is a Fourier integral operator of negative order \( \mu \), so the statement of Theorem 2.2 follows immediately from Corollary 3.3.

**Proof of Theorem 2.3.** Let \( L \) be as in (2.5) and let \( u \) be the solution of
\[
Lu = f, \quad D_j^l u(0, x) = g_j(x), \quad 0 \leq j \leq m - 1.
\]
Let
\[
U = (u, \partial_1 u, \partial_2 u, \ldots, \partial_1 \partial_2 u, \partial_2 \partial_1 u, \ldots, \partial^J u, \ldots)^T,
\]
where \( J = \{j_1, \ldots, j_k\} \) is a permutation of some elements of \( \{1, \ldots, m\} \), \( |J| = k \leq m - 1 \). Vector \( U \) has \( 1 + \sum_{j=1}^{m-1} m!/j! \) components. Here we can write
\[
\partial^J = D_1^k + \sum_{j=0}^{k-1} c_j(t, x, D_x) D_1^j,
\]
where \( c_j(t) \in \Psi^{k-j} \). We set \( |J| = k \). It was shown by Morimoto in [22] that \( U \) solves the system
\[
D_i U + AU + BU = F, \quad U(0, x) = G(x),
\]
where
\[ F = (0, \ldots, 0, f, \ldots, f) \]
and
\[ G = (g_0, \ldots, g_J) + \sum_{j=0}^{J-1} c_j^f g_j, \ldots, \]

A is a diagonal matrix with \( \lambda_j \)'s at the diagonal and \( B \) is a matrix of pseudo-differential operators of order zero. Matrix \( B \) has some operators in the last row, zeros, and \(-1\) above the diagonal. If \( \lambda_j \) satisfy Condition C, Theorem 2.2 implies
\[ \|U(t, \cdot)\|_{L^p_{loc}} \leq C\|G\|_{L^p}. \]

Since \( c_j^f(t) \in \Psi^{k-j} \), we get
\[ \|G\|_{L^p} \leq C \sum_{j=0}^{m-1} \|g_j\|_{L^p_{\alpha}}, \]

which implies the estimate of the Theorem.

\[ \square \]

4. ANALYSIS OF THE PICARD SERIES

In this Section we will prove Theorem 2.1 on the smoothing properties of terms \( Q_l \) of the Picard series (1.4). In the notation associated to groups of symbols as in the beginning of the proof of Theorem 2.2, we have \( r \) operators \( \tilde{A}_i \) with principal symbols \( \tilde{a}_i \). By definition these symbols are not identically equal to each other. To simplify the notation in this section, we drop the tilde, so that we let \( A_j \in \Psi^1, j = 1, \ldots, r \), be pseudo-differential operators of order one with principal symbols \( a_j(x, \xi) \). In assumptions of Theorem 2.1 we can assume that \( r \geq 2 \) and that there are no identical symbols among these \( a_j \)'s and that they satisfy Condition C. Let now
\[ H(t) = e^{i A_j t_1} e^{i A_j (t_2 - t_1)} \cdots e^{i A_{j+1} t_l}, \]
where \( 1 \leq j_k \leq r, j_k \neq j_{k+1}, k = 1, \ldots, l + 1 \), and \( \bar{t} = (t_1, \ldots, t_l) \). Let us define
\[ Q = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{l-1}} B(\bar{t}) H(\bar{t}) dt_l \cdots dt_1, \]
where \( B(\bar{t}) \in \Psi^0 \) is a pseudo-differential operator of order zero smoothly dependent on \( t \). Such operators \( Q \) appear in the Picard series (1.4), (1.5). In this section we will give a detailed description of operator \( Q \) in order to prove that it has the required regularising properties when \( l \) is sufficiently large. First of all let us note that \( H \) is a Fourier integral operator and due to the theorem on compositions of Fourier integral operators the canonical relation \( \Lambda^f \subset T^*X \times T^*X \) of \( H(\bar{t}) \) is given by
\[ \Lambda^f = \{(x, p, y, \xi) : (x, p) = \Phi^f(y, \xi)\}, \]
where
\[ \Psi^f = \Phi^f_{j_1} \circ \cdots \Phi^f_{j_l} \circ \Phi^f_{j_{l+1}} \]
and \( \Phi^f_{j_i} \) is the Hamiltonian flow defined by \( a_j \).

It can be easily checked that \( H \) is a solution operator for the system of equations
\[ (4.1) \quad \frac{\partial H}{\partial t_k} = iT_k(t_1, \ldots, t_k) H, \quad k = 1, \ldots, l, \]
where $T_k \in \Psi^1$ is a pseudo-differential operator of order one. In view of Egorov's theorem its principle symbol is equal to

$$T_k^0(t_1, \ldots, t_k, x, \xi) = (a_{j_k} - a_{j_{k+1}}) \circ \Phi_{j_k}^t \circ \cdots \circ \Phi_{j_1}^t - t_k + 1(x, \xi), \quad (x, \xi) \in T^*X$$

for all $k = 1, \ldots, l$. Let us now investigate the smoothing properties of operator $Q$. We will look for it in the form

$$\varphi(\bar{t}, x, y, \xi) = \psi(\bar{t}, x, \xi) - y \cdot \xi.$$

It follows from (4.1) and (4.2) that $\psi$ satisfies a system of Hamilton-Jacobi equations

$$\frac{\partial \psi}{\partial \tau} = T_k^0(\bar{t}, x, \frac{\partial \psi}{\partial x}), \quad \psi(0, x, \xi) = x \cdot \xi.$$  

In [24] it was checked that Frobenius conditions for system (4.3) are satisfied. Solving this system we obtain a non-degenerate phase function. This phase function defines a Lagrangian manifold $\Lambda$, so that we have

$$\left(x, \frac{\partial \psi}{\partial x}\right) = \Psi^*(y, \xi) = (x^\dagger(y, \xi), p^\dagger(y, \xi)), \quad y = \frac{\partial \psi}{\partial \xi}.$$  

Now we are going to investigate the smoothing properties of operator $Q$. We can write $Q$ as

$$Qu(x) = \int_\Delta \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\varphi(\bar{t}, x, y, \xi)} b(\bar{t}, x, y, \xi) u(y) dy d\xi d\bar{t},$$

where $\varphi$ satisfies (4.4) and $b$ is an amplitude of order zero, which we may assume to be compactly supported with respect to $x$ and $y$. Here $\Delta$ is the simplex $\{0 \leq t_1 \leq t_{l-1} \leq \ldots \leq t_1 \leq l\}$.

Let $\chi \in C^\infty_0$ be a cut-off function such that $\chi(\tau) = 1$ for $|\tau| < 1$ and $\chi(\tau) = 0$ for $|\tau| > 2$. Operator $Q$ can be decomposed as $Q = R_1 + R_2 + R_3$, where

$$R_j u(x) = \int_\Delta \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\varphi(\bar{t}, x, y, \xi)} \mu_j(\epsilon, \bar{t}, x, y, \xi) u(y) dy d\xi d\bar{t}, \quad j = 1, 2, 3,$$

where

$$\mu_1(\epsilon, \bar{t}, x, y, \xi) = \left(1 - \chi(\epsilon^{-1} \left| \frac{\partial \varphi}{\partial \xi} \right|)\right) b(\bar{t}, x, y, \xi),$$

$$\mu_2(\epsilon, \bar{t}, x, y, \xi) = \chi(\epsilon^{-1} \left| \frac{\partial \varphi}{\partial \xi} \right|) \left(1 - \chi(\epsilon^{-1} \left| \frac{\partial \varphi}{\partial \xi} \right|) \left| \xi^{-1} \right|\right) b(\bar{t}, x, y, \xi),$$

$$\mu_3(\epsilon, \bar{t}, x, y, \xi) = \chi(\epsilon^{-1} \left| \frac{\partial \varphi}{\partial \xi} \right|) \chi(\epsilon^{-1} \left| \frac{\partial \varphi}{\partial \xi} \right|) \left| \xi^{-1} \right| b(\bar{t}, x, y, \xi).$$

Let us first consider operator $R_1$. On the support of $\mu_1$ we have the estimate

$$\left| \frac{\partial \varphi}{\partial \xi} \right| \geq \epsilon.$$

Therefore, there exists an operator $L(e^{i\varphi})$ of order 1, such that $Le^{i\varphi} = e^{i\varphi}$, with coefficients estimated by $C\epsilon^{-1}$ on the support of the amplitude of $\mu_1$. When integrating by parts with $L$ there may appear an additional factor of $\epsilon^{-1}$ when differentiating $\chi$. Integrating by parts $p$ times with this operator $L$ we obtain an operator with an
amplitude of order \(-p\) and coefficients that can be estimated by \(\epsilon^{-2p}\). From Lemma 5.2 with \(q = -p\) we obtain the following estimate

\[
(4.5) \quad \| R_1 \|_{p-n-1} \leq C\epsilon^{-3p+n+1}.
\]

The same procedure with integrating by parts with respect to \(\xi\) can not be applied to \(R_2\). But here there is a possibility to integrate by parts with respect to \(\bar{t}\). Indeed, there exists an operator \(M (\partial / \partial \bar{t})\), such that \(Me^{i\varphi} = e^{i\varphi}\), with coefficients not greater than \(\epsilon^{-1}|\xi|^{-1}\) on the support of \(\mu_2\). Integrating by parts with \(M\) we obtain an operator with an amplitude of order \(-1\) with coefficients that can be estimated by \(\epsilon^{-2}\), where another \(\epsilon^{-1}\) may appear from differentiating \(\chi\). The boundary integrals have the same form as \(Q\) but they have amplitudes of order \(-1\) and depend on not less than \(l - 2\) time variables. The reason for possibly losing two variables is that after restriction to the boundary, say \(t_2 = t_1\), it may happen that \(a_{j_1}\) and \(a_{j_3}\) are the same roots. It follows that we can apply such procedure \(\lfloor l/2 \rfloor\) times. As a result we obtain operators of order \(-\lfloor l/2 \rfloor\) with coefficients not greater than \(C\epsilon^{-2\lfloor l/2 \rfloor}\). Then in view of Lemma 5.2 for \(\lfloor l/2 \rfloor > n + 1\) we have

\[
(4.6) \quad \| R_2 \|_{\lfloor l/2 \rfloor-n-1} \leq C\epsilon^{-3\lfloor l/2 \rfloor+n+1}.
\]

Let us now consider the last integral \(R_3\). It is not possible to apply procedures with integrating by parts as above either with respect to \(\xi\) or with respect to \(\bar{t}\). But in this case it turns out that the support of amplitude \(\mu_3\) is small. The singular support of the integral kernel of \(Q\) may be very irregular in this case, so a more delicate analysis is necessary to show the smoothing properties of \(R_3\). First we will show that

\[
(4.7) \quad | T_j^0 (\bar{t}, x, \partial_\psi / \partial x) | \leq C\epsilon |\xi|, j = 1, \ldots, l,
\]

on the support of \(\mu_3\). We notice that \(\mu_3\) differs from 0 only if

\[
(4.8) \quad \left| \frac{\partial \varphi}{\partial \xi} \right| \leq 2\epsilon, \quad \left| \frac{\partial \varphi}{\partial \bar{t}} \right| \leq 2\epsilon|\xi|.
\]

It follows from (4.3) and (4.8) that

\[
(4.9) \quad \left| T_j^0 (\bar{t}, x, \partial_\psi / \partial x) \right| \leq C\epsilon |\xi|, j = 1, \ldots, l.
\]

Because of the homogeneity of \(T_j^0\) with respect to \(\xi\) we also have the following trivial estimates

\[
\left| \partial_x T_j^0 (\bar{t}, x, \partial_\psi / \partial x) \right| \leq C|\xi|, \quad \left| \partial_{\bar{t}} T_j^0 (\bar{t}, x, \partial_\psi / \partial x) \right| \leq C, \quad j = 1, \ldots, l.
\]

Consequently, we obtain

\[
\left| T_j^0 (\bar{t}, x, \partial_\psi / \partial x) - T_j^0 (\bar{t}, x, x^f(y, \xi), p^f(y, \xi)) \right| \leq C \left( |\xi||x - x^f(y, \xi)| + \left| \frac{\partial \psi}{\partial x} - p^f(y, \xi) \right| \right), \quad j = 1, \ldots, l.
\]
Since $\frac{\partial \varphi}{\partial \xi}(\bar{t}, x, y, \xi, y, \xi) = 0$, we get
\[
\frac{\partial \varphi}{\partial \xi}(\bar{t}, x, y, \xi) = \frac{\partial \varphi}{\partial \xi}(\bar{t}, x, y, \xi) - \frac{\partial \varphi}{\partial \xi}(\bar{t}, x^\prime(y, \xi), y, \xi)
= \frac{\partial^2 \varphi}{\partial x \partial \xi}(\bar{t}, x^*, y, \xi)(x - x^\prime(y, \xi)),
\]
for some points $x^*$. By this notation we mean a more careful argument here with an alternative representation of the mean value theorem in the integral form. Now, since $\det \frac{\partial^2 \varphi}{\partial x \partial \xi} \neq 0$ for small $\bar{t}$, we obtain
\[
|x - x^\prime(y, \xi)| \leq C \left| \frac{\partial \varphi}{\partial \xi}(\bar{t}, x, y, \xi) \right|.
\]
From the properties of Lagrangian manifold $\Lambda^\bar{t}$ in (4.4) we also obtain
\[
|\frac{\partial \psi}{\partial x}(\bar{t}, x, \xi) - p^\bar{t}(y, \xi)| = \left| p^\bar{t}(\frac{\partial \psi}{\partial \xi}, \xi) - p^\bar{t}(y, \xi) \right| 
\leq C \left| \frac{\partial \varphi}{\partial \xi}(\bar{t}, x, y, \xi) \right| |\xi|, \ \forall (x, y, \xi).
\]
Finally, taking into consideration (4.8)-(4.11) we obtain (4.7) on the support of $\mu_3$.

Estimate (4.7) shows us that the support of the amplitude $\mu_3$ of operator $R_3$ is contained in
\[
\Xi_1 = \left\{ (\bar{t}, x, y, \xi) : x \in X, \xi \neq 0, (\bar{t}, y, \xi) \in \Xi \right\},
\]
where
\[
\Xi = \left\{ (\bar{t}, y, \xi) : \left| T^0_j(\bar{t}, y, \xi) \right|, c \leq |\xi| \leq C, j = 1, \ldots, l \right\}.
\]
That allows us to apply estimate (5.4) of Lemma 5.2 with $q = 0$ and $\delta = 1$ to the operator $R_3$ to get
\[
\|R_3\|_{-3n/2-2} \leq C \epsilon^{-n-1} \text{meas}(\Xi),
\]
where meas is the natural measure on $[0,T]^l \times S^*X$. To use this inequality we need to estimate the measure $\text{meas}(\Xi)$. The set $\Xi$ maps to
\[
\Xi_2 = \left\{ (\bar{t}, y, \xi) : |T^0_j(\bar{t}, y, \xi)| \leq C \epsilon, |p^\bar{t}(y, \xi)| = 1, j = 1, \ldots, l \right\}
\]
under the Hamiltonian flow $\Psi^\bar{t}$ which preserve measure. The measure of $\Xi_2$ may be estimated by the measure (in $[0,T]^l \times X \times \mathbb{R}^n$) of
\[
\Xi_3 = \left\{ (\bar{t}, y, \xi) : |T^0_j(\bar{t}, y, \xi)| \leq C \epsilon, c \leq |\xi| \leq C, j = 1, \ldots, l \right\}.
\]
Let us now introduce sets
\[
\Sigma_j(t_1, \ldots, t_{j-1}, y, \xi, c) = \{ t_j : |T^0_j(\bar{t}, y, \xi)| \leq C \epsilon, c \leq |\xi| \leq C \},
\]
for $j = 1, \ldots, l$. To estimate measures of these sets we will use Proposition 5.3. For this, in the notation of Proposition 5.3, we set $w = (y, \xi)$ and consider the function
\[
t \mapsto f(t, w) = T^0_j(t_1, \ldots, t_{j-1}, t, t_{j+1}, \ldots, t_l, y, \xi),
\]
where $t$ takes the position of $t_j$ in $\bar{t}$. 

21
Differentiating equation (4.2) with respect to $t_j$ and using the Hamilton-Jacobi equations for the flows, we get that derivatives of $f$ with respect to $t$ equal to the iterated Poisson brackets of the symbols. Hence Condition $C$ on zeros of Poisson brackets implies that function $f$ may have zeros in $t$ of order not greater than $M$. It follows now from Proposition 5.3 that

\begin{equation}
\sup_{y,\xi} \text{meas}\{t_j \in [0, T] : T^0_j(\bar{t}, y, \xi) \leq C\varepsilon\} \leq C\varepsilon^{1/2M},
\end{equation}

where meas is just the Lebesgue measure, $y$ varies over a compact set and $c \leq |\xi| \leq C$. Now we can use a simple observation that if we have two functions $f_1(t_1)$ and $f_2(t_1, t_2)$ such that

\[ \text{meas}\{t_1 : |f_1(t_1)| \leq \varepsilon, \sup_{t_1} \text{meas}\{t_2 : |f_2(t_1, t_2)| \leq \varepsilon\} \leq C\varepsilon^\alpha, \]

then

\[ \text{meas}\{(t_1, t_2) : |f_1(t_1)| \leq \varepsilon, |f_2(t_1, t_2)| \leq \varepsilon\} \]

\[ = \int_{|f_1(t_1)| \leq \varepsilon, |f_2(t_1, t_2)| \leq \varepsilon} dt_1 dt_2 \]

\[ = \int_{|f_1(t_1)| \leq \varepsilon} \left( \int_{|f_2(t_1, t_2)| \leq \varepsilon} dt_2 \right) dt_1 \]

\[ \leq C\varepsilon^{2\alpha}. \]

Applying this argument $l$ times to the estimate (4.13), we get

\[ \text{meas}(\Xi_3) \leq C \text{vol} (\text{supp}_y b) \sup_{(y, \xi)} \prod_{j=1}^l \sup_{t_1, \ldots, t_{j-1}} \text{meas}(\Sigma_j(\bar{t}, y, \xi, \varepsilon)), \]

which implies

\[ \text{meas}(\Xi) \leq C\varepsilon^{l/2M}. \]

Here we used that the support of the amplitude $b(\bar{t}, x, y, \xi)$ is compact with respect to $x$ and $y$, and that $T$ is small. Finally, combining this with estimate (4.12), we get

\begin{equation}
\|R_3\|_{-3n/2-2} \leq C\varepsilon^{l/2M-n-1}. \end{equation}

An application of Lemma 5.6 to (4.5), (4.6) and (4.14) yields formula (2.1) for the Sobolev index $N(l)$. Formula (2.2) then implies that $N(l) \to \infty$ as $l \to \infty$, which proves Theorem 2.1.

5. Various auxiliary results

First we describe the lifting of the problem to insure that characteristic roots of the principal symbol of $P$ are globally uniquely defined. This will mean that the microlocal diagonalizability of matrices implies a local diagonalizability on a suitable covering space. Since the dimensions are the same we are able to reduce all the results back to the original space. The proof of Lemma 5.1 is similar to the proof of Rozenblum [24], who considered the self-adjoint version of this result.

We note that the following lemma is also the reason for us to restrict the considerations to manifolds of dimension $n \geq 3$. For $n < 3$ we may either assume that
the principal part of $P$ is locally smoothly diagonalizable or make all the arguments microlocal.

**Lemma 5.1.** Let $X$ be a smooth compact manifold without boundary of dimension $n \geq 3$. Let a pseudo-differential operator $P(x, D)$ of order one act on sections of an $m$-dimensional vector bundle $E$. Let $L^2(E)$ be the space of sections of half-densities on $E$. Let $E'$ be the lifting of $E$ to $T^*X$. Let us assume that the principal symbol $A(x, \xi)$ is microlocally smoothly diagonalizable, i.e. microlocally in $\Lambda \subset T^*X$ such that $E'|_{\Lambda} \cong \Lambda \times \mathbb{C}^m$, principal symbol $A(x, \xi)|_{\Lambda}$ has $m$ smooth real eigenvalues $a_j(x, \xi)$ and one dimensional eigenspaces $V_j(x, \xi)$, and such diagonalizations are compatible in intersecting cones.

Then there exists a cover $\tilde{X}$ of $X$ with finitely many sheets, such that on the lifting $\tilde{E}$ of $E$ to $T^*\tilde{X}$ branches of eigenvalues $a_j(x, \xi)$ and eigenspaces $V_j(x, \xi)$ of the principal symbol $A(x, \xi)$ are smooth and globally well defined. The space $L^2(\tilde{E})$ has a decomposition into a direct sum of $m$ spaces such that the matrix representation of $P(\tilde{x}, \tilde{\xi})$ with respect to this decomposition consists of pseudo-differential operators, and its principal symbol is a diagonal matrix with $a_j(\tilde{x}, \tilde{\xi})$ at the diagonal.

**Proof.** Let us fix $z_0 = (x_0, \xi_0) \in S^*X$. For each path in $S^*X$ beginning at $z_0$ we look at the continuation of the diagonalization along this path. At each point $z \in S^*X$ we obtain up to permutation several possible collections $a_j(z), V_j(z)$, such that homotopic paths from $z_0$ to $z$ correspond to the same collection. Since for $n \geq 3$ homotopy of paths in $S^*X$ and their projections to $X$ is equivalent, we obtain a homomorphism $\pi_1(\tilde{X})$ to the permutation group of order $m$. A construction in the homotopy theory (see [23]) gives a cover $p : \tilde{X} \to X$ with finitely many sheets, such that the lifting of this homomorphism to $\pi_1(\tilde{X})$ is trivial. So, each closed path in $S^*\tilde{X}$ takes eigenvalues and, therefore, also eigenspaces, to themselves. This means that eigenvalues and eigenspaces have global smooth branches on $S^*\tilde{X}$ and hence also on $T^*\tilde{X}$.

Let $p_j^0(\tilde{x}, \tilde{\xi})$ be a family of orthogonal projectors on $V_j(\tilde{x}, \tilde{\xi})$. By the standard Gram-Schmidt process, we can add lower order terms to $p_j^0(\tilde{x}, \tilde{\xi})$ to obtain symbols $p_j(\tilde{x}, \tilde{\xi})$, for which $p_j(\tilde{x}, \tilde{D})p_k(\tilde{x}, \tilde{D}) = \delta_{jk}p_j(\tilde{x}, \tilde{D})$ and $\sum p_j(\tilde{x}, \tilde{D}) = 1$. For $u \in L^2(E)$, let $u_j = p_j(\tilde{x}, \tilde{D})u$. Since

$$P(\tilde{x}, \tilde{D})p_j(\tilde{x}, \tilde{D}) = p_j(\tilde{x}, \tilde{D})P(\tilde{x}, \tilde{D})p_j(\tilde{x}, \tilde{D})$$

modulo lower order terms, $P = \sum p_jPp_k$ yields the desired diagonalization. \hfill \Box

Let $a_j$ still denote the smooth global branches of characteristics of $A$ lifted to $\tilde{X}$. Let $\tilde{U}(t)$ be the fundamental solution to (1.1) with operator $P$ lifted to $\tilde{X}$. Let $\tilde{U}(t, \tilde{x}, \tilde{y})$ be the integral kernel of $\tilde{U}(t)$, $\tilde{x}, \tilde{y} \in \tilde{X}$. Let

$$U(t, x, y) = \sum_{p\tilde{x} = y} \tilde{U}(t, \tilde{x}, \tilde{y}), \hspace{1em} x, y \in X,$$

(5.1) where $\tilde{x}$ is any point of $\tilde{X}$ such that $p\tilde{x} = x$ and the summation is carried out over all preimages $\tilde{y}$ of $y$. Since (5.1) is invariant under permutations of the sheets of the cover $\tilde{X}$, the kernel $U(t, x, y)$ is independent of the choice of $\tilde{x}$. The equation and
Cauchy data are automatically satisfied, so (5.1) gives a fundamental solution to the system (1.1) on \( X \).

The following Lemma gives some estimates for the operator norm from \( L^2 \) to \( H^s \) for Fourier integral operators in terms of \( L^\infty \) norms of the amplitude and its derivatives. Recall that by \( ||T||_s \) we denote the operator norm of \( T \) from \( L^2 \) to \( H^s \).

**Lemma 5.2.** Let \( T \) be a Fourier integral operator

\[
(Tu)(x) = \int_{\mathbb{R}^n} \int_{X} e^{i\varphi(x,y,\xi)} a(x,y,\xi) u(y) dy d\xi,
\]

where \( X \) is an open set in \( \mathbb{R}^n \). Assume that \( a \in S^q, q \in \mathbb{Z} \), is an amplitude of order \( q \) and has compact support with respect to \( x \) and \( y \) in \( X \). Assume also that \( \partial \varphi / \partial x \neq 0 \) for \( \xi \neq 0 \). Then \( T \) extends to a bounded operator from \( L^2(X) \) to \( H^{-q-n-1}(X) \) with

\[
||T||_{-q-n-1} \leq C ||a(\xi)^{-q}||_{C^{[q+n+1]}};
\]

where \( \langle \xi \rangle = (1 + |\xi|^2)^{1/2} \).

Moreover, let us assume in addition that the support of \( a \) belongs to a conical set with respect to \( \xi \) which does not depend on \( x \), i.e.

\[
\text{supp } a \subset \Xi = \left\{ (x, y, \xi) : x \in X, (y, \frac{\xi}{|\xi|}) \in \Xi, \xi \neq 0 \right\},
\]

where \( \Xi \) is subset of \( \mathbb{R}^n \). Then for any \( \delta > 0 \) we have

\[
||T||_{-q-3n/2-1-\delta} \leq C \text{meas}(\Xi) ||a(\xi)^{-q}||_{C^{[q+n+1]}},
\]

where \( \text{meas} \) is the canonical induced measure on \( \mathbb{R}^n \) and constant \( C \) may depend on the size of the support of \( a \) with respect to \( x \) and \( y \).

**Proof.** First we consider the case \( q + n + 1 < 0 \). Differentiating (5.2) \(-(q + n + 1)\) times with respect to \( x \) we obtain the integral with amplitude of order \( -(n + 1) \). The integral with respect to \( \xi \) converges absolutely, so \( T \) extends as a bounded operator from \( L^2(X) \) to \( C^{-(q+n+1)} \) with estimate (5.3). Estimate (5.4) clearly follows from this as well.

We will now consider the case \( q + n + 1 \geq 0 \). For a smooth function \( v \), let us consider the bilinear form \( (Tu, v) \). Let us define operator \( L \) as the transpose of

\[
^t L = (1 + |\partial_x \varphi|^2)^{-1} (1 - i \partial_x \varphi \cdot \partial_x).
\]

Since \( \partial \varphi / \partial x \neq 0 \), integrating by parts \( q + n + 1 \) times with operator \( L \), we obtain an absolutely convergent integral with respect to \( \xi \), and the estimate

\[
|(Tu, v)| \leq C ||u||_{L^2} ||v||_{q+n+1} ||a(\xi)^{-q}||_{C^{[q+n+1]}}.
\]

This implies (5.3).

We can slightly modify this argument to obtain an estimate of operator \( T \) acting from \( L^\infty(X) \). Indeed,

\[
|(Tu, v)| \leq \left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\varphi(x,y,\xi)} L^{q+n+1} (a(x,y,\xi) \bar{v}(x)) u(y) dy d\xi dx \right|
\leq C ||v||_{q+n+1} ||u||_{L^\infty} I,
\]

where
where
\[ T^2 = \int_X I_1^2(x)dx, \quad I_1(x) = \int_X \int_{\mathbb{R}^n} |\tilde{a}(x, y, \xi)| d\xi dy, \]
with some amplitude \( \tilde{a} \) of order \(-n - 1\). We can estimate \( \tilde{a} \) by \( (\xi)^{-n-1} \) and from the
embedding theorems it follows that
\[ |(Tu, v)| \leq C\|v\|_{q+n+1}\|u\|_{n/2 + \delta} \text{meas}(\Xi)\|a(\xi)^{-q}\|_{C^{\gamma+n+1}}, \]
implying estimate (5.4) for the norm of operator \( T \) acting on \( L^2(X) \).

The following result shows that if a smooth function on a bounded interval has
zeros only of finite order, then the measure of the set where this function is small is
also small. Moreover, if we have a family of such functions continuously dependent
on a parameter, we can estimate measures of sets where functions are small uniformly
for all parameters varying over compact sets.

**Proposition 5.3.** Let \( W \subset \mathbb{R}^n \) be compact and let \( 0 < T < \infty \). Let a real valued
function \( f = f(t, w) \) be continuous in \( w \in W \) and smooth in \( t \in [0, T] \) up to the
boundary of \([0, T] \). Let \( M \in \mathbb{N} \) and suppose that for each \( w \in W \) function \( f(\cdot, w) \)
has zeros with respect to \( t \) of order not greater than \( M \). Then there exist \( C > 0 \) and \( \epsilon_0 > 0 \) such that for all \( 0 < \epsilon < \epsilon_0 \) we have
\[ \sup_{w \in W} \text{meas}\{t \in [0, T] : |f(t, w)| \leq \epsilon\} \leq C\epsilon^{1/M}. \]

In fact, a more precise estimate is possible:

**Remark 5.1.** Under conditions of Proposition 5.3 for any \( \delta > 0 \) there exist \( C > 0 \) and \( \epsilon_0 > 0 \) such that for all \( 0 < \epsilon < \epsilon_0 \) we have
\[ \sup_{w \in W} \text{meas}\{t \in [0, T] : |f(t, w)| \leq \epsilon\} \leq C\epsilon^{1/M-\delta}. \]

**Proof.** For \( \epsilon > 0 \) let us define
\[ \Sigma(w, \epsilon) = \{t \in [0, T] : |f(t, w)| \leq \epsilon\}. \]
Let \( K(w) \) be the number of zeros of function \( f(\cdot, w) \) with respect to \( t \in [0, T] \) and
let \( K \) be the maximum of \( K(w) \) over \( w \in W \). It is obvious that \( K \) is a finite number
due to the condition on zeroes of \( f \) and compactness of \( W \).

Let \( \alpha_p > 0, p \in \mathbb{N}, \) be a decreasing sequence of positive numbers which we will
choose later. Let us define sets \( \Sigma^p(w, \epsilon) \) by setting
\[ \Sigma^p(w, \epsilon) = \{t \in [0, T] : |f(t, w)| \leq \epsilon, \ldots, \\
|\partial_t^{p-1} f(t, w)| \leq \epsilon^{\alpha_p-1}, |\partial_t^p f(t, w)| \geq \epsilon^{\alpha_p}\}, p \in \mathbb{N}. \]

We claim now that there exists \( \epsilon_0 > 0 \) such that for all \( 0 < \epsilon < \epsilon_0 \), and all \( w \in W \),
we have \( \Sigma^p(w, \epsilon) = \emptyset \) for all \( p > M \). Indeed, if it is not so, then due to compactness
of \( W \) there are sequences \( t_n, w_n \) and \( \epsilon_n \), converging to \( t^* \in [0, T], w^* \in W \) and zero,
respectively, such that
\[ |f(t_n, w_n)| \leq \epsilon_n, \ldots, |\partial_t^{M+1} f(t_n, w_n)| \leq \epsilon_n^{\alpha_{M+1}}, \]
and consequently 
\[ \partial_t^p f(t^*, w^*) = 0, \ p = 0, \ldots, M + 1, \]
which is impossible. It follows now that the set \( \Sigma(w, \epsilon) \) may be presented as the following union of sets
\[
\Sigma(w, \epsilon) = \bigcup_{p=1}^{M} \Sigma^p(w, \epsilon).
\]
The idea of the proof now is to show first that the number of connected components of sets \( \Sigma^p(w, \epsilon) \) is finite and can be estimated uniformly over all \( w \) and \( \epsilon \). Then we will show that the size of each connected component is small and can be estimated by \( \epsilon^{1/2M} \), which will imply Proposition 5.3. These statements are proved in the following two lemma.

**Lemma 5.4.** There exists \( \epsilon_0 > 0 \) such that for all \( 0 < \epsilon < \epsilon_0 \) the inequality
\[
\Delta(\Sigma^p(w, \epsilon)) \leq K(M + 1)^2, \ \forall w \in W, 1 \leq p \leq M,
\]
holds, where \( \Delta(\Sigma^p) \) is the number of connected components of \( \Sigma^p \).

In particular, it follows that the numbers of connected components of \( \Sigma^p(w, \epsilon) \) are uniformly bounded.

**Proof.** For simplicity let us first consider the case \( p = 1 \). Let us assume that (5.7) is not valid. Then there exist sequences \( w_n \) converging to some \( w^* \in W \) and \( \epsilon_n \) converging to 0 such that
\[
\Delta(\Sigma^1(w_n, \epsilon_n)) > K(M + 1)^2.
\]
Let us now choose small enough \( \epsilon_1 > 0 \) such that each connected interval in the closure of \( \Sigma(w^*, \epsilon_1) \) will include one and only one zero of function \( f(t, w^*) \) with respect to \( t \) and will not include zeros of derivative of \( \partial_t f(t, w^*) \) different from zeros of \( f(t, w^*) \). This is possible because if we had an infinite number of zeros of \( \partial_t f(t, w^*) \) approaching a zero \( t^* \) of \( f(t, w^*) \), it would mean that \( \partial_t f(t^*, w^*) = 0 \) and that in fact \( t^* \) is zero of \( \partial^k f(t, w^*) \) for all \( k \geq 1 \), which is impossible since we assumed that all zeros of \( f(t, w^*) \) are of finite order not exceeding \( M \). It follows now that \( \Delta(\Sigma(w^*, \epsilon_1)) \leq K \). Since \( w_n \) converges to \( w^* \), by continuity we also have
\[
\Sigma^1(w_n, \epsilon_n) \subset \Sigma(w^*, \epsilon_1),
\]
for sufficiently large \( n \). From (5.8) and (5.9) it follows that there exist \( (M + 1)^2 + 1 \geq 2M + 1 \) connected components of \( \Sigma^1(w_n, \epsilon_n) \) which are all contained in one of the connected components of closure of \( \Sigma(w^*, \epsilon_1) \). Let us denote this connected component of \( \Sigma(w^*, \epsilon_1) \) by \( I \).

According to the definition of sets \( \Sigma^1(w_n, \epsilon_n) \), between two of its connected components function \( f(t, w_n) \) must become relatively large (i.e. \( > \epsilon \)) or its derivative \( \partial_t f(t, w_n) \) must become relatively small (i.e. \( < \epsilon^{\alpha_1} \)). From this we see that in the first case \( \partial_t f(\cdot, w_n) \) must become zero at some point between these components, while in the second case \( \partial^2_t f(\cdot, w_n) \) must become zero at some point there. Since the number of components of \( \Sigma^1(w_n, \epsilon_n) \) in \( I \) is at least \( 2M + 1 \), it follows that at least one of these two cases occurs at least \( M + 1 \) times.
Let us consider these cases separately. In the first case, for sufficiently large \( n \), we have at least \( M \) zeros \( s^1_n < \ldots < s^n_M \) of the derivative \( \partial_t f(t, w_n) \) contained in \( I \). It follows that \( \partial^2_t f(\cdot, w_n) \) has at least \( M - 1 \) different zeros in \( I \), \( \partial^3_t f(\cdot, w_n) \) has at least \( M - 2 \) different zeros in \( I \), etc. In particular, there are \( \tau^k_n \in [s^1_n, s^n_M] \) such that \( \partial^k_t f(\tau^k_n, w_n) = 0 \) for \( k = 1, \ldots, M \).

Using compactness of \( I \) and continuity of \( f \), it follows that there are subsequences \( s^1_n, \ldots, s^n_M, \tau^k_n \) and \( w_n \) converging to \( s^1_s, \ldots, s^s_s \) and \( \tau^k_s \), respectively, which are all contained in \( I \), and \( w_n \to w^* \). Since functions \( f(t, w) \) are smooth in \( t \) we have

\[
\partial_t f(s^1_s, w^*) = 0, \ldots, \partial_t f(s^s_s, w^*) = 0, \quad \text{and} \quad \partial^k_t f(\tau^k_s, w^*) = 0, \quad k = 1, \ldots, M.
\]

Since there are no zeros of derivative \( \partial_t f(\cdot, w^*) \) on the interval \( I \) except may be a point \( t^* \) which is zero of function \( f(\cdot, w^*) \), we obtain \( s^1_s = \ldots = s^s_s = t^* \). From \( \tau^k_s \in [s^1_s, s^s_s] \) it follows that \( \tau^k_s = t^* \) for all \( k = 1, \ldots, M \), which means

\[
\partial^k_t f(t^*, w^*) = 0 \quad \text{for all} \quad k = 0, \ldots, M.
\]

But this is impossible since function \( f(\cdot, w^*) \) may have zeros only of order \( M \) in view of our assumption.

The second case is slightly different. Here, for sufficiently large \( n \), we have at least \( M \) zeros \( s^1_n < \ldots < s^n_M \) of the second order derivative \( \partial^2_t f(t, w_n) \) contained in \( I \setminus \Sigma^1(w_n, \epsilon_n) \). Because in this case we assumed that \( \Sigma^1(w_n, \epsilon_n) \) breaks into at least \( M + 1 \) components due to the failure of condition \( |\partial_t f(t, w)| \geq \epsilon_n^1 \), it follows that

\[
|\partial_t f(s^i_n, w_n)| \leq \epsilon_n^1, \quad i = 1, \ldots, M.
\]

As \( n \) tends to infinity, we can choose subsequences of \( s^i_n \) converging to some \( s^i_s \in I \). Because \( \partial_t f(\cdot, w^*) \) does not have zeros in \( I \) except may be for some \( t^* \in I \) which is also the unique zero of \( f(\cdot, w^*) \) in \( I \), we get that \( s^1_s = \ldots = s^s_s = t^* \). From (5.10) we also have \( \partial_t f(t^*, w^*) = 0 \).

Now, to deal with higher order derivatives of \( f \) at \( t^* \), similar to the first case, we get a collection of \( \tau^k_s \in [s^1_s, s^s_s] \) such that \( \partial^k_t f(\tau^k_s, w_n) = 0 \) for \( k = 2, \ldots, M + 1 \). Because of compactness \( \tau^k_s \) has a subsequence converging to some \( \tau^k_s \), \( k = 2, \ldots, M + 1 \). Again, we must have \( \tau^k_s = t^* \) and hence also \( \partial^k_t f(t^*, w^*) = 0 \), for all \( k = 2, \ldots, M + 1 \). Since we already showed that \( f(t^*, w^*) = \partial_t f(t^*, w^*) = 0 \), this contradicts the assumption that \( f(\cdot, w^*) \) may have zeros of order up to \( M \).

The argument for \( p \geq 2 \) is similar. For \( t \) between two connected components of \( \Sigma^p \), at least one of conditions in (5.5) breaks down. Note that since we assumed that the total number of components of \( \Sigma^p(w_n, \epsilon_n) \) is larger than \( K(M + 1)^2 \) and the number of components of the larger set \( \Sigma(w^*, \epsilon_1) \) is at most \( K \), we will have at least \( (M + 1)^2 + 1 \) components of \( \Sigma^p(w_n, \epsilon_n) \) in \( I \). This means that these \( p + 1 \) conditions fail at least \( (M + 1)^2 \) times. Since \( p + 1 \leq M + 1 \), there is a condition that will fail at least \( M + 1 \) times.

In the case this is the last condition \( |\partial^p_t f(t, w)| \geq \epsilon_n^p \) that fails while conditions \( |\partial^k_t f(t, w)| \leq \epsilon_n^k \) remain valid, we can argue similar to the second case of the argument with \( p = 1 \). In this we need to have at least \( M - p \) different zeros of \( \partial^{p+1}_t f(t, w) \), which is the case if we have at least \( M - p + 1 \) such components in \( I \). This is true since we have at least \( M + 1 \) such components.

If one of the other conditions fails, let us take the smallest \( i \) for which condition \( |\partial^i_t f(t, w)| \leq \epsilon_n^i \) fails \( M + 1 \) times. Again, we need to have at least \( M - i \) different
Lemma 5.5. The length of each connected component of $\Sigma^p(w, \epsilon)$, for $p = 1, \ldots, M$, is no greater than $C\epsilon^{\alpha_p - 1 - \alpha_p}$.

Proof. Let $I$ be a connected component of $\Sigma^p(w, \epsilon)$ and let $t^* \in I$. We are going to estimate the maximal shift $\delta_0 > 0$ such that $t^* + \delta \in I$ for all $0 < \delta < \delta_0$. Recalling the definition of $\Sigma^p(w, \epsilon)$, we see that

\begin{equation}
|\partial^p_{\delta} f(t^* + \delta, w)| \leq \epsilon^{\alpha_p - 1}, \quad |\partial^p_{\delta} f(t^* + \delta, w)| \geq \epsilon^{\alpha_p}.
\end{equation}

Using the Taylor expansion of $\partial^p_{\delta} f(\cdot, w)$ at $t^*$, we have

$$
\partial^p_{\delta} f(t^* + \delta, w) = \partial^p_{\delta} f(t^*, w) + \partial^p_{\delta} f(t^*, w)\delta,
$$

where $t^{**}$ is some point between $t^*$ and $t^* + \delta$. Since $t^*, t^* + \delta \in I$, it follows that $t^{**} \in I$ and

$$
|\partial^p_{\delta} f(t^{**}, w)| |\delta| \leq |\partial^p_{\delta} f(t^* + \delta, w)| + |\partial^p_{\delta} f(t^*, w)| \leq 2\epsilon^{\alpha_p - 1}.
$$

From this and (5.11) we obtain that $|\delta_0| \leq C\epsilon^{\alpha_p - 1 - \alpha_p}$. Consequently, the length of each connected component of $\Sigma^p(w, \epsilon)$ can be estimated by $C\epsilon^{\alpha_p - 1 - \alpha_p}$. \hfill \Box

Our next step is to show that each connected component of $\Sigma^p(w, \epsilon)$, for $p = 1, \ldots, M$, is small enough.

Now we can finish the proof of Proposition 5.3. Let us choose $\alpha_k = 1 - k/2M$, $k = 0, \ldots, M$. According to Lemma 5.5 the length of each connected component of $\Sigma^p(w, \epsilon)$ can be estimated by $C\epsilon^{1/2M}$. Then according to Lemma 5.4, the size of $\Sigma^p(w, \epsilon)$ can be estimated by $CK(M + 1)^2\epsilon^{1/2M}$. Because of decomposition (5.6) the size of $\Sigma(w, \epsilon)$ is estimated by $C\epsilon^{1/2M}$. Statement of Proposition 5.3 is now a consequence of continuity of $f$ with respect to $w$ and compactness of $W$. Remark 5.1 follows with the choice $\alpha_M = \delta$ and $\alpha_k = 1 - k/ M$, $0 \leq k \leq M - 1$. \hfill \Box

The following interpolation lemma shows that if a function $u$ can be decomposed for all sufficiently small $\epsilon$ into a sum of $u_1^{(\epsilon)} + u_2^{(\epsilon)}$ with a good estimate for the norm of $u_2^{(\epsilon)}$ in a “bad” Sobolev space $H^p$ with small index $p$, and with a bad estimate for the norm of $u_1^{(\epsilon)}$ in a “good” Sobolev space $H^r$ with large index $r$, then $u$ belongs to some intermediate space $H^q$ with $p < q < r$.

Lemma 5.6. Let $p < r$ and $S, T > 0$. Let $u \in H^p(\mathbb{R}^n)$. Suppose that for all small enough $\epsilon > 0$ there is a representation $u = u_1^{(\epsilon)} + u_2^{(\epsilon)}$ such that

\begin{equation}
\|u_1^{(\epsilon)}\|_{H^r} \leq C\epsilon^{-T}, \quad \|u_2^{(\epsilon)}\|_{H^r} \leq C\epsilon^S,
\end{equation}

Then $u \in H^q(\mathbb{R}^n)$ for any

\begin{equation}
q < (pT + rS)(T + S)^{-1}.
\end{equation}
Proof. Let a sequence \( \{a_j\} \) be such that \( 0 = a_0 < a_1 < \ldots \) and \( a_j \to +\infty \) as \( j \to +\infty \), and let a sequence \( \{b_j\} \) of positive numbers \( b_j > 0 \) tend to zero. Then for each \( q \) we have the following estimates

\[
\|u\|_{H^n}^2 = \int_{\mathbb{R}^n} |\hat{u}|^2(1 + |\xi|^2)^q d\xi
\]

\[
= 2 \sum_{j=0}^{+\infty} \left( \int_{a_j \leq |\xi| \leq a_{j+1}} |u_2^{(b_j)}|^2(1 + |\xi|^2)^q d\xi \right)
\]

\[
\leq 2 \sum_{j=0}^{+\infty} \left( \max \{\langle a_j \rangle^{2(q-p)}, \langle a_{j+1} \rangle^{2(q-p)} \} \times \right.
\]

\[
\left. \times \int_{a_j \leq |\xi| \leq a_{j+1}} |u_2^{(b_j)}|^2(1 + |\xi|^2)^p d\xi \right)
\]

\[
+ \max \{\langle a_j \rangle^{2(q-r)}, \langle a_{j+1} \rangle^{2(q-r)} \} \int_{a_j \leq |\xi| \leq a_{j+1}} |u_2^{(b_j)}|^2(1 + |\xi|^2)^r d\xi .
\]

It follows from this estimate and (5.12) that

\[
\|u\|_{H^n}^2 \leq C \sum_{j=0}^{+\infty} \left( \max \{\langle a_j \rangle^{2(q-p)}, \langle a_{j+1} \rangle^{2(q-p)} \} b_j^{2S} \right.
\]

\[
\left. + \max \{\langle a_j \rangle^{2(q-r)}, \langle a_{j+1} \rangle^{2(q-r)} \} b_j^{-2T} \right) .
\]

(5.14)

Now we are going to demonstrate that under hypothesis (5.13) we may choose sequences \( \{a_j\} \) and \( \{b_j\} \) in such way that the right hand side of (5.14) will be finite, so that conclusion of Lemma 5.6 will follow. Let us set \( a_j = j^\alpha \) and \( b_j = j^{-\beta} \) with some \( \alpha, \beta > 0 \). Then the series in (5.14) will converge if the following inequalities are fulfilled

\[ 2\alpha(q-p) - 2\beta S < -1, \quad 2\alpha(q-r) + 2\beta T < -1. \]

These inequalities can be transformed into

\[ \alpha(2T(q-p) + 2S(q-r)) < -T - S, \quad 2\beta N > \alpha 2(q-p) + 1, \]

which hold with positive constants \( \alpha \) and \( \beta \) if and only if (5.13) is valid. \( \square \)

6. Spectral asymptotics

In this section we will prove Theorem 2.5. Let \( X \) be a smooth compact manifold without boundary of dimension \( n \geq 3 \). Then an elliptic self-adjoint operator \( P(x,D) \) has a collection of eigenfunctions and eigenvalues \( \lambda_j \to \infty \). We will be interested in distribution of eigenvalues and will find the asymptotics for the spectral function \( N(\lambda) = \sharp \{ \lambda_j : \lambda_j < \lambda \} \). One of the most effective methods to study such asymptotics is to use an explicit representation for the fundamental solution of the corresponding hyperbolic problem (1.1). We will follow the method developed in [8], [7], [26], etc., for scalar operators, and in [24] for the case \( M = 1 \) in Condition C.
In previous sections, and in particular in the proof of Theorem 2.2, we have represented the propagator \( U(t) \) in (1.1) as an infinite sum of some extensions of Fourier integral operators, i.e.

\[
U(t) = \sum_{j=0}^{\infty} e^{-i\hat{A}t} Q_j,
\]

where \( Q_0 = I \) and \( \text{Tr} Q_1 = 0 \). So we also have

\[
\text{Tr} e^{-i\hat{A}t} Q_1 = 0.
\]

The following proposition was implicitly proved in [7] and formulated in [24].

**Proposition 6.1.** Let \( \chi_1, \chi_2 \in C_0^\infty(\mathbb{R}) \) be such that \( \chi_1(0) = 1 \) and \( 0 \not\in \text{supp} \chi_2 \). Suppose that

\[
\text{Tr} \mathcal{F}_{t-\mu}^{-1}(\chi_1(t) U(t)) = c_1\mu^{n-1} + c_2\mu^{n-2} + o(\mu^{n-2}),
\]

(6.3)

\[
\text{Tr} \mathcal{F}_{t-\mu}^{-1}(\chi_2(t) U(t)) = o(\mu^{n-1}), \quad \mu \to \infty.
\]

Then \( N(\lambda) = c_1n^{-1}\lambda^n + c_2(n-1)^{-1}\lambda^{n-1} + o(\lambda^{n-1}) \).

Let us first consider the contribution of the first term, which is \( e^{-i\hat{A}t} \). It is the propagator for a block-diagonal system, so the asymptotic behaviour of \( \mathcal{F}_{t-\mu}^{-1}(\chi_\sigma(t) e^{-i\hat{A}t}) \), \( \sigma = 1, 2 \), determines the spectral distribution for a system of independent scalar equations. Thus, this is the sum of spectral distributions for scalar operators, which are well-known (e.g. [8], [7]). So

\[
\mathcal{F}_{t-\mu}^{-1} \text{Tr}(\chi_1(t) e^{-i\hat{A}t}) = c_1\mu^{n-1} + c_2\mu^{n-2} + o(\mu^{n-2}).
\]

We will show that under suitable conditions the asymptotics in (6.3) are determined only by the first term of the series (6.1). Let us first observe that in view of Theorem 2.1 operator \( \sum_{j=0}^{\infty} e^{-i\hat{A}t} Q_j \) is compact for sufficiently large \( K \), so it does not contribute to the singularity of \( \text{Tr} \chi_\sigma U \), \( \sigma = 1, 2 \). Therefore, we can replace \( U(t) \) in (6.3) by a sum of finitely many terms in (6.1). Also, in view of (6.2), the second term of (6.1) does not contribute to (6.3). Other terms of the sum (6.1) are of the form

\[
e^{-i\hat{A}t} \int_0^t \int_0^{t_1} \cdots \int_0^{t_l} Z(t_1) Z(t_2) \cdots Z(t_{l+1}) dt_{l+1} \cdots dt_1, \quad l \geq 2,
\]

where \( Z(\tau) = -ie^{i\hat{A}\tau} B e^{-i\hat{A}\tau} \). In this way, after a change of variables \( s_1 = t - t_1 \), \( s_2 = t_1 - t_2 \), \ldots, \( s_{l+1} = t_l \), we obtain a sum of terms of the form

\[
I_\sigma(\mu) = \text{Tr} \int \chi_\sigma(s_1 + \cdots + s_{l+1}) e^{i(s_1 + \cdots + s_{l+1})\mu} B(s) L(s) ds, \quad \sigma = 1, 2.
\]

where \( B(s) \in \Psi^0, L(s) = e^{-i\hat{A}_j s_1} e^{-i\hat{A}_j s_2} \cdots e^{-i\hat{A}_j s_{l+1}} \), and \( \hat{A}_j \) is the \( j \)-th block of \( \hat{A} \) corresponding to \( a_j \). As before, \( L(s) \) may be represented as a locally finite sum of oscillatory integrals with phase \( \varphi(s, x, y, \xi) \), for which

\[
\frac{\partial \varphi}{\partial s_k} = S_{jk}(s, x, \frac{\partial \varphi}{\partial x}),
\]

where

\[
S_{jk}(s, x, p) = a_{jk}(\Phi_{j_{k-1}} s_1 \cdots \Phi_{j_1}(x, p)).
\]
Therefore,

\[ I_\sigma(\mu) = \int \int_X \int_{\mathbb{R}^n} \chi_\sigma(s_1 + \cdots + s_{l+1}) e^{i(s_1 + \cdots + s_{l+1}) \mu + i \varphi(s, x, \xi)} b(s, x, \xi) d\xi dxds. \]

Substituting \( \xi = \mu \tau \omega, \tau > 0, |\omega| = 1 \), we get

(6.4) \[ I_\sigma(\mu) = \mu^n \int \int_X \int_{S^{n-1}} \int_{0}^{\infty} \chi_\sigma(s_1 + \cdots + s_{l+1}) e^{i \mu (s_1 + \cdots + s_{l+1} + \tau \varphi)} b \tau^{n-1} d\tau d\omega dxds. \]

To finish the argument for this operator, we can follow [24] to show that smoothing of \( Q_l \)'s implies (6.3). If we change variables again by \( \rho = \sum s_j, s_j = \rho \kappa_j \) and introduce \( K = \{ \kappa_j \geq 0, \sum \kappa_j = 1 \} \), we get

\[ I_1(\mu) = \mu^n \int_K \int_X \int_{S^{n-1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi_1(\rho) e^{i \mu (\rho + \tau \varphi(\rho \kappa, x, \omega))} \rho \tau^{n-1} d\tau d\rho d\omega dxd\kappa. \]

For fixed \( \omega, x, \kappa \) the point

\[ \tau = -\left( \sum \kappa_k S_{jk}(\rho \kappa, x, \partial \varphi/\partial x) \right)^{-1}, \rho = 1, \]

is a non-degenerate stationary point of this oscillatory integral, implying that

\[ I_1(\mu) = O(\mu^{n-1-l}) = o(\mu^{n-2}). \]

This already gives Hörmander’s first term of \( N(\lambda) \). Now we will show that singularities at \( t \neq 0 \) do not give essential contributions to the second term of spectral asymptotics. In the analysis of \( \text{Tr} \chi_2 U \), the contribution of the first term of (6.1) was established in [7].

Let us look at the integral (6.4) with respect to \( s_1, s_2, \tau \) with fixed \( x, \omega, s_3, \ldots, s_{l+1} \). The stationary phase method with respect to \( \tau, s_1 \) gives a stationary point

\[ \varphi(s, x, x, \omega) = 0, \tau = -(\partial \varphi/\partial s_1)^{-1}. \]

It is non-degenerate because \( \det (\partial_\tau \partial s_1 (\tau \varphi)) = -a_1(s, x, \partial_x \varphi)^2 \). This gives the estimate

\[ I_2(\mu) = O(\mu^{n-1}). \]

Since \( \partial \varphi/\partial s_2 = 0 \) only at \( \tau = -(\partial \varphi/\partial s_2)^{-1} = a_2(s, x, \partial_x \varphi)^{-1} \), the phase is stationary with respect to \( s_2 \) only at \( (x, \xi) \) for which \( a_1(s, x, \partial_x \varphi) = a_2(s, x, \partial_x \varphi) \). This is the set of measure zero and, therefore,

\[ I_2(\mu) = o(\mu^{n-1}), \]

which shows (6.3). Finally we note that if the support of \( \chi_1 \) is sufficiently small (i.e. when \( s \) is small), terms in (5.1), corresponding to \( \tilde{x} \) and \( \tilde{y} \) in different sheets of \( \tilde{X} \) will produce only smoothing operators (5.1) in view of the finite propagation speed. So their contribution to \( I(\mu) \) is rapidly decreasing and we obtain the result also on \( X \).
REFERENCES


ILIA KAMOTSKI:
MATHEMATICAL SCIENCES
UNIVERSITY OF BATH
BATH BA2 7AY
UNITED KINGDOM

MICHAEL RUZHANSKY:
DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE LONDON
180 QUEEN’S GATE, LONDON SW7 2AZ
UNITED KINGDOM
E-mail addresses I.Kamotski@maths.bath.ac.uk, m.ruzhansky@imperial.ac.uk