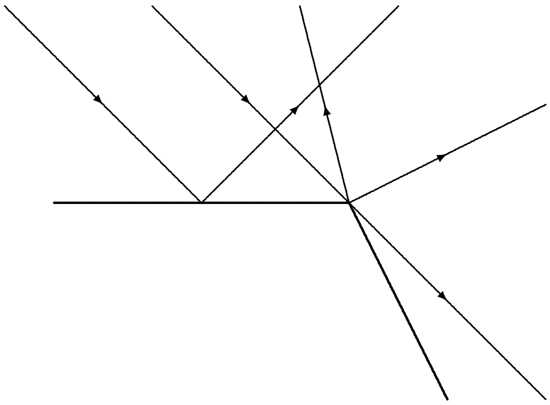
Geometric optics and the wave equation on manifolds with corners

András Vasy

Helsinki, June 2004

According to geometric optics, light propagates in straight lines (in homogeneous media), reflects/refracts from surfaces according to Snell's law: energy and tangential momentum are conserved.



But light also satisfies the wave equation,

$$Pu = 0, \quad Pu = D_t^2 u - \Delta_g u,$$

Δ_g is the Laplacian, so it is $c^2 \sum_{j=1}^n D_{x_j}^2$ in $\mathbb{R}^n$, where c is the speed of light (this corresponds to a Riemannian metric $g = c^{-2} \sum dx_j^2$), $D_{x_j} = \frac{1}{i}\partial_{x_j}$, with suitable boundary conditions.

One way of discussing the relationship between these is that singularities (lack of smoothness) of solutions of $Pu = 0$ follow geometric optics rays. In order to describe this precisely, I discuss a more general setting.

Let's start with the boundaryless case. So suppose X is a manifold without boundary of dimension n. The standard setting for microlocal analysis is then T^*X. If z_j are local coordinates on X, and we write one-forms as $\sum \zeta_j \, dz_j$, then (z_j, ζ_j), $j = 1, \ldots, n$, are local coordinates on T^*X. (Actually, they are global on the fibres of $T^*X \to X$.)

Now, T^*X, as a vector bundle, is equipped with an $\mathbb{R}^+$-action: $\mathbb{R}^+_s \times T^*X \ni (s, z, \zeta) \mapsto (z, s\zeta)$.

It is also a symplectic manifold, equipped with a canonical symplectic form ω, $\omega = \sum d\zeta_j \wedge dz_j$ in local coordinates.

If P is a differential operator on X, say $P = \sum_{|\alpha| \leq m} a_\alpha(z) D_z^\alpha$, one can associate a principal symbol $p = \sigma_m(P) = \sum_{|\alpha|=m} a_\alpha(z) \zeta^\alpha$ to P; this is homogeneous degree m with respect to the $\mathbb{R}^+$-action.

In fact, the same works for pseudodifferential operators, $P \in \Psi^m_{cl}(X)$; now $p = \sigma_m(P)$ is homogeneous degree m on $T^*X \setminus o$, o denoting the zero section.

The symplectic form ω turns p, or dp, into a vector field H_p on T^*X via $\omega(V, H_p) = Vp$ for all vector fields V. Thus,

$$H_p = \sum_j \frac{\partial p}{\partial \zeta_j} \frac{\partial}{\partial z_j} - \frac{\partial p}{\partial z_j} \frac{\partial}{\partial \zeta_j}.$$

Note that H_p is homogeneous of degree $m - 1$.

Bicharacteristics are then integral curves of H_p inside the characteristic set $\Sigma = p^{-1}(\{0\})$.

To do analysis, we also need a notion of singularity of a function or distribution u. The roughest notion is that of the wave front set $\mathrm{WF}(u)$, which locates at which points and in which direction a function u is not smooth, here meaning C^∞. This is defined as a closed conic subset of $T^*X \setminus o$; u is C^∞ if and only if $\mathrm{WF}(u) = \emptyset$. In fact, for any point $z_0 \in X$, z_0 has a neigborhood in X on which u is C^∞ if and only if $\mathrm{WF}(u) \cap (T^*_{z_0}X \setminus o) = \emptyset$. We do not give the precise definition here, although it is easily done analogously to $\mathrm{WF}_b(u)$, which we define later.

One can measure singularities with respect to other spaces: e.g. the Sobolev spaces $H^s_{\mathrm{loc}}(X)$, where we would write $\mathrm{WF}^s(u)$, or with respect to real analytic functions, where we would write $\mathrm{WF}_A(u)$.

As an example: $\mathrm{WF}(\delta_0) = \{(0, \zeta) : \zeta \neq 0\} = N^*\{0\} \setminus o$, and if Ω has a C^∞ boundary, and χ_Ω is the characteristic function of Ω, then $\mathrm{WF}(\chi_\Omega) = N^*\partial\Omega \setminus o$.

The main facts about the analysis of P are:

1. Microlocal elliptic regularity. Let $\Sigma(P) = p^{-1}(\{0\})$ be the characteristic set of P. If $u \in \mathcal{D}'(X)$ then $\mathrm{WF}(u) \subset \mathrm{WF}(Pu) \cup \Sigma(P)$. In particular, if $Pu \in C^\infty(X)$ then $\mathrm{WF}(u) \subset \Sigma(P)$.

2. Propagation of singularities. Suppose that p is real, $Pu \in C^\infty(X)$. Then $\mathrm{WF}(u)$ is a union of maximally extended bicharacteristics in $\Sigma(P)$. That is, if $q \in \mathrm{WF}(u)$ (hence in $\Sigma(P)$) then so is the whole bicharacteristic through q.

Note that 2 may be vacuous; indeed, if H_p is radial, i.e. tangent to the orbits of the $\mathbb{R}^+$-action, then it does not give any information on $\mathrm{WF}(u)$, as the latter is already known to be conic. Such points are called radial points, and in recent work with Hassell and Melrose, they have been extensively analyzed under non-degeneracy assumptions. If P is the wave operator, there are no radial points in $\Sigma = \Sigma(P)$, but such points are very important in scattering theory (where the $\mathbb{R}^+$-action, or its remnants, are in the base variables z).

Example: $P = D_t^2 - \Delta_g$ the wave operator, $X = M \times \mathbb{R}$, M a manifold without boundary. Then $p = \sigma_2(P) = \tau^2 - |\xi|_g^2$, where (x, t, ξ, τ) are coordinates on T^*X, and the projection of bicharacteristics to M are geodesics. If $M \subset \mathbb{R}^n$ and g is the Euclidean metric, then $H_p = 2\tau\partial_t - 2\xi \cdot \partial_x$, and bicharacteristics inside $p = 0$, i.e. $|\tau| = |\xi|$, are straight lines

$$s \mapsto (x_0 - 2s\xi_0, t_0 + 2\tau_0 s, \xi_0, \tau_0),$$

which explains geometric optics in the absence of boundaries.

We can now turn to boundaries and corners. Roughly, the results have the same form, but the definitions of wave front set and the bicharacteristics change significantly. In particular, the relevant wave front set is $\mathrm{WF}_b(u)$, introduced by Melrose. Both $\mathrm{WF}_b(u)$ and the image of the (generalized broken) bicharacteristics are subsets of the b-cotangent bundle, which we now define.

So suppose X is a manifold with corners. Locally this means that X is diffeomorphic to an open subset of $[0,\infty)^k \times \mathbb{R}^{n-k}$; we denote (x,y) the corresponding coordinates. Let $\mathcal{V}(X)$ be the set of C^∞ vector fields on X, and $\mathcal{V}_b(X)$ the subset consisting of vector fields tangent to every face of X. Here 'b' stands for 'boundary'. Note that both $\mathcal{V}(X)$ and $\mathcal{V}_b(X)$ are Lie algebras. Locally, over a coordinate chart U, $V \in \mathcal{V}_b(X)$ means that

$$V = \sum a_j x_j \, \partial_{x_j} + \sum b_j \, \partial_{y_j},$$

with $a, b \in C^\infty(U)$. Thus, $\mathcal{V}_b(X)$ is the set of all C^∞ sections of a vector bundle, ${}^bTX$.

Its dual bundle is ${}^bT^*X$, with basis $\frac{dx_j}{x_j}$, dy_j (over $C^\infty(U)$), so $\alpha \in {}^bT^*X$ means that

$$\alpha = \sum \sigma_j \, \frac{dx_j}{x_j} + \sum \zeta_j \, dy_j.$$

Notice that ${}^bT^*_{X^\circ}X$ is naturally identified with T^*X°, and there is a map $\pi : T^*X \to {}^bT^*X$ which in local coordinates takes the form

$$\pi(x,y,\xi,\zeta) = (x,y,x\xi,\zeta),$$
$$\text{with } x\xi = (x_1\xi_1,\ldots,x_k\xi_k).$$

Thus, π is a C^∞ map, but at ∂X, it is not a diffeomorphism.

There is an algebra $\Psi_b(X)$ of pseudodifferential operators microlocalizing $\mathcal{V}_b(X)$, and the corresponding algebra of differential operators, $\mathrm{Diff}_b(X)$. Formally, these can be thought to be of the form $A = a(x,y,xD_x,D_y)$, where $xD_x = (x_1D_{x_1},\ldots,x_kD_{x_k})$, and a is a symbol in the last two slots (so e.g. it is, at least asymptotically, homogeneous). Key properties of $\Psi_b^m(X)$ include the existence of a principal symbol, $\sigma_{b,m}$, which for each $A \in \Psi_b^m(X)$ gives $a = \sigma_{b,m}(A)$ homogeneous degree m on ${}^bT^*X \setminus o$, and boundedness on spaces such as

$L^2(X)$ (at least if the Schwartz kernel has compact support).

The replacement spaces for $C^\infty(X)$ are spaces of conormal functions. These are functions which remain in a fixed function space $\mathcal{H}$ under iterated applications of b-vector fields, or b-differential operators. An example is the space $H^{0,\infty}_{b,\mathrm{loc}}(X)$ of u such that $Lu \in L^2_{\mathrm{loc}}(X)$ for all $L \in \mathrm{Diff}_b(X)$; here $\mathcal{H} = L^2_{\mathrm{loc}}(X)$.

This space microlocalizes, and, for $u \in L^2_{\mathrm{loc}}(X)$, we say that $q \in {}^bT^*X \setminus o$ is *not* in $\mathrm{WF}_b(u)$ if there exists $A \in \Psi^0_b(X)$ such that $\sigma_{b,0}(A)(q) \neq 0$ (i.e. A is elliptic at q) and $Au \in H^{0,\infty}_{b,\mathrm{loc}}(X)$.

In fact, for solutions of the wave equation, it is more natural to measure b-regularity relative to $H^1_{\mathrm{loc}}(X)$, and ask whether u is microlocally (in the above sense) in $H^{1,\infty}_{b,\mathrm{loc}}(X)$. It turns out that this actually makes no difference for the statement of the propagation theorems for solutions of the wave equation, although it does for the proofs, so we do not emphasize this point here.

We now start considering $P = D_t^2 - \Delta_g$ on $X = M \times \mathbb{R}$, where g is a C^∞ Riemannian metric on M. Let $\Sigma(P) = p^{-1}(\{0\}) \subset T^*X \setminus o$. Although p is a function on T^*X, the locus of microlocal analysis is ${}^bT^*X$, which is the cause of some technical complications.

There are three kinds of points in ${}^bT^*X$:

- elliptic points:

$$\mathcal{E} = \{q \in {}^bT^*X \setminus o : \\ \pi^{-1}(\{q\}) \cap \Sigma(P) = \emptyset\},$$

- glancing points:

$$\mathcal{G} = \{q \in {}^bT^*X \setminus o : \\ \mathrm{Card}(\pi^{-1}(\{q\}) \cap \Sigma(P)) = 1\},$$

- hyperbolic points:

$$\mathcal{H} = \{q \in {}^bT^*X \setminus o : \\ \mathrm{Card}(\pi^{-1}(\{q\}) \cap \Sigma(P)) \geq 2\}.$$

In T^*X°, π is the identity map, so $q \in T^*X^\circ$ is either in $\mathcal{E}$ or in $\mathcal{G}$ depending on whether $q \notin \Sigma(P)$ or $q \in \Sigma(P)$.

Note also that there are two ways for $q \in {}^bT^*X \setminus o$ to lie in $\mathcal{E}$: either q is not in the image of $\pi : T^*X \to {}^bT^*X$, which happens if $\sigma(q) \neq 0$, and this is independent of what P is, or it does lie in the image of π, but not in the image of $\pi|_{\Sigma(P)}$.

To give a feeling for these sets, we describe them in local coordinates. One can arrange, by adjusting the y_j if needed, that

$$
\begin{aligned}
&g(x,y,\xi,\zeta)\\
&\;=\sum A_{ij}(x,y)\xi_i\xi_j\\
&\qquad\quad+\sum C_{ij}(x,y)\xi_i\zeta_j+\sum B_{ij}(x,y)\zeta_i\zeta_j,
\end{aligned}
$$

with A,B,C in C^∞, $C_{ij}(0,y)=0$. Then

$$
p|_{x=0}=\tau^2-\xi\cdot A(0,y)\xi-\zeta\cdot B(y)\zeta,
$$

and A is positive definite, so:

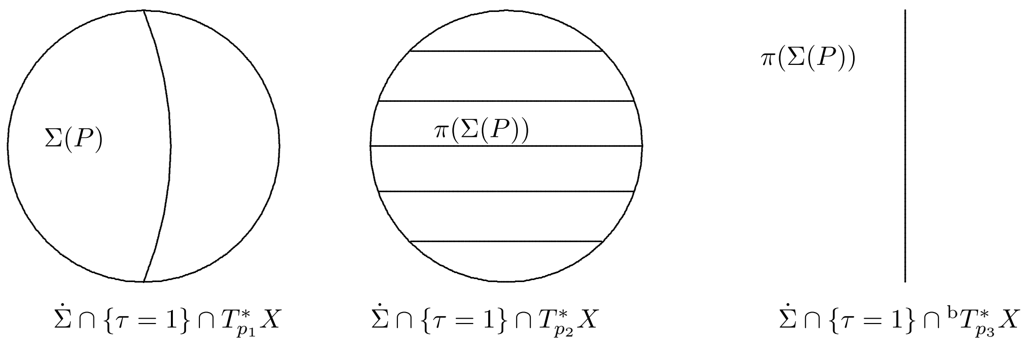
- $$\mathcal{E} \cap \{x = 0\} = \{(0, y, t, \sigma, \zeta, \tau) : \\ \sigma \neq 0 \text{ or } \tau^2 < \zeta \cdot B(0, y)\zeta\},$$

- $$\mathcal{G} \cap \{x = 0\} = \{(0, y, t, 0, \zeta, \tau) : \\ \tau^2 = \zeta \cdot B(0, y)\zeta\},$$

- $$\mathcal{H} \cap \{x = 0\} = \{(0, y, t, 0, \zeta, \tau) : \\ \tau^2 > \zeta \cdot B(0, y)\zeta\}.$$

We let $\dot{\Sigma} = \pi(\Sigma) = \mathcal{G} \cup \mathcal{H}$ be the compressed characteristic set.

Below we illustrate these sets for a 3-dimensional manifold with corners M. In the pictures we depict $\dot\Sigma$ inside ${}^bT^*X$ over points p_1, p_2, p_3 which are, respectively, in the interior of X, in the interior of a boundary hypersurface, and in the interior of a codimension 2 corner. Since the fiber of ${}^bT^*X$ over each p_j is 4-dimensional, we only show its intersection with $\tau = 1$. Then $\dot\Sigma \cap \{\tau = 1\} \cap {}^bT^*_{p_j}X$ is, respectively, a sphere ($|\zeta|^2 = 1$), a disk and an interval. The variables normal to the plane of the disk, resp. to the interval, are σ_1, resp. (σ_1, σ_2). Thus the sphere was compressed into the disk, resp. interval.



Over p_1, $\mathcal{H}$ is empty, and all of $\Sigma(P)$ is $\mathcal{G}$ (the complement of which is $\mathcal{E}$). Over p_2 and p_3, $\mathcal{H}$ is the interior of the disk, resp. the interval, while $\mathcal{G}$ is its boundary.

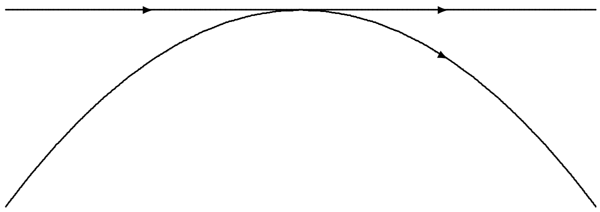
Generalized broken bicharacteristics are curves inside $\dot{\Sigma}$, namely they are continuous maps $\gamma : I \to \dot{\Sigma}$, where I is an interval (here $\dot{\Sigma}$ has the subspace topology of ${}^bT^*X$), such that

- If $q_0 = \gamma(s_0) \in \mathcal{G}$, and $\hat{q}_0$ is the unique point in $\Sigma(P)$ with $\pi(\hat{q}_0) = q_0$, then for all f, C^∞ on a neighborhood of q_0 in ${}^bT^*X$,
$$\frac{d}{ds}(f \circ \gamma)(s_0) = (H_p \pi^* f)(\hat{q}_0).$$

- If $q_0 = \gamma(s_0) \in \mathcal{H}$, and q_0 is in the interior of a face $x = 0$, then there is $\epsilon > 0$ such that for $s \in I$ with $0 < |s - s_0| < \epsilon$, $x(\gamma(s)) \neq 0$.

- If $q_0 = \gamma(s_0) \in \mathcal{G}$ lies over the interior of a boundary hypersurface, then there is $\epsilon > 0$ such that $\gamma|_{(s_0-\epsilon, s_0+\epsilon)}$ is a generalized broken bicharacteristic in the sense of Melrose and Sjöstrand.

The last part is relevant because in a C^∞ setting, there is a difference between gliding and diffractive points inside $\mathcal{G}$, unlike in the analytic setting.



Leaving out the last condition would thus weaken the following theorem.

Snell's law is encoded in the statement that γ is continuous. Thus, any (locally defined) smooth functions on ${}^bT^*X$, such as $x, y, t, \sigma, \zeta, \tau$, are continuous along γ. However, $\xi_j = x_j^{-1}\sigma_j$ is *not* continuous, so the normal momentum may jump.

The first two criteria in the definition can be unified, which in fact allows generalizations such as N-body scattering with bound states, but we do not need it here.

Recall that $H_0^1(X)$ is the completion of $C_c^\infty(X^\circ)$ in the norm $\|u\|^2_{H^1(X)} = \|du\|^2_{L^2(X)} + \|u\|^2_{L^2(X)}$, and that elements of $H_0^1(X)$ restrict as 0 to ∂X, i.e. $u \in H^1_{0,\mathrm{loc}}(X)$ means that u satisfies the Dirichlet boundary conditions.

Theorem 1 *Suppose $Pu = 0$, $u \in H^1_{0,\mathrm{loc}}(X)$. Then $\mathrm{WF}_b(u) \subset \dot{\Sigma}$, and it is a union of maximally extended generalized broken bicharacteristics of P.*

This theorem can be stated in a completely microlocal manner. In addition, it also holds for Neumann boundary conditions.

It was proved in the real analytic setting by Lebeau, and in the C^∞ setting with C^∞ boundaries (and no corners) by Melrose, Sjöstrand and Taylor. This result is thus the C^∞ version of Lebeau's theorem: the geometry is similar in the real analytic vs. C^∞ settings, but the analysis is quite different.

Ideas of the proof: use positive commutator estimates, b-microlocalization, i.e. use elements $A \in \Psi_b^m(X)$ as the commutants, and the Dirichlet form. Thus, regularity that is gained is b-regularity relative to $H^1_{\mathrm{loc}}(X)$. In particular, at hyperbolic points at boundary hypersurfaces, this proof differs from the Melrose-Sjöstrand-Taylor proofs, which rely on a parametrix construction. Roughly, at $\mathcal{H}$, the key positivity arises from

$$H_p \sum \sigma_j = H_p \sum x_j \xi_j = -2\xi \cdot A(0,y)\xi$$

at $x = 0$, which is $2p - 2\tau^2 + 2\zeta \cdot B(0,y)\zeta$, hence negative at $\Sigma(P) \cap \pi^{-1}(\mathcal{H})$. Not coincidentally, this is quite similar to Mourre-type estimates in the N-body setting, though simpler in some ways and more complicated in other ways!

We give the arguments in a little more detail. They gain regularity for u iteratively, 1/2 a b-derivative at a time, relative to $H^1_{\mathrm{loc}}(X)$. Thus, we work with the spaces $H^{1,m}_{b,\mathrm{loc}}(X)$, consisting of functions u such that $Lu \in H^1_{\mathrm{loc}}(X)$ for $L \in \Psi^m_b(X)$. We can define a wave front set relative to $\mathcal{H} = H^{1,m}_{b,\mathrm{loc}}(X)$, just as before.

The other piece of technical machinery is that operators $A \in \Psi^m_b(X)$ also have an operator wave front set associated to them,

$$\mathrm{WF}'_b(A) \subset {}^bT^*X \setminus o,$$

which is again closed and conic. Roughly, if $A = a(x, y, xD_x, D_y)$, then a point $q = (x, y, \sigma, \zeta)$ is not in $\mathrm{WF}'_b(A)$ if a is a symbol of order $-\infty$, i.e. decays rapidly, in an open cone around q.

Then WF_b and WF'_b have the standard properties, e.g.

$$\mathrm{WF}_b(Au) \subset \mathrm{WF}'_b(A) \cap \mathrm{WF}_b(u),$$

which is to say that $\Psi^m_b(X)$ is 'b-microlocal'.

We are now ready to give some details of the proof.

In $\mathcal{E}$, we use the Dirichlet form:

$$\|d_M Au\|^2 - \|D_t Au\|^2 = \int_X |d_M Au|^2 - |D_t Au|^2 \, dg,$$

which is equal to $-\langle PAu, Au\rangle$ for $u \in H^1_{0,\mathrm{loc}}(X)$. Thus, modulo a commutator, which is negligible (because it is lower order), the Dirichlet form is equal to $-\langle APu, Au\rangle$.

A priori, this works if $A \in \Psi_b^0(X)$, and gives a bound for the left hand side in terms of APu and Au (which can be rearranged, if needed). We actually want to use it for $A \in \Psi_b^{1/2}(X)$, and this can be achieved by a regularization argument, approximating A by a family $A_r \in \Psi_b^0(X)$ which stays uniformly bounded in $\Psi_b^{1/2}(X)$ and converges to A in $\Psi_b^{1/2+\epsilon}(X)$, $\epsilon > 0$, as $r \to 0$. (In fact, we also want $A_r \to A$ 'microlocally in $\Psi_b^{-\infty}(X)$' away from $\mathrm{WF}_b'(A)$.)

Then

$$\|d_M A_r u\|^2 - \|D_t A_r u\|^2$$

stays uniformly bounded as $r \to 0$. If A has operator wave front set in $\mathcal{E}$, then it is straightforward to get that in fact $\|d_X A_r u\|^2$ remains bounded as $r \to 0$: either the normal derivatives D_{x_j} dominate all other derivatives, since $D_{x_j} = x_j^{-1}(x_j D_{x_j})$, or the D_{y_j} can be used to dominate D_t.

Then a standard argument gives that $Au \in H^1_{\mathrm{loc}}(X)$, so $\mathrm{WF}_b^{1,1/2}(u)$ is disjoint from the elliptic set of A. As A is arbitrary, as long as its operator wave front set lies in $\mathcal{E}$, this gives that $\mathrm{WF}_b^{1,1/2}(u) \cap \mathcal{E} = \emptyset$. An iterative argument gives $\mathrm{WF}_b^{1,\infty}(u) \cap \mathcal{E} = \emptyset$.

In $\mathcal{G} \cup \mathcal{H}$, we use positive commutator estimates, i.e. construct A such that $i[A^*A, P]$ is positive, modulo terms we can control. This is useful since

$$\begin{aligned}&\langle i[A^*A, P]u, u\rangle \\ &\qquad = \langle iA^*APu, u\rangle - \langle iPA^*Au, u\rangle \\ &\qquad = \langle iA^*APu, u\rangle + \langle u, iA^*APu\rangle,\end{aligned}$$

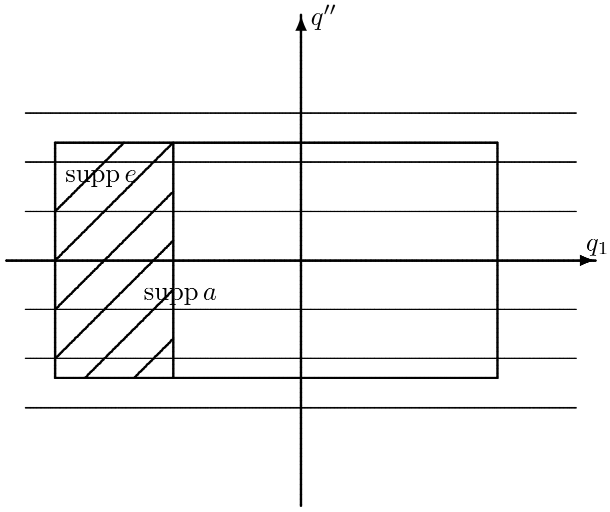
and the PDE gives information about Pu. (This argument needs to be regularized, as above.) Now, if $i[A^*A, P] = B^*B + E + F$ where F is lower order, hence negligible, we deduce that $\|Bu\|^2$ can be controlled provided we have information about $\langle Eu, u\rangle$, i.e. about u microlocally on $\mathrm{WF}_b'(E)$.

The shape this takes in the positive commutator estimates is that we need to assume that $\mathrm{WF}_b(u)$ is disjoint from $\mathrm{WF}_b'(E)$, and then we *propagate* the regularity of u and conclude that $\mathrm{WF}_b(u)$ is disjoint from the elliptic set of B. (At least after an iterative argument, as before.)

Apart from technical issues (namely that $P \notin \mathrm{Diff}_b^2(X)$), $i[A^*A, P] = B^*B + E + F$ as above is a condition on the principal symbol $\sigma_b(A)$, namely that $H_p \pi^* \sigma_b(A) = -b^2 + e$, with b and e playing the role of B and E. Thus, we want $a = \sigma_b(A)$ to be decreasing along bicharacteristics except in a region in which we have a priori information about u. It also needs to have small support.

In the simplest setting, namely in the interior of X, H_p can locally be made into a constant vector field $\frac{\partial}{\partial q_1}$ by a change of coordinates on $S^*X = (T^*X \backslash o)/\mathbb{R}^+$. Thus, one works with homogeneous degree zero functions $q = (q_1, q'')$ on $T^*X \setminus o$ which give coordinates on S^*X. In such coordinates it is straightforward to construct such an a as a C^∞ function on S^*X, and then regard it as a homogeneous function on $T^*X \setminus o$: one can take a to be the product of a function of q_1, on the derivative of which we have sign conditions, and an arbitrary function of q''.

Below we illustrate the bicharacteristics of H_p, which are now straight lines, and the support of a and e.



In general, at ∂X, the problem is microlocalization: there is no analogue of the functions of q'' – we cannot microlocalize by taking a product of two functions.

Instead, one follows a much more robust construction due to Melrose and Sjöstrand, which they used to prove propagation of singularities at $\mathcal{G}$ when ∂X was smooth, which would, however, take us beyond the limits of this talk.