DISTRIBUTIONS, FOURIER TRANSFORMS AND MICROLOCAL ANALYSIS

Notation

- $\mathbb{R}$ and $\mathbb{C}$ are the sets of real and complex numbers, respectively;
- $\mathbb{R}^n$ denotes the n-dimensional Euclidean space;
- $\operatorname{supp} f$ denotes the support of the function f; by definition, $\operatorname{supp} f$ is the closure of the set $\{x : f(x) \neq 0\}$;
- a multi-index α is a set of n non-negative integers, $\alpha := \{\alpha_1, \alpha_2, \dots, \alpha_n\}$;
- if α, β are multi-indices then $|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n$, $\alpha! := \alpha_1!\,\alpha_2! \dots \alpha_n!$ and $\alpha + \beta := \{\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n\}$;
- $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n)$, $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ are elements of $\mathbb{R}^n$;
- if $x \in \mathbb{R}^n$ and α is a multi-index then $x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$, $\partial_{x_k} := \frac{\partial}{\partial x_k}$, $\partial_x^\alpha := \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \dots \partial_{x_n}^{\alpha_n}$, $D_{x_k} := -i\,\partial_{x_k}$ and $D_x^\alpha := (-i)^{|\alpha|}\,\partial_x^\alpha$ where $i = \sqrt{-1}$;
- $C^\infty(\mathbb{R}^n)$ is the linear space of all infinitely differentiable functions on $\mathbb{R}^n$;
- $C_0^\infty(\mathbb{R}^n)$ is the linear space of all infinitely differentiable functions on $\mathbb{R}^n$ with compact supports.

References

[DS] M. Dimassi and J. Sjöstrand. *Spectral asymptotics in the semi-classical Limit.* LMS Lecture Notes Series, vol. 268, Cambridge University Press, 1999.

[H] L. Hörmander *The analysis of linear partial differential operators,* I–IV. Springer–Verlag, New York, 1984.

[Sh] M. Shubin. *Pseudodifferential operators and spectral theory.* “Nauka”, Moscow, 1978. English transl. Springer-Verlag, 1987.

[T] M. Taylor *Pseudodifferential operators.* Princeton Univ. Press, Princeton, New Jersey, 1981.

[Tr] F. Treves. *Introduction to pseudodifferential and Fourier integral operators,* I, II. Plenum Press, New York and London, 1980.

[SV] Y. Safarov and D. Vassiliev. *The asymptotic distribution of eigenvalues of partial differential operators.* American Mathematical Society, Providence, Rhode Island, 1996.

Course web site

http://www.mth.kcl.ac.uk/∼ysafarov/Lectures/LTCC2014a/index.html

Date: September 2014.

1. Fourier transform

1.1. Schwartz space $\mathcal{S}(\mathbb{R}^n)$.

Definition 1.1. We say that $f \in \mathcal{S}(\mathbb{R}^n)$ if the function f is infinitely differentiable and

$$\|f\|_{\alpha,m} := \sup_{x\in\mathbb{R}^n} (1+|x|)^m\, |\partial_x^\alpha f(x)| < \infty \tag{1.1}$$

for all multi-indices α and all $m = 0, 1, 2, \ldots$

Obviously, $\mathcal{S}(\mathbb{R}^n)$ is a linear space which contains $C_0^\infty(\mathbb{R}^n)$. If $f \in \mathcal{S}(\mathbb{R}^n)$ then, for all multi-indices α and all positive integers k, we have

$$|\partial_x^\alpha f(x)| \;\le\; \|f\|_{\alpha,m}\, (1+|x|)^{-m}\,.$$

In other words, the functions $f \in \mathcal{S}(\mathbb{R}^n)$ and all their derivatives decay faster than any negative power of $|x|$ as $|x| \to \infty$. Therefore these functions are said to be *rapidly decreasing*.

Example 1.2. The function $f(x) = e^{-|x|^2}$ belongs to $\mathcal{S}(\mathbb{R}^n)$.

Lemma 1.3.

- *If* $f \in \mathcal{S}(\mathbb{R}^n)$ *then* $x^\beta \partial_x^\gamma f(x) \in \mathcal{S}(\mathbb{R}^n)$.
- *If* $\|f\|_{\alpha,\beta} := \sup_{x\in\mathbb{R}^n} |x^\beta \partial_x^\alpha f(x)| < \infty$ *for all multi-indices* α, β *then* $f \in \mathcal{S}(\mathbb{R}^n)$.
- $\|f\|_{\alpha,m} \le \mathrm{const} \sum_{|\beta|\le m} \|f\|_{\alpha,\beta}$ *and* $\|f\|_{\alpha,\beta} \le \|f\|_{\alpha,|\beta|}$.

Proof is obvious.

We shall need the following version of Taylor's formula.

Lemma 1.4. *Let m be a non-negative integer, $f \in \mathcal{S}(\mathbb{R}^n)$ and $y \in \mathbb{R}^n$ be a fixed point. If f and all its derivatives up to the order m vanish at y then there exist functions $h_\beta \in \mathcal{S}(\mathbb{R}^n)$ such that*

$$f(x) \;=\; \sum_{\beta\,:\,|\beta|=m+1} (x-y)^\beta\, h_\beta(x)\,, \qquad \forall x \in \mathbb{R}^n\,. \tag{1.2}$$

Proof. Let $\zeta \in C_0^\infty(\mathbb{R}^n)$ and $\zeta \equiv 1$ in a neighbourhood of the point y. Denote $f_1 = (1-\zeta)f$ and $f_2 = \zeta f$. Obviously, the function $h(x) := |x-y|^{-2m-2} f_1(x)$ belongs to $\mathcal{S}(\mathbb{R}^n)$. We have

$$f_1(x) \;=\; |x-y|^{2m+2}\, h(x) \;=\; \sum_{\beta\,:\,|\beta|=m+1} (x-y)^\beta\, P_\beta(x-y)\, h(x),$$

where P_β are some polynomials. Thus, the function f_1 can be represented in the form (1.2). By Taylor's formula,

$$f_2(x) \;=\; \sum_{\beta\,:\,|\beta|=m+1} (x-y)^\beta\, \tilde h_\beta(x)\,,$$

where $\tilde h_\beta$ are some infinitely smooth functions. If $\tilde\zeta \in C_0^\infty(\mathbb{R}^n)$ and $\tilde\zeta \equiv 1$ on $\operatorname{supp}\zeta$ then, multiplying both parts of the above identity by $\tilde\zeta$, we obtain the expansion (1.2) for f_2. □

1.2. Convergence in the space $\mathcal{S}(\mathbb{R}^n)$.

Definition 1.5. We say that a sequence $\{f_k\} \subset \mathcal{S}(\mathbb{R}^n)$ converges to $f \in \mathcal{S}(\mathbb{R}^n)$ in the space $\mathcal{S}(\mathbb{R}^n)$ and write $f_k \overset{\mathcal{S}}{\to} f$ if $\|f - f_k\|_{\alpha,m} \to 0$ as $k \to \infty$ for all α and m.

The space $\mathcal{S}(\mathbb{R}^n)$ can be provided with a metric ρ such that $f_k \overset{\mathcal{S}}{\to} f$ if and only if $\rho(f, f_k) \to 0$. In particular, one can take

$$\rho(f,g) \ := \ \sum_{\alpha,m} \frac{\|f-g\|_{\alpha,m}}{(\alpha! + m^2)\,(1 + \|f-g\|_{\alpha,m})}\,.$$

1.3. Fourier transform in $\mathcal{S}(\mathbb{R}^n)$.

Definition 1.6. Let $f \in \mathcal{S}(\mathbb{R}^n)$. The function

$$\hat f(\xi) \ := \ \mathcal{F}_{x\to\xi} f(x) \ := \ (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\, f(x)\,\mathrm{d}x\,, \qquad \xi \in \mathbb{R}^n\,, \tag{1.3}$$

is called the *Fourier transform* of f.

The Fourier transform is well defined whenever the integral on the right hand side of (1.3) exists and is finite for all $\xi \in \mathbb{R}^n$. Obviously, this is true if $f \in \mathcal{S}(\mathbb{R}^n)$.

Remark 1.7. By means of (1.3) one can define the Fourier transform for functions f from the Lebesgue space $L_1(\mathbb{R}^n)$ (which contains $\mathcal{S}(\mathbb{R}^n)$ as a subspace). However, as we shall see later, the Fourier transform can be extended to $L_1(\mathbb{R}^n)$ and even more general classes of functions in a different, more elegant way.

Lemma 1.8. *For all $f \in \mathcal{S}(\mathbb{R}^n)$ and all multi-indices α we have*

$$\mathcal{F}_{x\to\xi}(D_x^\alpha f(x)) \ = \ \xi^\alpha\, \hat f(\xi)\,, \qquad \mathcal{F}_{x\to\xi}(x^\alpha f(x)) \ = \ (-1)^{|\alpha|}\, D_\xi^\alpha \hat f(\xi)\,. \tag{1.4}$$

Proof. The identities (1.4) are proved by integration by parts and differentiation under the integral sign. □

Corollary 1.9. *If $f \in \mathcal{S}(\mathbb{R}^n)$ then $\hat f \in \mathcal{S}(\mathbb{R}^n)$, and the map $\mathcal{F} : f \to \hat f$ is continuous in the space $\mathcal{S}(\mathbb{R}^n)$.*

Proof. For all $g \in \mathcal{S}(\mathbb{R}^n)$ we have

$$\begin{aligned}\sup_{\xi\in\mathbb{R}^n} |\hat g(\xi)| \ &= \ (2\pi)^{-n/2} \sup_{\xi\in\mathbb{R}^n} \left| \int e^{ix\cdot\xi}\, g(x)\,\mathrm{d}x \right| \ \le \ (2\pi)^{-n/2} \int |g(x)|\,\mathrm{d}x \\ &\le (2\pi)^{-n/2} \left(\int (1+|x|)^{-n-1}\,\mathrm{d}x \right) \sup_{x\in\mathbb{R}^n} (1+|x|)^{n+1}\,|g(x)|\,.\end{aligned} \tag{1.5}$$

Therefore the corollary follows from Lemmas 1.3 and 1.8. □

Example 1.10. Let us calculate the Fourier transform of the function $f(x) = \exp(-|x|^2/2)$. First, we consider the function $f_0(t) = \exp(-t^2/2)$ on $\mathbb{R}$. This function is a solution of the differential equation

$$f_0'(t) \ = \ -t\, f_0(t)\,. \tag{1.6}$$

Applying the (one dimensional) Fourier transform to (1.6) and taking into account (1.4), we obtain

$$it\,\hat f_0(t) \;=\; -i\,\hat f_0'(t)\,.$$

where $\hat f_0'$ is the derivative of the Fourier transform $\hat f_0$. Now we see that

$$\left(\frac{\hat f_0(t)}{f_0(t)}\right)' \;=\; \frac{\hat f_0'(t)\,f_0(t)-\hat f_0(t)\,f_0'(t)}{f_0^2(t)} \;=\; \frac{-t\,\hat f_0(t)\,f_0(t)+t\,\hat f_0(t)\,f_0(t)}{f_0^2(t)} \;=\; 0\,,$$

which implies $\hat f_0(t)=c_0\,f_0(t)=c_0\,\exp\left(-t^2/2\right)$ with some constant $c_0\ge 0$. Passing to the polar coordinates, we obtain

$$c_0^2 \;=\; \left(\hat f_0(0)\right)^2 \;=\; (2\pi)^{-1}\left(\int e^{-t^2/2}\,\mathrm{d}t\right)^2 \;=\; (2\pi)^{-1}\iint e^{-(t^2+\tau^2)/2}\,\mathrm{d}t\,d\tau$$
$$=\; (2\pi)^{-1}\int_0^\infty\int_{\mathbf{S}^1} e^{-r^2/2}\,r\,\mathrm{d}\theta\,\mathrm{d}r \;=\; \int_0^\infty e^{-r^2/2}\,r\,\mathrm{d}r \;=\; \frac12\int_0^\infty e^{-s/2}\,\mathrm{d}s \;=\; 1\,,$$

so $\hat f_0(t)=f_0(t)=\exp\left(-t^2/2\right)$. Finally,

$$(2\pi)^{-n/2}\int e^{-ix\cdot\xi}e^{-|x|^2/2}\,\mathrm{d}x=\hat f_0(\xi_1)\,\hat f_0(\xi_2)\dots\hat f_0(\xi_n)=\exp\left(-|\xi|^2/2\right)\,.$$

Thus, $\hat f(\xi)=f(\xi)=\exp\left(-|\xi|^2/2\right)$.

1.4. **Inversion formula.**

Theorem 1.11. *Let $T:\mathcal S(\mathbb R^n)\to\mathcal S(\mathbb R^n)$ be a linear map commuting with multiplication by x_k and differentiation D_{x_k} for all $k=1,\dots,n$, that is,*

$$T(x_kf) \;=\; x_k\,(Tf)\,,\qquad T(D_{x_k}f) \;=\; D_{x_k}(Tf)\,,\forall\,,\quad k=1,2,\dots,n\,, \tag{1.7}$$

for all $f\in\mathcal S(\mathbb R^n)$. Then there exists a constant c such that $Tf=c\,f$ for all $f\in\mathcal S(\mathbb R^n)$.

Proof. Let $f,g\in\mathcal S(\mathbb R^n)$ and $y\in\mathbb R^n$ be a fixed point. If $f(y)=g(y)$ then, by Lemma 1.4,

$$f(x)-g(x) \;=\; \sum_{k=1}^n(x_k-y_k)\,h_k(x)\,,$$

where $h_k\in\mathcal S(\mathbb R^n)$. Now the first identity (1.7) implies that $(Tf)(y)=(Tg)(y)$. Thus, the value of Tf at any point y depends only on the value of f at the point y. Since T is a linear map, this implies that $(Tf)(y)=c(y)\,f(y)$, where $c(y)$ is some constant depending on y.

Since Tf is an infinitely differentiable function for every $f\in\mathcal S(\mathbb R^n)$, the constant $c(y)$ smoothly depends on y. Applying the second identity (1.7), we obtain

$$c(y)\,\partial_{y_k}f(y)=\partial_{y_k}\left(c(y)\,f(y)\right)=f(y)\,\partial_{y_k}c(y)+c(y)\,\partial_{y_k}f(y)$$

for all $f\in\mathcal S(\mathbb R^n)$ and $k=1,2,\dots,n$. Therefore all first derivatives of c are identically equal to zero, which means that c does not depend on y. □

Let $Jf(x):=f(-x)$. Obviously, J is a continuous operator in $\mathcal S(\mathbb R^n)$ and $J\mathcal F=\mathcal FJ$ (the latter is proved by changing variables $\xi=-\eta$ in (1.3)).

Corollary 1.12. *If $f \in \mathcal{S}(\mathbb{R}^n)$ then*

$$f(x) \;=\; J\left(\mathcal{F}_{\xi\to x}\hat{f}(\xi)\right) \;=\; (2\pi)^{-n/2}\int e^{ix\cdot\xi}\,\hat{f}(\xi)\,\mathrm{d}\xi \tag{1.8}$$

Proof. Define $(Tf)(x) := J\left(\mathcal{F}_{\xi\to x}\hat{f}(\xi)\right)$. In view of Corollary 1.9, T is a linear operator from $\mathcal{S}(\mathbb{R}^n)$ into $\mathcal{S}(\mathbb{R}^n)$. Lemma 1.8 implies that

$$T(D_{x_k}f) = J\left(\mathcal{F}_{\xi\to x}(\xi_k\hat{f}(\xi))\right) = J\left(-D_{x_k}\left(\mathcal{F}_{\xi\to x}\hat{f}(\xi)\right)\right) = D_{x_k}(Tf)\,,$$

$$T(x_kf) = J\left(\mathcal{F}_{\xi\to x}(-D_{x_k}\hat{f}(\xi))\right) = J\left(-x_k\left(\mathcal{F}_{\xi\to x}\hat{f}(\xi)\right)\right) = x_k(Tf)\,.$$

Therefore, by Theorem 1.11, there exists a constant c such that $Tf = cf$ for all $f \in \mathcal{S}(\mathbb{R}^n)$. If $f(x) = \exp(-|x|^2/2)$ then

$$c\,f(x) = (Tf)(x) \;=\; J\left(\mathcal{F}_{\xi\to x}\hat{f}(\xi)\right) = f(x)$$

(see Example 1.10). This implies that $c = 1$. □

The linear operator

$$f(x) \to J(\mathcal{F}_{x\to\xi}f(x)) = (2\pi)^{-n/2}\int e^{ix\cdot\xi}\,\hat{f}(x)\,\mathrm{d}x$$

is called the *inverse Fourier transform*. By Corollary 1.9, $J\mathcal{F}f \in \mathcal{S}(\mathbb{R}^n)$ and, by (1.8), $J\mathcal{F}\mathcal{F}f = \mathcal{F}J\mathcal{F}f = f$ whenever $f \in \mathcal{S}(\mathbb{R}^n)$. Thus, we have proved

Theorem 1.13. *The Fourier transform $\mathcal{F}$ is a one-to-one map from $\mathcal{S}(\mathbb{R}^n)$ onto $\mathcal{S}(\mathbb{R}^n)$ and $\mathcal{F}^{-1} = J\mathcal{F} = \mathcal{F}J$.*

Corollary 1.14. (Parseval's formula) *If $f, g \in \mathcal{S}(\mathbb{R}^n)$ then*

$$\int f(x)\,\overline{g(x)}\,\mathrm{d}x \;=\; \int \hat{f}(x)\,\overline{\hat{g}(x)}\,\mathrm{d}x\,.$$

Proof. Parseval's formula follows from (1.8) and the obvious identities

$$\overline{\hat{g}(x)} = \mathcal{F}^{-1}(\bar{g})\,,\qquad \int \hat{f}(x)\,g(x)\,\mathrm{d}x = \int f(x)\,\hat{g}(x)\,\mathrm{d}x\,,\qquad \forall f, g \in \mathcal{S}(\mathbb{R}^n)\,. \tag{1.9}$$

□

2. Tempered distributions

2.1. **Definition and examples.** A map $u:\mathcal{S}(\mathbb{R}^n)\to\mathbb{C}$ is said to be a functional on $\mathcal{S}(\mathbb{R}^n)$. The value of functional u on the function $f\in\mathcal{S}(\mathbb{R})$ is denoted by $\langle u,f\rangle$. We say that the functional u is linear if

$$\langle u, c_1 f + c_2 g\rangle \;=\; c_1\,\langle u,f\rangle \;+\; c_2\,\langle u,g\rangle\,, \qquad \forall f,g\in\mathcal{S}(\mathbb{R}^n)\,,\quad \forall\, c_1,c_2\in\mathbb{C}\,,$$

and u is continuous if $\langle u,f_k\rangle\to\langle u,f\rangle$ whenever $f_k\overset{\mathcal{S}}{\to} f$.

Definition 2.1. A linear continuous functional on $\mathcal{S}(\mathbb{R}^n)$ is said to be a *tempered distribution.*

If u,v are tempered distribution and $c_1,c_2\in\mathbb{C}$, let us define the distribution c_1u+c_2v by

$$\langle c_1u+c_2v,f\rangle \;=\; c_1\,\langle u,\varphi\rangle \;+\; c_2\,\langle v,f\rangle\,, \qquad \forall\, f\in\mathcal{S}(\mathbb{R}^n)\,.$$

Then the set of tempered distributions becomes a linear space. This space is denoted by $\mathcal{S}'(\mathbb{R}^n)$.

Example 2.2. Let u be a polynomially bounded function on $\mathbb{R}^n$. If u is sufficiently nice (continuous, piecewise continuous or, more generally, measurable) then the functional defined by

$$\langle u,f\rangle \;:=\; \int u(x)\,f(x)\,\mathrm{d}x\,, \qquad \forall f\in\mathcal{S}(\mathbb{R}^n)\,, \tag{2.1}$$

is a tempered distribution. This allows us to identify the 'regular' polynomially bounded functions with distributions. Obviously, two functions u_1 and u_2 define the same distribution then $u_1=u_2$ 'almost everywhere' (with respect to the Lebesgue measure). Further on we shall use the same notation u for the function on $\mathbb{R}^n$ and the corresponding distribution.

If u is not polynomially bounded then the integrals on the right hand side of (2.1) may not converge in the usual sense. However, in many cases one can use a suitable regularization of these integrals in order to define a distribution generated by u. This distribution may, of course, depend on the choice of regularization.

Example 2.3. Let $x\in\mathbb{R}^n$ be a fixed point. The tempered distribution δ_x defined by

$$\langle\delta_x,f\rangle \;=\; f(x)\,, \qquad \forall\, f\in\mathcal{S}(\mathbb{R}^n)\,,$$

is said to be the δ-function at x. The δ-function at the origin is usually denoted by δ or $\delta(y)$, where y indicates that δ is considered as a functional on the space of functions depending on the variables y.

Theorem 2.4. *A linear functional u on $\mathcal{S}(\mathbb{R}^n)$ is continuous if and only if there exists a constant C and a non-negative integer p such that*

$$|\langle u,f\rangle| \;\le\; C\sum_{|\alpha|+m\le p}\|f\|_{\alpha,m}\,, \qquad \forall f\in\mathcal{S}(\mathbb{R}^n)\,.$$

Proof. Assume first that there exist constants C and p for which the above estimate holds. If $f_j \overset{\mathcal{S}}{\to} f$ then $\|f - f_j\|_{\alpha,m} \to 0$ for all α, m and, consequently, $\langle u, f - f_j\rangle \to 0$. This implies that u is continuous.

Conversely, let us assume that such constants C and p do not exist. Then there is a sequence of functions $f_j \in \mathcal{S}(\mathbb{R}^n)$ such that

$$|\langle u, f_j\rangle| \;>\; j \sum_{|\alpha|+m\le j} \|f_j\|_{\alpha,m}\,.$$

If $g_j(x) = (\langle u, f_j\rangle)^{-1} f_j(x)$ then $\|g_j\|_{\alpha,m} \le j^{-1}$ for all $j \ge |\alpha| + m$. This implies that $\|g_j\|_{\alpha,m} \to 0$ as $j \to \infty$ for all α and m or, in other words, that $g_j \overset{\mathcal{S}}{\to} 0$. On the other hand, $\langle u, g_j\rangle = 1$ for all j, so u is not a continuous functional on $\mathcal{S}(\mathbb{R}^n)$. □

Exercise 1. Prove that the functional $f \mapsto \lim_{\varepsilon\to 0} \int_{|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t$ belongs to $\mathcal{S}'(\mathbb{R})$.

Exercise 2. Find all distributions $u \in \mathcal{S}'(\mathbb{R})$ such that $\langle u, f\rangle = \int_{-\infty}^{\infty} t^{-1} f(t)\,\mathrm{d}t$ for all functions $f \in \mathcal{S}(\mathbb{R})$ satisfying the condition $f(0) = 0$.

2.2. **Operators in the space of distributions.** Let A be a linear operator acting in the space $\mathcal{S}(\mathbb{R}^n)$.

Condition 2.5. There exists a continuous operator $A^T : \mathcal{S}(\mathbb{R}^n) \mapsto \mathcal{S}(\mathbb{R}^n)$ such that

$$\int (Au)(x)\, v(x)\,\mathrm{d}x \;=\; \int u(x)\,(A^T v)(x)\,\mathrm{d}x\,, \qquad \forall u, v \in \mathcal{S}(\mathbb{R}^n)\,. \tag{2.2}$$

Remark. It is clear from (2.2) that $(A^T)^T = A$

Lemma 2.6. *If Condition* 2.5 *is fulfilled then one can extend* A *from* $\mathcal{S}(\mathbb{R}^n)$ *to the space of distributions* $\mathcal{S}'(\mathbb{R}^n)$.

Proof. If $u \in \mathcal{S}'(\mathbb{R}^n)$, we define Au by

$$\langle Au, v\rangle \;:=\; \langle u, A^T v\rangle\,, \qquad \forall\, v \in \mathcal{S}(\mathbb{R}^n)\,. \tag{2.3}$$

One can easily see that, under Condition 2.5, Au is a tempered distribution. If $u \in \mathcal{S}(\mathbb{R}^n)$ then (2.3) turns into (2.2), so (2.3) defines the same operator A on the space $\mathcal{S}(\mathbb{R}^n)$. □

Lemma 2.6 justifies the following definitions.

Definition 2.7. Let h be an infinitely smooth function on $\mathbb{R}^n$ which is polynomially bounded with all its derivatives. If $u \in \mathcal{S}'(\mathbb{R}^n)$ then hu is the distribution defined by

$$\langle hu, f\rangle \;:=\; \langle u, hf\rangle\,, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n)\,.$$

Definition 2.8. If $u \in \mathcal{S}'(\mathbb{R}^n)$ and α is a multi-index then $\partial_x^\alpha u$ is the distribution defined by

$$\langle \partial_x^\alpha u, f\rangle \;:=\; (-1)^{|\alpha|}\langle u, \partial_x^\alpha f\rangle\,, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n)\,.$$

In the same manner one can define other operators in $\mathcal{S}'(\mathbb{R}^n)$, in particular, the change of variables operator $u(x) \to v(x) = u(\tilde{x}(x))$ where $\tilde{x}(x)$ is a smooth vector function satisfying certain conditions at infinity.

Example 2.9. Let $x \in \mathbb{R}^n$ be a fixed point and $A_x f(y) := f(x-y)$. Then A_x is a continuous operator in $\mathcal{S}(\mathbb{R}^n)$ and $A_x^T = A_x$. If $u(y)$ is a distribution (y indicates that we apply u to functions depending on y) then, by (2.3),

$$\langle u(x-y), f(y)\rangle \;:=\; \langle u(y), f(x-y)\rangle\,, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n)\,.$$

In particular, for the δ-function (see Example 2.3) we have

$$\langle \delta(x-y), f(y)\rangle \;:=\; \langle \delta(y), f(x-y)\rangle \;=\; f(x) \;=\; \langle \delta_x, f\rangle\,, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n)\,,$$

that is, $\delta_x(y) = \delta(x-y)$. In a similar way one can show that $\delta_x(y) = \delta(y-x)$ which implies that $\delta(x-y) = \delta(y-x)$.

Example 2.10. Let $u(t)$ be the characteristic function of the positive half-line. Then, for every $s \in \mathbb{R}$, the derivative of the function $u(t-s)$ coincides with $\delta(t-s)$. For the second derivative of $u(t-s)$ we have

$$\langle u''(t-s), f(t)\rangle \;=\; \langle \delta'(t-s), f(t)\rangle \;=\; -f'(s)\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n)\,.$$

The distribution $\delta'(t-s)$ cannot be described in any simpler way. It is called the derivative of the δ-function at the point s.

Exercise 3. Let $-\infty < a_1 < a_2 < \cdots < a_m < \infty$ and u be a function on $\mathbb{R}$ with the following properties:

(1) u vanishes outside the interval $[a_1, a_m]$;
(2) u is continuously differentiable on every interval (a_k, a_{k+1});
(3) u has finite left and right limits at the points a_k.

Evaluate the derivative $u' \in \mathcal{S}'(\mathbb{R})$ of the function u.

2.3. **Supports of distributions.** Generally speaking, a distribution does not take any particular value at one fixed point. However, two distributions may coincide on an open set.

Definition 2.11. We say that the distribution u vanishes on an open set Ω and write $u|_\Omega = 0$ if $\langle u, f\rangle = 0$ for all $f \in \mathcal{S}(\mathbb{R}^n)$ with $\operatorname{supp} f \subset \Omega$. We say that u coincides with another distribution v on Ω if $(u-v)|_\Omega = 0$.

In particular, the distribution u coincides with a function v on Ω if $\langle u, f\rangle = \int_\Omega v f\,\mathrm{d}x$ whenever $\operatorname{supp} f \subset \Omega$.

Definition 2.12. If $u \in \mathcal{S}'(\mathbb{R}^n)$ then $\operatorname{supp} u := \mathbb{R}^n \setminus \Omega_u$, where Ω_u is the union of all open sets Ω such that $u|_\Omega = 0$.

Example 2.13. The support of any derivative of the δ-function at y coincides with the point y.

The support of a continuous function u coincides with the support of the corresponding distribution (if u is not continuous then this statement is correct modulo a set of Lebesgue's measure zero). If h is a function satisfying conditions of Definition 2.7 then

$$\operatorname{supp}(hu) \;\subset\; (\operatorname{supp} h) \cap (\operatorname{supp} u)\,, \qquad \forall u \in \mathcal{S}'(\mathbb{R}^n)\,.$$

In particular, if $h = 0$ in a neighbourhood of $\operatorname{supp} u$ then $hu = 0$. This is not necessarily true if $h = 0$ only on $\operatorname{supp} u$.

Example 2.14. If $h = 0$ at the origin then $h(x)\delta(x) \equiv 0$. However,

$$\langle h(x)\,\partial_{x_k}\delta(x), f(x)\rangle \;=\; -\,\partial_{x_k}(h(x)f(x))|_{x=0} \;=\; -h_{x_k}(0)\,f(0)\,,$$

that is, $h(x)\,\partial_{x_k}\delta(x) = -(\partial_{x_k}h(0))\,\delta(x)$.

The set of distributions with compact supports is denoted by $\mathcal{E}'(\mathbb{R}^n)$. Theorem 2.4 implies the following result (see [H, Theorem 4.4.7]).

Theorem 2.15. *If $u \in \mathcal{E}'(\mathbb{R}^n)$ then there exists a non-negative integer m such that*

$$u(x) \;=\; \sum_{|\alpha|\le m} \partial_x^\alpha u_\alpha(x)\,, \tag{2.4}$$

where u_α are some continuous functions on $\mathbb{R}^n$.

One can always choose C_0^∞-functions ψ_j such that $\sum_j \psi_j(x) \equiv 1$ (it is called a partition of unity). Then an arbitrary distribution $u \in \mathcal{S}'(\mathbb{R}^n)$ is represented as the sum of distributions $u_j = \psi_j u$ with compact supports, that is, as a sum of distributions of the form (2.4).

2.4. **Fourier transform in $\mathcal{S}'(\mathbb{R}^n)$.** By Corollary 1.9, the Fourier transform $\mathcal{F}$ and the inverse Fourier transform $\mathcal{F}^{-1} = J\mathcal{F}$ are linear continuous operators in $\mathcal{S}(\mathbb{R}^n)$. Obviously, $\mathcal{F}$ and $\mathcal{F}^{-1}$ satisfy Condition 2.5 with $\mathcal{F}^T = \mathcal{F}$ and $(\mathcal{F}^{-1})^T = \mathcal{F}^{-1}$. Therefore, according to Lemma 2.6, the operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ can be extended to $\mathcal{S}'(\mathbb{R}^n)$.

Definition 2.16. If $u \in \mathcal{S}'(\mathbb{R}^n)$ then $\hat{u} = \mathcal{F}u$ and $\mathcal{F}^{-1}u$ are the tempered distributions defined by

$$\langle \mathcal{F}u, f\rangle \;:=\; \langle u, \mathcal{F}f\rangle\,, \quad \langle \mathcal{F}^{-1}u, f\rangle \;:=\; \langle u, \mathcal{F}^{-1}f\rangle\,, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n)\,.$$

Lemma 1.8, Theorem 1.13 and Definition 2.16 immediately imply

Lemma 2.17. *For all $u \in \mathcal{S}'(\mathbb{R}^n)$ we have $\mathcal{F}^{-1}\mathcal{F}u = \mathcal{F}\mathcal{F}^{-1}u = u$ and*

$$\mathcal{F}_{x\to\xi}(D_x^\alpha u) \;=\; \xi^\alpha\,\hat{u}(\xi)\,, \qquad \mathcal{F}_{x\to\xi}(x^\alpha u) \;=\; (-1)^{|\alpha|}\,D_\xi^\alpha \hat{u}(\xi)\,.$$

Example 2.18. Let u be a 'nice' polynomially bounded function (as in Example 2.2). Then

$$\langle \hat{u}, f\rangle = \langle u, \hat{f}\rangle = (2\pi)^{-n/2}\int u(x)\left(\int e^{-ix\cdot\xi}\,f(\xi)\,d\xi\right)\mathrm{d}x\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n)\,.$$

If we can change the order of integration then

$$\langle \hat{u}, f\rangle = (2\pi)^{-n/2}\int\left(\int e^{-ix\cdot\xi}\,u(x)\,\mathrm{d}x\right)f(\xi)\,\mathrm{d}\xi\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n)\,,$$

which implies that $\hat{u}(\xi) = (2\pi)^{-n/2}\int e^{-ix\cdot\xi}\,u(x)\,\mathrm{d}x$. In particular, this formula holds for all functions u from the Lebesgue space $L_1(\mathbb{R}^n)$ (see Remark 1.7).

Example 2.19. If $\delta(x)$ is the δ-function then $\mathcal{F}_{x\to\xi}(D_x^\alpha\delta(x)) = (2\pi)^{-n/2}\,\xi^\alpha$. Indeed,

$$\langle \mathcal{F}_{x\to\xi}(D_x^\alpha\delta(x)), f(\xi)\rangle \;=\; (-1)^{|\alpha|}\langle \delta(x), D_x^\alpha\hat{f}(x)\rangle \;=\; (2\pi)^{-n/2}\int \xi^\alpha\,f(\xi)\,d\xi\,.$$

2.5. **Divergent integrals.** We have defined the Fourier transform for all distributions $u \in \mathcal{S}'(\mathbb{R}^n)$, in particular, for all continuous polynomially bounded functions u. This means, in fact, that we have defined the integral $\int e^{-ix\cdot\xi}\, u(x)\,\mathrm{d}x$ for every such a function u. Of course, this integral may not converge in the classical sense, but can be understood as a distribution in ξ. This idea can be generalized as follows.

Definition 2.20. Let $z \in \mathbb{R}^N$, $\xi \in \mathbb{R}^n$ and $G(z,\xi)$ be a continuous polynomially bounded function on $\mathbb{R}^N \times \mathbb{R}^n$. We shall say that the integral $\int G(z,\xi)\,\mathrm{d}\xi$ converges in the sense of distributions if the consecutive integral $\int \big(\int G(z,\xi)\, f(z)\,\mathrm{d}z\big)\,\mathrm{d}\xi$ converges for every $f \in \mathcal{S}(\mathbb{R}^N)$ and the linear functional $\int G(z,\xi)\, d\xi$ defined by

$$(2.5) \qquad \Big\langle \int G(z,\xi)\,\mathrm{d}\xi\ ,\ f\Big\rangle \ :=\ \int\Big(\int G(z,x)\, f(z)\,\mathrm{d}z\Big)\mathrm{d}x\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^N)\,,$$

belongs to $\mathcal{S}'(\mathbb{R}^N)$.

Considering the integral $\int G(z,\xi)\,\mathrm{d}\xi$ as a distribution, one can operate with it as with an absolutely convergent integral: formally integrate by parts, differentiate under the integral sign, etc. A rigorous justification of all these operations is obtained with the use of Definition 2.20.

Example 2.21. We have $(2\pi)^{-n}\int e^{-ix\cdot\xi}\, d\xi = \delta(x)$. Indeed, if $f \in \mathcal{S}(\mathbb{R}^n)$ then

$$(2\pi)^{-n}\int\left(\int e^{-ix\cdot\xi}\, f(x)\,\mathrm{d}x\right)\mathrm{d}\xi \ =\ (2\pi)^{-n/2}\int \hat f(\xi)\,\mathrm{d}\xi \ =\ \mathcal{F}^{-1}_{\xi\to y}\hat f(\xi)\Big|_{y=0} \ =\ f(0)\,.$$

It may well happen that the distribution $\int G(z,\xi)\,\mathrm{d}\xi$ coincides with a function even if the integral does not converge in the usual sense.

Example 2.22. For all nonzero $z \in \mathbb{C}$ with $\operatorname{Re} z \le 0$ we have

$$(2.6) \qquad \mathcal{F}_{x\to\xi}\exp(z|x|^2/2) \ =\ (2\pi)^{-n/2}\int e^{z|x|^2/2}\, e^{-ix\cdot\xi}\,\mathrm{d}x \ =\ z^{-n/2}\exp(z^{-1}|\xi|^2/2)\,,$$

where $z^{-n/2} = |z|^{-n/2}\exp(-\frac{n}{2}\arg z)$ and $\arg z \in [-\pi/2,\pi/2]$. In particular, for $z = i$,

$$\mathcal{F}_{x\to\xi}\exp(i|x|^2/2) \ =\ (2\pi)^{-n/2}\int e^{i|x|^2/2 - ix\cdot\xi}\,\mathrm{d}x = e^{i\pi n/4}e^{-i|\xi|^2/2}\,.$$

Proof. If $z = 1$ then $(2\pi)^{-n/2}\int e^{-|y|^2/2}\, e^{-iy\cdot\xi}\,\mathrm{d}y = \exp(-|\xi|^2/2)$ (see Example 1.10). Changing variables $y = |z|^{1/2}\,x$, we see that (2.6) holds for all real negative z.

Let us fix an arbitrary complex number z_0 with $\operatorname{Re} z_0 < 0$, substitute $e^{-z|x|^2/2} = e^{-(z-z_0)|x|^2/2}e^{-z_0|x|^2/2}$ and expand the function $e^{-(z-z_0)|x|^2/2}$ into its Taylor series at the point $z = z_0$. Integrating the obtained series term by term, we see that for each fixed $\xi \in \mathbb{R}^n$ the function $\mathcal{F}_{x\to\xi}\exp(-z|x|^2/2)$ is given by an absolutely convergent power series in a neighbourhood of z_0. This implies that $\mathcal{F}_{x\to\xi}\exp(-z|x|^2/2)$ is analytic in the open half-plane $\{z \in \mathbb{C} : \operatorname{Re} z < 0\}$. The function $z^{-n/2}\exp(-z^{-1}|\xi|^2/2)$ is also analytic in this half-plane and, by the above, coincides with $\mathcal{F}_{x\to\xi}\exp(-z|x|^2/2)$ on the negative half-line. Now, from the identity theorem for analytic functions, it follows that (2.6) holds for all z with $\operatorname{Re} z < 0$.

Finally, letting $\operatorname{Re} z \to 0$, we obtain (2.6) for all imaginary numbers $z \ne 0$. □

3. Schwartz kernels, oscillatory integrals and pseudodifferential operators

3.1. Schwartz kernels.

Theorem 3.1. *For every linear continuous operator A in the space $\mathcal{S}(\mathbb{R}^n)$ there exists a family of tempered distributions $\mathcal{A}(x,\cdot)$ depending on the parameter $x\in\mathbb{R}^n$ such that*

$$Av(x) \;=\; \langle\mathcal{A}(x,y),v(y)\rangle\,,\qquad \forall x\in\mathbb{R}^n.$$

Proof. For every $x\in\mathbb{R}^n$ the map $v\to Av(x)$ is a linear continuous functional on $\mathcal{S}(\mathbb{R}^n)$, that is, a tempered distribution which we denote $\mathcal{A}(x,\cdot)$. □

It is clear from the proof that $\mathcal{A}$ is uniquely defined by the operator A.

Definition 3.2. The family of distributions $\mathcal{A}$ is said to be the *Schwartz kernel* of the operator A.

If A is a linear continuous operator in $\mathcal{S}(\mathbb{R}^n)$ then, for every $u\in\mathcal{S}'(\mathbb{R}^n)$, the map

$$v(x)\;\to\;\langle u(x),Av(x)\rangle=\langle u(x),\langle\mathcal{A}(x,y),v(y)\rangle\rangle\,,\qquad v\in\mathcal{S}(\mathbb{R}^n)\,,$$

is a tempered distribution.

Definition 3.3. The linear operator A^T in $\mathcal{S}'(\mathbb{R}^n)$ defined by

$$\langle A^Tu(x),v(x)\rangle \;=\; \langle u(x),Av(x)\rangle\,,\qquad \forall u\in\mathcal{S}'(\mathbb{R}^n)\,,\quad \forall v\in\mathcal{S}(\mathbb{R}^n)\,,$$

is said to be the *transposed* to A.

Now Condition 2.5 be rewritten as follows.

Condition 3.4. The transposed operator A^T continuously maps $\mathcal{S}(\mathbb{R}^n)$ into itself.

If Condition 3.4 is fulfilled then A^T also has a Schwartz kernel $\mathcal{A}^T(x,y)$.

Lemma 3.5. *Let A be a linear continuous operator in $\mathcal{S}(\mathbb{R}^n)$. If its Schwartz kernel $\mathcal{A}$ can be considered as a distribution $\mathcal{A}(\cdot,y)$ smoothly depending on $y\in\mathbb{R}^n$, that is, if there exists a family of distributions $\mathcal{A}(\cdot,y)$ such that $u\mapsto\langle\mathcal{A}(x,y),u(x)\rangle$ is a continuous mapping from $\mathcal{S}(\mathbb{R}^n)$ into itself and*

$$\int\langle\mathcal{A}(x,y),u(x)\rangle v(y)\,\mathrm{d}y \;=\; \int u(x)\,\langle\mathcal{A}(x,y),v(y)\rangle\mathrm{d}x \tag{3.1}$$

for all $u,v\in\mathcal{S}(\mathbb{R}^n)$, then A satisfies Condition 3.4 *and $\mathcal{A}^T(x,y)=\mathcal{A}(y,x)$.*

Proof. The identity (3.1) implies that

$$\langle Bu(x),v(x)\rangle \;=\; \langle u(x),Av(x)\rangle \;=\; \langle A^Tu(x),v(x)\rangle\,,\qquad \forall u,v\in\mathcal{S}(\mathbb{R}^n)\,,$$

where B is the operator in $\mathcal{S}(\mathbb{R}^n)$ given by the Schwartz kernel $\mathcal{B}(x,y)=\mathcal{A}(y,x)$. Therefore $A^Tu=Bu\in\mathcal{S}(\mathbb{R}^n)$ for all $u\in\mathcal{S}(\mathbb{R}^n)$. □

Example 3.6. The δ-function $\delta(x-y)$ (see Example 2.9) can be considered either as a distribution in x depending on the parameter y, or as a distribution in y depending on the parameter x. We have

$$\langle \delta(x-y), f(y)\rangle = f(x)\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n),$$

that is, $\delta(x-y)$ is the Schwartz kernel of the identity operator.

3.2. **Oscillatory integrals.**

Definition 3.7. We say that a function $a(x,y,\xi)$ on $\mathbb{R}^n_x \times \mathbb{R}^n_y \times \mathbb{R}^n_\xi$ belongs to the class S^m if a is infinitely smooth and

$$|\partial_\xi^\alpha \partial_x^\beta \partial_y^\gamma a(x,y,\xi)| \ \le\ \mathrm{const}_{\alpha,\beta,\gamma}\,(1+|\xi|)^{m-|\alpha|}\,.$$

for all multi-indices α,β,γ. We define $S^{-\infty} := \cap_m S^m$, where the intersection is taken over all $m \in \mathbb{R}$.

Obviously, $\xi^{\alpha_1}\partial_\xi^{\alpha_2}\partial_x^\beta\partial_y^\gamma a \in S^{m-|\alpha_2|+|\alpha_1|}$ whenever $a \in S^m$.

Example 3.8. The polynomial $\sum_{|\alpha|\le m} a_\alpha(x,y)\,\xi^\alpha$ with smooth coefficients a_α belongs to S^m if a_α are bounded with all their derivatives.

Definition 3.9. A function $a(x,y,\xi)$ is said to be *positively homogeneous* of degree m in ξ if $a(x,y,\lambda\xi) = \lambda^m a(x,y,\xi)$ for all $\lambda > 0$.

Example 3.10. Let $a(x,y,\xi)$ be a positively homogeneous of degree m function such that

$$|\partial_x^\beta \partial_y^\gamma a(x,y,\xi)| \ \le\ \mathrm{const}_{\beta,\gamma}\,, \qquad \forall \xi : |\xi| = 1\,.$$

Then, for every smooth cut-off function $\zeta(\xi)$ vanishing in a neighbourhood of zero and equal to 1 for large ξ, we have $\zeta a \in S^m$.

Definition 3.11. Let m_k be a sequence of real numbers such that $m_k \to -\infty$ as $k \to \infty$, and let $a_{m-k} \in S^{m_k}$. We say that the function $a \in S^m$ admits an asymptotic expansion

$$a(x,y,\xi) \sim \sum_{k=0}^{\infty} a_{m-k}(x,y,\xi), \qquad |\xi| \to \infty\,, \tag{3.2}$$

if $\left(a - \sum_{k=0}^{l} a_{m-k}\right) \in S^{p_l}$ where $p_l \to -\infty$ as $l \to \infty$ for all $l = 1, 2, \ldots$ We say that a admits the asymptotic expansion (3.2) with a_{m-k} positively homogeneous of degree m_k in ξ if $a \sim \sum_{k=0}^{\infty} \zeta\, a_{m-k}$, where $\zeta = \zeta(\xi)$ is the same cut-off function as in Example 3.10.

Lemma 3.12. *Let m_k be as in Definition* 3.11*, and let $m = \max\{m_k\}$. Then for any sequence of functions $a_{m_k} \in S^{m_k}$ there exists a function $a \in S^m$ such that* (3.2) *holds. This function is determined uniquely modulo $S^{-\infty}$.*

Proof. See [H, Proposition 18.1.3]. □

Definition 3.13. The integral

$$\mathcal{I}_a(x,y) \ =\ (2\pi)^{-n} \int e^{i(x-y)\cdot\xi} a(x,y,\xi)\,\mathrm{d}\xi \tag{3.3}$$

with $a \in S^m$ is called *oscillatory integral* and the function a is called its *amplitude*.

Remark 3.14. One can replace $(x-y)\cdot\xi$ in (3.3) with a more general *phase function* $\varphi(x,y,\xi)$ which has to be positively homogeneous in ξ of degree 1 and non-degenerate in some appropriate sense (see, for example, [Sh], [SV] or [T]).

One can easily see that for every fixed y the integral (3.3) converges in the sense of distributions in x and, the other way round, for every fixed x it converges in the sense of distributions in y. Thus, $\mathcal{I}_a$ can be considered either as a distribution in x depending on the parameter y or as a distribution in y depending on the parameter x. If $m<-n$ then the integral (3.3) is absolutely convergent, so the distribution $\mathcal{I}_a$ coincides with a function. Clearly, this function gets smoother and smoother as $m\to-\infty$; if $a\in S^{-\infty}$ then it is infinitely smooth and bounded with all its derivatives. As a rule, the oscillatory integrals are used for the study of singularities of functions and distributions, and therefore all calculations are carried out modulo $S^{-\infty}$.

3.3. **Pseudodifferential operators.**

Lemma 3.15. *Let $a(x,y,\xi)\in S^m$, and let $\sigma(x,\xi)$ be an arbitrary amplitude from S^m such that*

$$\sigma(x,\xi)\ \sim\ \sum_\alpha \frac{1}{\alpha!}\ D_\xi^\alpha\partial_y^\alpha a(x,y,\xi)\big|_{y=x}\,. \tag{3.4}$$

Then $\mathcal{R}(x,y):=\mathcal{I}_a(x,y)-\mathcal{I}_\sigma(x,y)$ is an infinitely differentiable function on $\mathbb{R}^n_x\times\mathbb{R}^n_y$ such that

$$|\partial_x^\beta\partial_y^\gamma\mathcal{R}(x,y)|\ \le\ \mathrm{const}_{\beta,\gamma,N}\,(1+|x-y|)^{-N} \tag{3.5}$$

for all multi-indices β,γ and positive integers N.

Proof. By Taylor's formula, for all $l=1,2,\dots$, we have

$$a(x,y,\xi)\ =\ \sum_{|\alpha|\le l}\frac{1}{\alpha!}\,(y-x)^\alpha\ \partial_y^\alpha a(x,y,\xi)\big|_{y=x}+\sum_{|\alpha|=l+1}(y-x)^\alpha\,\tilde a_\alpha(x,y,\xi)\,,$$

where

$$\tilde a_\alpha(x,y,\xi)\ =\ \frac{l+1}{\alpha!}\int_0^1(1-t)^l\ \partial_z^\alpha a(x,z,\xi)|_{z=x+t(y-x)}\,\mathrm{d}t\,.$$

If we substitute this expansion into (3.3), replace $(y-x)^\alpha e^{i(x-y)\cdot\xi}$ with $(-1)^{|\alpha|}D_\xi^\alpha e^{i(x-y)\cdot\xi}$ and integrate by parts, then we obtain an oscillatory integral with the amplitude

$$\sum_{|\alpha|\le l}\frac{1}{\alpha!}\ D_\xi^\alpha\partial_y^\alpha a(x,y,\xi)\big|_{y=x}\ +\sum_{|\alpha|=l+1}D_\xi^\alpha\tilde a_\alpha(x,y,\xi)\,d\xi\,. \tag{3.6}$$

One can easily see that the second sum in (3.6) belongs to S^{m-l-1}.

The above arguments show that, for every σ satisfying (3.4) and every positive integer l, the difference $\mathcal{R}(x,y)=\mathcal{I}_a(x,y)-\mathcal{I}_\sigma(x,y)$ can be represented by the oscillatory integral

$$\int e^{i(x-y)\cdot\xi}b_l(x,y,\xi)\,\mathrm{d}\xi$$

with an amplitude $b_l \in S^{m-l-1}$. If $l \geqslant |\beta| + |\gamma| + m + n$ then

$$\begin{aligned}(3.7)\quad \partial_x^\beta \partial_y^\gamma \left((x-y)^\alpha \int e^{i(x-y)\cdot\xi} b_l(x,y,\xi)\,\mathrm{d}\xi \right) &= \partial_x^\beta \partial_y^\gamma \int (D_\xi^\alpha e^{i(x-y)\cdot\xi})\, b_l(x,y,\xi)\,\mathrm{d}\xi \\ &= (-1)^{|\alpha|} \int \partial_x^\beta \partial_y^\gamma \left(e^{i(x-y)\cdot\xi} D_\xi^\alpha b_l(x,y,\xi)\right) \mathrm{d}\xi\end{aligned}$$

and the integral on the right hand side is absolutely convergent and bounded uniformly with respect to x and y. Since l can be chosen arbitrarily large, it follows that

$$\left|\partial_x^\beta \partial_y^\gamma \left((x-y)^\alpha \mathcal{R}(x,y)\right)\right| \leqslant \mathrm{const}_{\alpha,\beta,\gamma}, \qquad \forall \alpha,\beta,\gamma\,.$$

This implies (3.5). □

In a similar way one can prove that $\mathcal{I}_a(x,y) - \mathcal{I}_{\sigma'}(x,y)$ is an infinitely smooth function satisfying (3.5) if

$$(3.8)\qquad \sigma' = \sigma'(y,\xi) \sim \sum_\alpha \frac{1}{\alpha!}\, (-1)^{|\alpha|} D_\xi^\alpha \partial_x^\alpha a(x,y,\xi)\big|_{x=y}\,.$$

Thus, the distribution $\mathcal{I}_a(x,y)$ can be represented, modulo a smooth function satisfying (3.5), by the oscillatory integral with an amplitude independent either of y or of x.

Definition 3.16. We say that an operator A belongs to the class Ψ^m if its Schwartz kernel is given by an oscillatory integral $\mathcal{I}_a(x,y)$ with some amplitude $a \in S^m$. The operator $A \in \Psi^m$ is said to be a *pseudodifferential operator* (PDO) of order m, and the functions σ and σ' satisfying (3.4) and (3.8) are said to be its *symbol* and *dual symbol* respectively.

Lemma 3.15 and (3.8) imply that

$$(3.9)\qquad \begin{aligned}\sigma'(y,\xi) &\sim \sum_\alpha \frac{1}{\alpha!}\, (-1)^{|\alpha|} D_\xi^\alpha \partial_y^\alpha \sigma(y,\xi)\,, \\ \sigma(x,\xi) &\sim \sum_\alpha \frac{1}{\alpha!}\, D_\xi^\alpha \partial_x^\alpha \sigma'(x,\xi)\,.\end{aligned}$$

Clearly, $\Psi^m \subset \Psi^l$ whenever $m \leqslant l$. We shall denote $\Psi^{-\infty} := \bigcap_{m\in\mathbb{R}} \Psi^m$.

Lemma 3.17. *An operator $R : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ belongs to $\Psi^{-\infty}$ if and only if its Schwartz kernel is an infinitely smooth function satisfying* (3.5)*.*

Exercise 4. Prove Lemma 3.17. **Hint:** *the Schwartz kernel of an operator from $\Psi^{-\infty}$ can be estimated with the use of* (3.7)*.*

Lemma 3.18. *A PDO $A \in \Psi^m$ continuously maps $\mathcal{S}(\mathbb{R}^n)$ into itself.*

Proof. According to Definitions 2.20, 3.2 and 3.16, if $A \in \Psi^m$ then

$$Au(x) = (2\pi)^{-n} \int e^{ix\cdot\xi} \left(\int e^{-iy\cdot\xi} a(x,y,\xi)\, u(y)\, dy \right) \mathrm{d}\xi$$

with some $a \in S^m$. If $u \in \mathcal{S}(\mathbb{R}^n)$ then the integral with respect to y absolutely converges for each fixed x, ξ and defines a smooth function of (x,ξ) rapidly decreasing with respect to ξ

with all its derivatives. The same is true for all integrals obtained by formal differentiation. We have

$$
\begin{aligned}
(3.10)\quad |Au(x)| \;&\le\; \int |\int e^{-iy\cdot\xi} a(x,y,\xi)\,u(y)\,\mathrm{d}y\,|\,\mathrm{d}\xi\\
&= \int |\int \big((1-\Delta_y)^N e^{-iy\cdot\xi}\big)\,(1+|\xi|^2)^{-N} a(x,y,\xi)\,u(y)\,\mathrm{d}y\,|\,\mathrm{d}\xi\\
&= \int |\int e^{-iy\cdot\xi}(1+|\xi|^2)^{-N}(1-\Delta_y)^N\,(a(x,y,\xi)\,u(y))\,\mathrm{d}y\,|\,\mathrm{d}\xi\\
&\le \iint (1+|\xi|^2)^{-N}|(1-\Delta_y)^N\,(a(x,y,\xi)\,u(y))\,|\mathrm{d}y\,\mathrm{d}\xi
\end{aligned}
$$

for all positive integers N, where $\Delta_y := \sum_k \partial_{y_k}^2$. The integral on the right hand side converges and is estimated by a finite linear combination of $\|u\|_{\alpha,\beta}$. Differentiating under the integral sign and integrating by parts, we see that

$$
\begin{aligned}
(3.11)\quad \partial_{x_k} Au(x) \;&=\; \int e^{ix\cdot\xi}\Big(\int e^{-iy\cdot\xi}(i\xi_k)\,a(x,y,\xi)\,u(y)\,\mathrm{d}y\Big)\mathrm{d}\xi\\
&\quad + \int e^{ix\cdot\xi}\Big(\int e^{-iy\cdot\xi}\partial_{x_k} a(x,y,\xi)\,u(y)\,\mathrm{d}y\Big)\mathrm{d}\xi\,,
\end{aligned}
$$

$$
\begin{aligned}
(3.12)\quad x_k\,Au(x) \;&=\; \int (D_{\xi_k} e^{ix\cdot\xi})\Big(\int e^{-iy\cdot\xi} a(x,y,\xi)\,u(y)\,\mathrm{d}y\Big)\mathrm{d}\xi\\
&= \int e^{ix\cdot\xi}\Big(\int e^{-iy\cdot\xi} a(x,y,\xi)\,y_k\,u(y)\,\mathrm{d}y\Big)\mathrm{d}\xi - \int e^{ix\cdot\xi}\Big(\int e^{-iy\cdot\xi}\partial_{\xi_k} a(x,y,\xi)\,u(y)\,\mathrm{d}y\Big)\mathrm{d}\xi\,.
\end{aligned}
$$

This implies that $x^\beta \partial_x^\alpha (Au)(x)$ coincides with a finite sum of integrals of the same type as $Au(x)$ and therefore is also estimated by a finite linear combination of $\|u\|_{\alpha',\beta'}$. □

By Lemma 3.5, if A is a PDO with an amplitude $a(x,y,\xi)$ then its transposed A^T is a PDO with the amplitude $a(y,x,-\xi)$. Therefore, in view of Lemma 2.6, a PDO can be extended to the space $\mathcal{S}'(\mathbb{R}^n)$.

Example 3.19. Every differential operator with smooth coefficients bounded with all their derivatives is a PDO. Indeed, if the Schwartz kernel of A is given by the oscillatory integral with an amplitude $a(x,y,\xi) = \sum_{|\alpha|\le m} a_\alpha(x,y)\,\xi^\alpha$ then, by Lemma 2.17,

$$
Au(x) \;=\; \sum_{|\alpha|\le m} D_y^\alpha\,(a_\alpha(x,y)\,u(y))\big|_{y=x}\,.
$$

In particular, if A is the PDO with symbol $\sigma(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)\,\xi^\alpha$ then

$$
Au(x) \;=\; \sum_{|\alpha|\le m} a_\alpha(x) D_x^\alpha\,u(x)\,,
$$

and if A is the PDO with dual symbol $\sigma(y,\xi)=\sum_{|\alpha|\le m} a_\alpha(y)\,\xi^\alpha$ then

$$Au(x) \;=\; \sum_{|\alpha|\le m} D_x^\alpha\,(a_\alpha(x)\,u(x))\,.$$

This explains the role of the factor $(2\pi)^{-n}$ appearing in (3.3).

Example 3.20. If A is a PDO with symbol $\sigma_A(x,\xi)$ then, by Lemma 3.15, the symbol σ_{A^T} of the transposed operator A^T admits the asymptotic expansion

$$\sigma_{A^T}(x,\xi) \;\sim\; \sum_\alpha \frac{1}{\alpha!}\,D_\xi^\alpha\partial_x^\alpha\sigma(x,-\xi)\,. \tag{3.13}$$

Remark 3.21. In the process of proving Lemmas 3.15, we have shown that $\mathcal{I}_a=\mathcal{I}_b$ where b is the amplitude given by (3.6). Thus two different amplitudes may define the same oscillatory integral. In particular, it may well happen that $\mathcal{I}_a=0$ but $a\ne0$, and even $a\notin S^{-\infty}$.

On the other hand, the symbol σ of the PDO A with Schwartz kernel $\mathcal{I}_a$ is defined uniquely modulo $S^{-\infty}$. Indeed, if $\tilde A$ is the PDO defined by the oscillatory integral $\mathcal{I}_\sigma$ then, by Lemmas 3.15 and 3.17, $R=A-\tilde A$ is an integral operator with smooth kernel $\mathcal{R}(x,y)$ satisfying (3.5).

Let us fix $\eta\in\mathbb{R}^n$ and consider the smooth function $u_\eta(y)=e^{iy\cdot\eta}$. Using (3.5), one can easily show that the function

$$r(x,\eta) \;:=\; e^{-ix\cdot\eta}\,Ru_\eta(x) \;=\; \int \mathcal{R}(x,y)\,e^{-i(x-y)\cdot\eta}\,\mathrm{d}y$$

lies in $S^{-\infty}$. Thus we have $e^{-ix\cdot\eta}\,Au_\eta=e^{-ix\cdot\eta}\,\tilde Au_\eta$ modulo $S^{-\infty}$. It remains to notice that

$$\begin{aligned}\tilde Au_\eta(x) \;&=\; (2\pi)^{-n}\int e^{i(x-y)\cdot\xi}\sigma(x,\xi)\,e^{iy\cdot\eta}\,\mathrm{d}y\,\mathrm{d}\xi\\ &=\; (2\pi)^{-n}\int e^{ix\cdot\xi}\sigma(x,\xi)\,e^{iy\cdot(\eta-\xi)}\,\mathrm{d}y\,\mathrm{d}\xi\\ &=\; \langle\delta(\eta-\xi),e^{ix\cdot\xi}\sigma(x,\xi)\rangle \;=\; e^{ix\cdot\eta}\sigma(x,\eta)\,,\end{aligned}$$

so that $\sigma(x,\eta)=e^{-ix\cdot\eta}Au_\eta(x)$ modulo $S^{-\infty}$.

3.4. **Other classes of PDOs.** Lemma 3.15 plays the key role in the theory of PDOs. One can consider much more general classes of amplitudes and the corresponding classes of PDOs (see, for example [H]), and usually all classical results remain valid as far as an analogue of Lemma 3.15 holds. For example, given a ‘weight’ functions $g(x,y,\xi)$ and two real numbers $\rho,\delta\in[0,1]$, we can consider the classes $S^{m,g}_{\rho,\delta}$ which consist of amplitudes satisfying the estimates

$$|\partial_\xi^\alpha\partial_x^\beta\partial_y^\gamma a(x,y,\xi)| \;\le\; \mathrm{const}_{\alpha,\beta,\gamma}\,(g(x,y,\xi))^{m-\rho|\alpha|+\delta|\beta|+\delta|\gamma|}\,. \tag{3.14}$$

These classes are more convenient than S^m if we want to control not only the smoothness properties of functions but also their behaviour at infinity.

One often has to deal with differential operators depending on an additional parameter h (for instance, the semi-classical parameter). In this case one can introduce a weight function g depending on this parameter and use the classes of amplitudes defined by (3.14) in order to study asymptotics with respect to h (see, for example, [DS]).

4. Solution of partial differential equations

Example 4.1. Consider the differential equation $t^2u'(t) = 0$ on the real line. Obviously, this equation does not have any classical solution apart from $u \equiv \text{const}$. However, if we rewrite this equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(t^2u(t)\right) - 2t\,u(t) = 0$$

then we see that any function of the form

$$u(t) = \begin{cases} c_1\,, & t \ge 0\,, \\ c_2\,, & t < 0\,, \end{cases} \tag{4.1}$$

where c_1, c_2 are constants, is also a solution.

This example shows that if we are looking only for classical solutions then the class of solutions may depend on the way we write down the equation. This problem does not arise if we understand solutions in the sense of distribution. For instance, the derivative of the function (4.1) is the δ-function multiplied by some constant, so $t\,u(t)$ is equal to 0 as a distribution.

4.1. Differential equations with constant coefficients.

Let $a(\xi) := \sum_{|\alpha|\le m} c_\alpha\,\xi^\alpha$ be a polynomial with constant coefficients c_α and $A = a(D_x) = \sum_{|\alpha|\le m} c_\alpha\,D_x^\alpha$ be the differential operator with symbol $a(\xi)$ (see Example 3.19). Then, by Lemma 2.17,

$$Au(x) \;=\; \mathcal{F}^{-1}_{\xi\to x}a(\xi)\mathcal{F}_{y\to\xi}u(y)\,, \qquad \forall u \in \mathcal{S}'(\mathbb{R}^n)\,.$$

Therefore $Au = f$ if and only if

$$a(\xi)\,\hat{u}(\xi) \;=\; \hat{f}(\xi)\,. \tag{4.2}$$

Thus, in order to solve partial differential equation $Au = f$ with $f \in \mathcal{S}'(\mathbb{R}^n)$ it is sufficient to solve the algebraic equation (4.2).

Example 4.2. If $(a(\xi))^{-1}$ is a polynomially bounded continuous function then the equation $Au = f$ has the only solution $u(x) = \mathcal{F}^{-1}_{\xi\to x}(a(\xi))^{-1}f(\xi) \in \mathcal{S}'(\mathbb{R}^n)$ for every $f \in \mathcal{S}(\mathbb{R}^n)$.

Example 4.3. If $(a(\xi))^{-1}$ is an infinitely differentiable function polynomially bounded with all its derivatives then the equation $Au = f$ has the only solution $u(x) = \mathcal{F}^{-1}_{\xi\to x}(a(\xi))^{-1}f(\xi) \in \mathcal{S}'(\mathbb{R}^n)$ for every $f \in \mathcal{S}'(\mathbb{R}^n)$.

Example 4.4. If $u \in \mathcal{S}'(\mathbb{R}^n)$ and $Au = 0$ then necessarily

$$\operatorname{supp}\hat{u} \;\subset\; \Sigma_a := \{\xi : a(\xi) = 0\}\,.$$

Every distribution $u \in \mathcal{S}'(\mathbb{R}^n)$ whose Fourier transform is given by

$$\langle \hat{u}(\xi), f(\xi)\rangle \;=\; \int_{\Sigma_a} v(\xi)\,f(\xi)\,\mathrm{d}\Sigma_a(\xi)\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n)\,, \tag{4.3}$$

where $\mathrm{d}\Sigma_a$ is an arbitrary measure and v is an arbitrary integrable function on Σ_a, solves the equation $Au = 0$. If (4.3) holds true and the function v is 'sufficiently nice' then

$$u(x) \;=\; (2\pi)^{-n/2} \int_{\Sigma_a} e^{ix\xi} v(\xi)\, \mathrm{d}\Sigma_a(\xi)$$

is a function on $\mathbb{R}^n$.

A comprehensive exposition of the theory of partial differential equations with constant coefficients can be found in [H, Volume II].

4.2. **Non-stationary equations with constant coefficients.** If the operator includes the time variable t and we want to solve the Cauchy problem then it is usually more convenient to consider the Fourier transform only with respect to the spatial variables. If, for example,

$$A(\partial_t, D_x) \;= \partial_t^m + \sum_{|\alpha|=1} c_{m-1,\alpha}\, D_x^\alpha \partial_t^{m-1} + \sum_{|\alpha|=2} c_{m-2,\alpha}\, D_x^\alpha \partial_t^{m-2} + \cdots + \sum_{|\alpha|=m} c_{0,\alpha}\, D_x^\alpha$$

where $c_{k,\alpha}$ are some constants, then $u(x,t)$ solves the Cauchy problem

$$A(\partial_t, D_x) u(t,x) = 0\,, \qquad \partial_t^k u(t,x)\big|_{t=0} = v_k(x)\,, \quad k = 0, 1, \ldots, m,$$

if and only if $\hat{u}(t,\xi) = \mathcal{F}_{x\to\xi} u(t,x)$ is a solution of the ordinary differential equation

$$A(\partial_t, \xi)\hat{u}(t,\xi) = 0\,, \qquad \partial_t^k \hat{u}(t,\xi)\big|_{t=0} = \hat{v}_k(\xi)\,, \quad k = 0, 1, \ldots, m,$$

where $A(\partial_t, \xi) = \partial_t^m + \sum_{|\alpha|=1} c_{m-1,\alpha}\, \xi^\alpha \partial_t^{m-1} + \cdots + \sum_{|\alpha|=m} c_{0,\alpha}\, \xi^\alpha$. In this case we understand $u(t,x)$ as a family of distributions in x depending on the parameter t, and $\partial_t^k u$ is the family of distributions such that

$$\langle \partial_t^k u(t,x), f(x)\rangle \;=\; \partial_t^k \langle u(t,x), f(x)\rangle\,, \qquad \forall f \in \mathcal{S}(\mathbb{R}^n)\,.$$

Example 4.5. (Heat equation.) Let $a(x)$ be a semibounded from below polynomial on $\mathbb{R}^n$ and $A = a(D_x)$. Then, for every $v \in \mathcal{S}'(\mathbb{R}^n)$ the distribution

$$u(t,x) \;=\; \mathcal{F}^{-1}_{\xi\to x} e^{-ta(\xi)} \hat{v}(\xi)$$

is the only solution of the Cauchy problem

$$\partial_t u + Au = 0\,, \qquad u(0,x) = v(x)\,.$$

If $v \in \mathcal{S}(\mathbb{R}^n)$ then, obviously, $u(t,\cdot) \in \mathcal{S}(\mathbb{R}^n)$ for every t, and the Schwartz kernel of the operator $\exp(-tA) : v(x) \mapsto u(t,x)$ (the so-called heat kernel) is given by the integral

$$(2\pi)^{-n} \int e^{i(x-y)\cdot\xi} e^{-ta(\xi)}\, \mathrm{d}\xi$$

which converges in the sense of distributions.

Example 4.6. (Wave equation.) If $v \in \mathcal{S}'(\mathbb{R}^n)$ then the distribution

$$u(t,x) \;=\; \mathcal{F}^{-1}_{\xi\to x} \cos(t|\xi|)\, \hat{v}(\xi)$$

is the only solution of the Cauchy problem

$$\partial_t^2 u - \Delta u = 0\,, \qquad u(0,x) = v(x)\,, \quad \partial_t u(0,x) = 0\,,$$

where $\Delta = \sum_k \partial^2_{x_k}$ is the Laplacian. As in the previous example, this implies that $u(t,\cdot) \in \mathcal{S}(\mathbb{R}^n)$ for every t whenever $v \in \mathcal{S}(\mathbb{R}^n)$.

4.3. **Elliptic (pseudo)differential equations.**

Lemma 4.7. (Composition of PDOs.) *If $A \in \Psi^{m_1}$ and $B \in \Psi^{m_2}$ then $AB \in \Psi^{m_1+m_2}$ and the symbol σ_{AB} of the PDO AB is given by the asymptotic series*

$$\sigma_{AB}(x,\xi) \;\sim\; \sum_\alpha \frac{1}{\alpha!} D_\xi^\alpha \sigma_A(x,\xi)\, \partial_x^\alpha \sigma_B(x,\xi) \tag{4.4}$$

where σ_A and σ_B are the symbols of A and B respectively.

Proof. If A and B are given by the oscillatory integrals (3.3) with $\sigma_A(x,\xi)$ and $\sigma'_B(y,\xi)$ respectively then

$$Av(x) = (2\pi)^{-n} \int e^{i(x-y)\cdot\xi}\, \sigma_A(x,\xi)\, v(y)\, \mathrm{d}y\, \mathrm{d}\xi = (2\pi)^{-n/2} \int e^{ix\cdot\xi}\, \sigma_A(x,\xi)\, \hat v(\xi)\, \mathrm{d}\xi$$

and

$$Bu(x) = (2\pi)^{-n} \int e^{i(x-y)\cdot\xi}\, \sigma'_B(y,\xi)\, u(y)\, \mathrm{d}y\, \mathrm{d}\xi = \mathcal{F}^{-1}_{\xi\to x}\Big((2\pi)^{-n/2} \int e^{-iy\cdot\xi}\, \sigma'_B(y,\xi)\, u(y)\, \mathrm{d}y\Big).$$

Therefore

$$ABu(x) \;=\; (2\pi)^{-n} \int e^{ix\cdot\xi}\, \sigma_A(x,\xi) \Big(\int e^{-iy\cdot\xi}\, \sigma'_B(y,\xi)\, u(y)\, \mathrm{d}y\Big) \mathrm{d}\xi\,,$$

that is, the Schwartz kernel of AB coincides with the oscillatory integral with the amplitude $\sigma_A(x,\xi)\,\sigma'_B(y,\xi)$. By (3.4), we have

$$\sigma_{AB}(x,\xi) \;\sim\; \sum_\alpha \frac{1}{\alpha!} D_\xi^\alpha \left(\sigma_A(x,\xi)\, \partial_x^\alpha \sigma'_B(x,\xi)\right).$$

Now (4.4) is obtained by substituting the first expansion (3.9) and rearranging terms in the asymptotic series. □

Definition 4.8. A PDO $A \in \Psi^m$ is said to be *classical* if its symbol admits an asymptotic expansion into the series (3.2) with a_{m-k} positively homogeneous in ξ of degree $m-k$. The spaces of classical PDOs $A \in \Psi^m$ and their symbols are denoted by Ψ^m_{cl} and S^m_{cl}, respectively.

Obviously, every differential operator is a classical PDO.

Definition 4.9. If $A \in \Psi^m_{\mathrm{cl}}$ then the leading homogeneous term a_m in the expansion of σ_A is said to be the *principal symbol* of the operator A. The operator A is said to be elliptic if $a_m(x,\xi) \neq 0$ whenever $\xi \neq 0$.

Definition 4.10. The operator B is said to be a *left parametrix* of A if $BA - I \in \Psi^{-\infty}$.

If B is a left parametrix of A and $Au = f$ then $(I+R)u = Bf$ where $R \in \Psi^{-\infty}$. The 'remainder' operator R often turns out to be compact in a suitable function space H. In this case the existence of a parametrix implies that the subspace $\{u \in H : Au = 0\}$ is of finite dimension, and that the equation $Au = f$ has a solution for all f from a subspace of finite codimension.

Theorem 4.11. *Every elliptic classical PDO $A \in \Psi^m_{\mathrm{cl}}$ has a left parametrix $B \in \Psi^{-m}_{\mathrm{cl}}$.*

Proof. We shall construct the symbol of B as an asymptotic series of positively homogeneous in ξ functions $b_{-m-k}(x,\xi)$ of degree $-m-k$. If we substitute the formal series $\sigma_A = \sum_{k=0}^m a_{m-k}$ and $\sigma_B = \sum_{k=0}^\infty b_{-m-k}$ into (4.4), collect together homogeneous terms of the same degree, equate the first term to 1 and others to zero, then we obtain a recurrent system of differential equations of the form

$$
\begin{aligned}
a_m\, b_{-m} &= 1\,,\\
a_m\, b_{-m-1} &= L_1(a_m, a_{m-1}, b_{-m})\,,\\
\dots &= \dots\\
a_m\, b_{-m-k} &= L_k(a_m, a_{m-1}\dots, a_{m-k}, b_{-m}, b_{-m-1}, \dots, b_{-m-k+1})\,,\\
\dots &= \dots
\end{aligned}
$$

where $L_k(a_m, a_{m-1}\dots, a_{m-k}, b_{-m}, b_{-m-1}, \dots, b_{-m-k+1})$ are some polynomials of the functions $a_m, a_{m-1}\dots, a_{m-k}, b_{-m}, b_{-m-1}, \dots, b_{-m-k+1}$ and their derivatives. If $b_{-m} = a_m^{-1}$,

$$
b_{-m-k} = a_m^{-1}\, L_k(a_m, \dots, a_{m-k}, b_{-m}, \dots, b_{-m-k+1})\,, \qquad k = 1, 2, \dots,
$$

and $\sigma_B \sim \sum_{k=0}^\infty b_{-m-k}$ then $\sigma_{AB} = 1$ modulo $S^{-\infty}$, that is, $AB - I \in \Psi^{-\infty}$. □

Exercise 5. Prove that

(1) $A \in \Psi^m_{\mathrm{cl}}$ if and only if $A^T \in \Psi^m_{\mathrm{cl}}$, where A^T is the transposed operator (see Definition 3.3);
(2) A is elliptic if and only if A^T is elliptic;
(3) if $A \in \Psi^m_{\mathrm{cl}}$ is elliptic then there exists an operator $B \in \Psi^{-m}_{\mathrm{cl}}$, called a *right parametrix* of A, such that $AB - I \in \Psi^{-\infty}$.

Hint: deduce (3) *from* (1), (2) *and Theorem* 4.11.

Remark 4.12. Let $A \in \Psi^m_{\mathrm{cl}}$ be a classical PDO and $\mathcal{O} \subset \mathbb{R}^n_x \times (\mathbb{R}^n_\xi \setminus \{0\})$ be a conic with respect to ξ subset (the word conic means that $(x, \lambda\,\xi) \in \mathcal{O}$ for all $\lambda > 0$ whenever $(x,\xi) \in \mathcal{O}$). If the principal symbol of A is separated from 0 on the set $\mathcal{O} \cap \{|\xi| = 1\}$ then, exactly in the same way, one can construct a PDO $B \in \Psi^{-m}_{\mathrm{cl}}$ such that $\sigma_{BA} = 1$ on $\mathcal{O}$. Such an operator is called a *microlocal parametrix* of A in $\mathcal{O}$.

4.4. **General partial differential equations with variable coefficients.** An arbitrary partial differential operator does not necessarily have a pseudodifferential parametrix. However, quite often one can construct a parametrix in the form of a general oscillatory integral (Remark 3.14) or a PDO which belongs to a more general class (Subsection 3.4). The procedure remains almost the same as in the proof Theorem 4.11: we formally replace the amplitude with an asymptotic series, substitute the integral into the equation, get rid of the variable y in the new amplitude (like we did in the proof of Lemma 3.15), collect together the terms of the same order, equate them to zero and try to solve these equations. Note that in the general case the equations may also involve the unknown phase function, and that the terms in the asymptotic expansions may not be homogeneous in ξ (in which case the words 'terms of the same order' simply mean that these terms satisfy estimates of the form (3.14) with the same m, ρ and δ).

5. Singularities of functions and distributions

5.1. **What is microlocal analysis.** Suppose that we want to describe singularities of a function $f(x)$ on $\mathbb{R}^n$. In classical analysis one only deals with the variables x, and the typical statements look like *"the function f has a singularity at the point x_0"* or *"f is smooth in a neighbourhood of x_0"*. However, the function may well be smooth in one direction and non-smooth in another direction, so such statements contain a limited information. More detailed description of singularities should involve additional variables ξ specifying the directions in which the function is not smooth. In other words, the set of singularities should be a subset of $\mathbb{R}^n_x \times \mathbb{R}^n_\xi$, and then we say that *"f is not smooth at the point $(x,\xi) \in \mathbb{R}^n_x \times \mathbb{R}^n_\xi$"* if f is not smooth at the point x in the direction ξ. This is the main idea of microlocal analysis; the word 'microlocal' simply means that we conduct analysis of functions in the space $\mathbb{R}^n_x \times \mathbb{R}^n_\xi$ of dimension $2n$, even though the functions themselves are defined on the n-dimensional space.

5.2. **Singular supports and wave front sets.**

Definition 5.1. If $u \in \mathcal{S}'(\mathbb{R}^n)$ then the *singular support* of u is defined by

$$\operatorname{sing\,supp} u \;:=\; \mathbb{R}^n \setminus \Omega_u,$$

where Ω_u is the union of all open sets Ω such that $u|_\Omega \in C^\infty(\Omega)$. In other words, Ω_u is the maximal open subset of $\mathbb{R}^n$ such that $u|_{\Omega_u} \in C^\infty(\Omega_u)$.

Definition 5.2. If $u \in \mathcal{S}'(\mathbb{R}^n)$ then the *wave front set* of u is defined by

$$\operatorname{WF} u \;:=\; (\mathbb{R}^n_x \times \mathbb{R}^n_\xi) \setminus \mathcal{O}_u$$

where $\mathcal{O}_u$ is the maximal open subset of $\mathbb{R}^n_x \times \mathbb{R}^n_\xi$ with following property: for every point $(x_0,\xi_0) \in \mathcal{O}_u$ there exist a cut-off function $\chi \in C_0^\infty(\mathbb{R}^n)$ equal to 1 in a neighbourhood of x_0 and a conic neighbourhood Ω_{ξ_0} of ξ_0 such that the Fourier transform $\mathcal{F}_{x\to\xi}\,(\chi(x)u(x))$ decays faster than any negative power of $|\xi|$ in Ω_{ξ_0} as $|\xi| \to \infty$.

If $(x_0,\xi_0) \in \mathcal{O}_u$ then $(x_0,\lambda\xi_0) \in \mathcal{O}_u$ for all $\lambda > 0$ because ξ_0 and $\lambda\xi_0$ have the same conic neighbourhoods. It follows that the sets $\mathcal{O}_u$ and $\operatorname{WF} u$ are invariant under the transformations $(x,\xi) \mapsto (x,\lambda\xi)$ for all $\lambda > 0$.

Definition 5.2 can be rewritten as follows.

Definition 5.3. Denote by $Q_{a,\chi}$ the PDO with dual symbol $a(\xi)\,\chi(y)$. The point (x_0,ξ_0) does not belong to $\operatorname{WF} u$ if there exist a C_0^∞-function χ equal to 1 in a neighbourhood of x_0 and a function $a \in S^m_{\mathrm{cl}}$ equal to 1 in a conic neighbourhood of ξ_0 such that $Q_{a,\chi}u \in \mathcal{S}(\mathbb{R}^n)$.

Indeed, if $\mathcal{F}_{x\to\xi}\,(\chi\,u)$ decays faster than any negative power of $|\xi|$ in Ω_{ξ_0} then

$$Q_{a,\chi}u(x) = \mathcal{F}^{-1}_{\xi\to x}\,(a(\xi)\,\mathcal{F}_{x\to\xi}\,(\chi\,u)) \in \mathcal{S}(\mathbb{R}^n)$$

for any $a \in S^m_{\mathrm{cl}}$ with $\operatorname{supp} a \in \Omega_{\xi_0}$. Conversely, if $Q_{a,\chi}u \in \mathcal{S}(\mathbb{R}^n)$ then the function

$$\mathcal{F}_{x\to\xi}\,(Q_{a,\chi}u) = a(\xi)\,\mathcal{F}_{x\to\xi}\,(\chi\,u)$$

is rapidly decreasing, which implies that $\mathcal{F}_{x\to\xi}\,(\chi\,u)$ decays faster than any negative power of $|\xi|$ in the conic neighbourhood Ω_{ξ_0} where $a = 1$.

Note that

(S) a compactly supported distribution u coincides with an infinitely smooth function if and only if its Fourier transform $\hat u(\xi)$ decays faster than any negative power of $|\xi|$ as $|\xi|\to\infty$ everywhere.

Indeed, if $u\in C^\infty(\mathbb{R}^n)\cap\mathcal{E}'(\mathbb{R}^n)$ then $u\in\mathcal{S}(\mathbb{R}^n)$ and $\hat u\in\mathcal{S}(\mathbb{R}^n)$; if $\hat u(\xi)$ satisfies the above condition then

$$u(x)\ =\ (2\pi)^{-n/2}\int e^{ix\cdot\xi}\,\hat u(\xi)\,\mathrm{d}\xi\ \in\ C^\infty(\mathbb{R}^n)$$

since we can differentiate under the integral sign infinitely many times. From **(S)** it follows that the projection of $\mathrm{WF}\,u$ onto $\mathbb{R}^n_x$ coincides with $\operatorname{sing\,supp} u$.

5.3. **Operators $R\in\Psi^{-\infty}$ in the space of distributions.**

Lemma 5.4. *Let $R\in\Psi^{-\infty}$ and $\mathcal{R}(x,y)$ be the Schwartz kernel of R. Then*

$$Ru(x)\ =\ \langle u(y),\mathcal{R}(x,y)\rangle\,,\qquad \forall u\in\mathcal{S}'(\mathbb{R}^n)\,. \tag{5.1}$$

Note that, in view of Lemma 3.17, we have $\mathcal{R}(x,\cdot)\in\mathcal{S}(\mathbb{R}^n)$ for each fixed $x\in\mathbb{R}^n$. Therefore the expression on the right hand side of (5.1) makes sense.

Proof. Recall that the Schwartz kernel of the transposed operator R^T is $\mathcal{R}(y,x)$, so that $R^Tv(y)=\int\mathcal{R}(x,y)\,v(x)\,\mathrm{d}x$ for all $v\in\mathcal{S}(\mathbb{R}^n)$.

Assume first that $u\in\mathcal{E}'(\mathbb{R}^n)$. Then, applying Theorem 2.15, we see that for all $v\in\mathcal{S}(\mathbb{R}^n)$

$$\begin{aligned}\langle Ru,v\rangle\ &=\ \langle u,R^Tv\rangle\ =\ \sum_{|\alpha|\le m}(-1)^{|\alpha|}\langle u_\alpha(y),\partial_y^\alpha(R^Tv)(y)\rangle\\ &=\ \sum_{|\alpha|\le m}(-1)^{|\alpha|}\int u_\alpha(y)\left(\partial_y^\alpha\int\mathcal{R}(x,y)\,v(x)\,\mathrm{d}x\right)\mathrm{d}y\\ &=\ \sum_{|\alpha|\le m}(-1)^{|\alpha|}\left(\int u_\alpha(y)\,\partial_y^\alpha\mathcal{R}(x,y)\,\mathrm{d}y\right)v(x)\,\mathrm{d}x\ =\ \int\langle u(y),\mathcal{R}(x,y)\rangle v(x)\,\mathrm{d}x\,,\end{aligned}$$

where u_α are the continuous compactly supported functions given by Theorem 2.15. Thus (5.1) holds true whenever $u\in\mathcal{E}'(\mathbb{R}^n)$.

If $u\notin\mathcal{E}'(\mathbb{R}^n)$, let us choose a function $\chi\in C_0^\infty(\mathbb{R}^n)$ which is equal to 1 on the ball $\{|x|\le1\}$ and consider the family of distributions $u_t(x):=\chi(tx)\,u(x)$. Since $\chi(tx)-1=0$ for $|x|<t^{-1}$, we have

$$\sup_{x\in\mathbb{R}^n}|x^\beta\partial_x^\alpha\big(\chi(tx)v(x)-v(x)\big)|=\sup_{x\,:\,|x|>t^{-1}}|x^\beta\partial_x^\alpha\big(\chi(tx)v(x)-v(x)\big)|\underset{t\to0}{\to}0$$

for all $v\in\mathcal{S}(\mathbb{R}^n)$ and all multi-indices α,β. It follows that $\chi(tx)v(x)\overset{\mathcal{S}}{\to}v(x)$ and, consequently, $\langle u_t,v\rangle\to\langle u,v\rangle$ for all $f\in\mathcal{S}(\mathbb{R}^n)$ as $t\to0$. Now the lemma is proved by applying (5.1) to the distribution $u_t\in\mathcal{E}'(\mathbb{R}^n)$ for each fixed t and letting $t\to0$ in the identity $\langle u_t,R^Tv\rangle=\int\langle u_t(y),\mathcal{R}(x,y)\rangle v(x)\,\mathrm{d}x\,.$ □

Corollary 5.5. *If $R\in\Psi^{-\infty}$ then $R:\mathcal{S}'(\mathbb{R}^n)\mapsto C^\infty(\mathbb{R}^n)$ and $R:\mathcal{E}'(\mathbb{R}^n)\mapsto\mathcal{S}(\mathbb{R}^n)$.*

Proof. By Taylor's formula

$$\varepsilon^{-1}\left(\mathcal{R}(x_1,\dots,x_k+\varepsilon,\dots,x_n,y)-\mathcal{R}(x,y)\right)-\partial_{x_k}\mathcal{R}(x,y)$$
$$=\ \varepsilon\int_0^1 \partial_{x_k}^2\mathcal{R}(x_1,\dots,x_k+t\varepsilon,\dots,x_n,y)\,(1-t)\,\mathrm{d}t\,.$$

If the Schwartz kernel $\mathcal{R}(x,y)$ satisfies (3.5) then the above identity implies that

$$\varepsilon^{-1}\left(\mathcal{R}(x_1,\dots,x_k+\varepsilon,\dots x_n,y)-\mathcal{R}(x,y)\right)\ \overset{\mathcal{S}}{\to}\ \partial_{x_k}\mathcal{R}(x,y)\,,\qquad \varepsilon\to 0\,,$$

for every fixed $x=(x_1,\dots,x_n)\in\mathbb{R}^n$ and all $k=1,2,\dots,n$. Therefore, for every $u\in\mathcal{S}'(\mathbb{R}^n)$, the function $\langle u(y),\mathcal{R}(x,y)\rangle$ of variable x is differentiable and

$$\partial_{x_k}\langle u(y),\mathcal{R}(x,y)\rangle\ =\ \langle u(y),\partial_{x_k}\mathcal{R}(x,y)\rangle$$

for all $k=1,2,\dots,n$. Since the derivatives of $\mathcal{R}$ also satisfy (3.5), the function $\langle u(y),\mathcal{R}(x,y)\rangle$ is infinitely differentiable and $\partial_x^\alpha\langle u(y),\mathcal{R}(x,y)\rangle=\langle u(y),\partial_x^\alpha\mathcal{R}(x,y)\rangle$ for all multi-indices α. If, in addition, $u\in\mathcal{E}'(\mathbb{R}^n)$ then, applying Theorem 2.15, one can easily show that this function and all its derivatives vanish faster than any power of $|x|$ as $|x|\to\infty$. □

5.4. **Wave front sets and PDOs.** The following lemma shows that $Q_{a,\chi}$ in Definition 5.3 can be replaced by a much more general PDO.

Lemma 5.6. *The point (x_0,ξ_0) does not belong to* WF u *if and only if there exists a classical PDO Q such that $Qu\in C^\infty(\mathbb{R}^n)$ and the principal symbol of Q does not vanish at (x_0,ξ_0).*

Proof. If $(x_0,\xi_0)\notin \operatorname{WF}u$ then we can take $Q=Q_{a,\chi}$, where $Q_{a,\chi}$ is the classical PDO from Definition 5.3.

Conversely, assume that $Qu\in C^\infty(\mathbb{R}^n)$ for some classical PDO $Q\in\Psi^m_{\mathrm{cl}}$ whose principal symbol does not vanish at (x_0,ξ_0). Let us fix arbitrary functions $\chi_1,\chi_2\in C_0^\infty(\mathbb{R}^n)$ such that $\chi_1=1$ in a neighbourhood of x_0 and $\chi_2=1$ in a neighbourhood of $\operatorname{supp}\chi_1$. Denote $Q_0=\chi_1Q\chi_2$. Lemma 4.7 implies that $\chi_1Q-Q_0\in\Psi^{-\infty}$. Since $Qu\in C^\infty(\mathbb{R}^n)$, from Corollary 5.5 it follows that $Q_0u\in C_0^\infty(\mathbb{R}^n)$.

As $\chi_1(x_0)=\chi_2(x_0)=1$, the principal symbols of Q and Q_0 coincide at (x_0,ξ_0). Therefore, there exist a neighbourhood Ω_{x_0} and a conic neighbourhood Ω_{ξ_0} such that the principal symbol of Q_0 is separated from zero on $\mathcal{O}\cap\{|\xi|=1\}$, where $\mathcal{O}:=\Omega_{x_0}\times\Omega_{\xi_0}$. Let P be a microlocal parametrix of Q_0 in $\mathcal{O}$ (see Remark 4.12), and let $R_{a,\chi}=Q_{a,\chi}PQ_0-Q_{a,\chi}$ where $\chi\in C_0^\infty(\mathbb{R}^n)$ and $a\in S^m_{\mathrm{cl}}$ are such that $\operatorname{supp}\chi\subset\Omega_{x_0}$ and $\operatorname{supp}a\subset\mathcal{O}_{\xi_0}$. Then, by Lemma 4.7, $R_{a,\chi}\in\Psi^{-\infty}$. Also, $R_{a,\chi}=R_{a,\chi}\tilde\chi$ for any function $\tilde\chi\in C_0^\infty(\mathbb{R}^n)$ equal to 1 on $\operatorname{supp}\chi$ and $\operatorname{supp}\chi_2$. In view of Corollary 5.5, this implies that $R_{a,\chi}:\mathcal{S}'(\mathbb{R}^n)\mapsto\mathcal{S}(\mathbb{R}^n)$. Since $Q_0u\in\mathcal{S}(\mathbb{R}^n)$ and, consequently $Q_{a,\chi}PQ_0u\in\mathcal{S}(\mathbb{R}^n)$, we have $Q_{a,\chi}u\in\mathcal{S}(\mathbb{R}^n)$. By Definition 5.3, it follows that $(x_0,\xi_0)\notin\operatorname{WF}u$. □

Corollary 5.7. *If P is a classical PDO whose symbol vanishes in a conic (with respect to ξ) neighbourhood $\mathcal{O}$ of (x_0,ξ_0) then $(x_0,\xi_0)\notin\operatorname{WF}(Pu)$.*

Proof. By Lemma 4.7, we have $QP\in\Psi^{-\infty}$ for every classical PDO Q whose symbol vanishes outside a smaller conic set $\mathcal{O}'\subset\mathcal{O}$. In view of Corollary 5.5, this implies that $Q(Pu)\in C^\infty(\mathbb{R}^n)$ for every such a PDO Q. □

The following is a version of the famous elliptic regularity theorem.

Corollary 5.8. *Let A be a classical PDO whose principal symbol does not vanish at (x_0,ξ_0). Then $(x_0,\xi_0)\in \mathrm{WF}\,u$ if and only if $(x_0,\xi_0)\in \mathrm{WF}(Au)$.*

Proof. If $(x_0,\xi_0)\notin \mathrm{WF}(Au)$ then there exists a classical PDO Q whose principal symbol does not vanish at (x_0,ξ_0), such that $QAu\in C^\infty(\mathbb{R}^n)$. Since the principal symbol of QA does not vanish at (x_0,ξ_0), it follows that $(x_0,\xi_0)\notin \mathrm{WF}\,u$.

Let us now assume that $(x_0,\xi_0)\in \mathrm{WF}(Au)$ and let B be a microlocal parametrix of A in a neighbourhood of (x_0,ξ_0). Then, applying Corollary 5.7 with $P=BA-I$, we see that $(x_0,\xi_0)\notin \mathrm{WF}(BAu-u)$. At the same time, by the above, $(x_0,\xi_0)\in \mathrm{WF}(BAu)$. This implies that $(x_0,\xi_0)\in \mathrm{WF}\,u$. □

Remark. PDOs play the same role in microlocal analysis as smooth cut-off functions in classical analysis. One can construct, for example, a microlocal partition of unity using the classical PDOs and represent an arbitrary function as the sum of functions with small wave front sets which are often easier to deal with (this procedure is called *microlocalization*).

5.5. **Propagation of singularities.** If A is an elliptic PDO then, by Corollary 5.8, we have $\mathrm{WF}(Au)=\mathrm{WF}\,u$ and, consequently, $\operatorname{sing\,supp}(Au)=\operatorname{sing\,supp}u$. In other words, a solution u of an elliptic (pseudo)differential equation $Au=f$ has the same singularities as the function f. This is usually not true if the operator A is not elliptic: the solutions u of a non-elliptic equation $Au=f$ may have additional singularities.

Corollary 5.8 immediately implies that $\mathrm{WF}\,u\subset \mathrm{WF}(Au)\cup \operatorname{Char}A$, where $\operatorname{Char}A$ is the set of zeros of the principal symbol of A. A more interesting question is what happens with the singularities inside the set $\operatorname{Char}A$. The following theorem answers this question (see, for example, [Sh, Appendix 1]).

Theorem 5.9. *Let B be a classical PDO with a real principal symbol $b_m(x,\xi)$ and let $(x(t),\xi(t))$ be a solution of the Hamiltonian system*

$$\dot{x}(t) = \partial_\xi b_m(x(t),\xi(t)), \qquad \dot{\xi}(t) = -\partial_x b_m(x(t),\xi(t)). \tag{5.2}$$

Assume that $b_m(x(t),\xi(t))=0$ and $(x(t),\xi(t))\notin \mathrm{WF}(Bu)$ for all $t\in(t_1,t_2)$. Then either $(x(t),\xi(t))\notin \mathrm{WF}\,u$ or $(x(t),\xi(t))\in \mathrm{WF}\,u$ for all $t\in(t_1,t_2)$.

Remark 5.10. One can easily see that $b_m(x(t),\xi(t))$ is constant, so $b_m(x(t),\xi(t))=0$ for all $t\in(t_1,t_2)$ provided that $b_m(x(t_0),\xi(t_0))=0$ for some fixed $t_0\in(t_1,t_2)$.

The Hamiltonian trajectories $(x(t),\xi(t))$ satisfying $b_m(x(t),\xi(t))=0$ are said to be the *bicharacteristics* of the operator B. By Theorem 5.9, if a solution u of the equation $Bu=0$ has at least one singularity on a bicharacteristic $(x(t),\xi(t))$ then the whole trajectory lies in $\mathrm{WF}\,u$. This effect is called propagation of singularities.

Solutions to exercises

Solution 1. If $\varepsilon < 1$ then

$$\int_{|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t = \int_{|t|\geq 1} t^{-1} f(t)\,\mathrm{d}t + \int_{1>|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t\,.$$

The first integral is estimated by

$$\int_{|t|\geq 1} t^{-2}\,|t\,f(t)|\,\mathrm{d}t \;\leq\; 2\sup_{s\in\mathbb{R}}|s\,f(s)|\int_1^\infty t^{-2}\,\mathrm{d}t \;=\; 2\sup_{s\in\mathbb{R}}|s\,f(s)|\,.$$

Since $|f(t)-f(-t)| \leq 2t\sup_{s\in[-t,t]}|f'(s)|$, the second integral is estimated as follows

$$|\int_{1>|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t\,| \;=\; |\int_\varepsilon^1 t^{-1}\,(f(t)-f(-t))\,\mathrm{d}t\,| \;\leq\; 2\sup_{s\in\mathbb{R}}|f'(s)|\,.$$

Thus $|\int_{|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t\,| \leq 2\sup_{s\in\mathbb{R}}|s\,f(s)| + 2\sup_{s\in\mathbb{R}}|f'(s)|$ for all $\varepsilon\in(0,1]$ and $f\in\mathcal{S}(\mathbb{R})$. This implies that $u_0 : f \mapsto \lim_{\varepsilon\to 0}\int_{|t|>\varepsilon} t^{-1} f(t)\,\mathrm{d}t$ is a continuous functional on $\mathcal{S}(\mathbb{R})$, that is, $u_0\in\mathcal{S}'(\mathbb{R})$.

Solution 2. Clearly, the distribution u_0 from Exercise 2 satisfies this condition. If u is another distribution with the same property then $\langle u-u_0, f\rangle$ depends only on the value of f at the origin. The map $f(0)\to\langle u-u_0,f\rangle$ is a linear functional on $\mathbb{C}$, and therefore $\langle u-u_0,f\rangle = cf(0)$ with some constant c. This implies that $u=u_0+c\delta$, where δ is the δ-function at the origin. Conversely, every distribution $u=u_0+c\delta$ has the required property.

Solution 3. Let $c_k = u(a_k+0)-u(a_k-0)$ be the jumps of u at the points a_k, and let

$$v(x) \;=\; \begin{cases} u'(x), & x\in(a_k,a_{k+1})\,, \qquad k=1,\dots,m-1,\\ 0, & x<a_1 \text{ or } x>a_m. \end{cases}$$

Integrating by parts we obtain $-\langle u,f'\rangle = -\int u\,f'\,\mathrm{d}x = \int v\,f\,\mathrm{d}x + \sum_{k=1}^m c_k\,f(a_k)$. Therefore $u' = v + \sum_{k=1}^m c_k\,\delta_{a_k}$ where δ_{a_k} are the δ-functions at the points a_k.

Solution 4 (proof of Lemma 3.17). Let $\mathcal{R}(x,y)$ be the Schwartz kernel of R. If $R\in\Psi^{-\infty}$ then $\mathcal{R}(x,y)$ is infinitely smooth (since we can differentiate under the integral sign), and the required estimates follow from (3.7).

Conversely, if $\mathcal{R}$ is smooth and satisfies (3.5) then it is represented by the oscillatory integral with the amplitude $a(x,\xi)=(2\pi)^{n/2}\mathcal{F}_{z\to\xi}\tilde{\mathcal{R}}(x,z)$, where $\tilde{\mathcal{R}}$ is defined by the equality $\tilde{\mathcal{R}}(x,x-y)=\mathcal{R}(x,y)$.

Solution 5. Let $A \in \Psi^m$, and let σ be the symbol of A, so that

$$\mathcal{A}(x,y) \;=\; (2\pi)^{-n} \int e^{i(x-y)\cdot\xi} \sigma(x,\xi)\, \mathrm{d}\xi$$

modulo a function $\mathcal{R}(x,y)$ defining an operator from $\Psi^{-\infty}$ (see Lemma 3.17). Lemma 3.5 implies that A^T is a PDO whose dual symbol coincides with σ. Now (1) and (2) follow from (3.9).

By (2), A is elliptic if and only if A^T is elliptic. Applying Theorem 4.11, let us find a PDO $B_1 \in \Psi^{-m}_{\mathrm{cl}}$ such that $B_1 A^T - I \in \Psi^{-\infty}$. In view of Lemmas 3.5 and 3.17, we have $A B_1^T - I = (B_1 A^T - I)^T \in \Psi^{-\infty}$. It remains to notice that, by (1), $B_1^T \in \Psi^{-m}_{\mathrm{cl}}$.