A proof of the Schwartz kernel theorem

Ethan Y. Jaffe

We show the following:

Theorem 1.1. *Let X, Y be smooth manifolds. For every $A : C_0^\infty(Y) \to \mathcal{D}'(X)$ continuous and linear, there exists a unique $K \in \mathcal{D}'(X \times Y)$ such that*

$$\langle Au, v \rangle = \langle K, u \otimes v \rangle.$$

By continuity, we mean sequential continuity, i.e. that if $\varphi_j \to \varphi$ in $C_0^\infty(Y)$, then $A\varphi_j \to A\varphi$ weakly in $\mathcal{D}'(X)$

Proof. Uniqueness is obvious since tensor products $C_c^\infty(Y) \otimes C_c^\infty(X)$ are dense in $C_c^\infty(Y \times X)$ (which can be seen for the case that Y and X are cubes by using Fourier series, and then employing a partition of unity which lives in the tensor product).

Now for existence.

Step 1: We first prove an analogous theorem for $\mathcal{S}(\mathbf{R}^d)$, where all the work will be done. Explicitly we prove the same theorem, except with $\mathcal{S}(\mathbf{R}^m)$ $\mathcal{S}'(\mathbf{R}^n)$ as the domain and codomain of A, respectively, first. Set $\langle x \rangle = (1 + |x|^2)^{1/2}$. We define $\langle x \rangle^s H^t(\mathbf{R}^d)$ to be the space of all distributions u of the form

$$u = \langle x \rangle^s v$$

for $v \in H^t(\mathbf{R}^d)$, with a norm $||u||_{\langle x \rangle^s H^t(\mathbf{R}^d)} = ||v||_{H^t(\mathbf{R}^d)}$. We have that

$$\mathcal{S}(\mathbf{R}^d) = \bigcap_{k \in \mathbf{Z}} \langle x \rangle^{-k} H^k(\mathbf{R}^d)$$

and

$$\mathcal{S}'(\mathbf{R}^d) = \bigcup_{k \in \mathbf{Z}} \langle x \rangle^k H^{-k}(\mathbf{R}^d),$$

moreover the intersection (resp. union) is decreasing (resp. increasing), and the inclusions maps are continuous. Furthermore, $\varphi_j \to \varphi$ in $\mathcal{S}(\mathbf{R}^d)$ iff $\varphi_j \to \varphi$ in all spaces $\langle x \rangle^{-k} H^k(\mathbf{R}^d)$ and $u_j \to u$ in $\mathcal{S}'(\mathbf{R}^d)$ iff $u_j \to u$ in some space $\langle x \rangle^k H^{-k}(\mathbf{R}^d)$. See the Appendix for details.

Using this, we show that there is some k, ℓ so that A extends to a bounded linear map

$$A : \langle x \rangle^{-k} H^k(\mathbf{R}^m) \to \langle x \rangle^\ell H^{-\ell}(\mathbf{R}^n).$$

The major idea in this note of passing to a δ function, was taken from notes written by Richard Melrose

Proof. Notice that there is some constant $C_{k,\ell,d}$ such that whenever $k < \ell$,

$$||\cdot||_{\langle x \rangle^k H^k(\mathbf{R}^d)} \le C_{k,\ell,d}||\cdot||_{\langle x \rangle^{-\ell} H^\ell(\mathbf{R}^d)}.$$

Dividing by a constant, we may thus inductively renormalize the norms $||\cdot||_{\langle x \rangle^k H^{-k}(\mathbf{R}^d)}$ for $\mathbf{Z} \ni k \ge 0$ so that $C_{-k-1,-k,d} = 1$.

Suppose A did not extend to a continuous map. Then taking $k = \ell$, for each ℓ fixed there would be $\varepsilon_\ell > 0$ and a sequence $\varphi_{j,\ell} \to 0$ in $\langle x \rangle^{-\ell} H^\ell(\mathbf{R}^m)$ as $j \to \infty$ such that

$$||A\varphi_{j,\ell}||_{\langle x \rangle^\ell H^{-\ell}(\mathbf{R}^n)} \ge \varepsilon_\ell.$$

Rescaling $\varphi_{j,\ell}$, we may assume that $\varepsilon_\ell = 1$ for all ℓ. Passing to a subsequence, we may also assume that $||\varphi_{j,\ell}||_{\langle x \rangle^{-\ell} H^\ell} \le 2^{-j}$. Set $\psi_j = \varphi_{j,j}$. Then for all ℓ and $j \ge \ell$,

$$||\varphi_{j,j}||_{\langle x \rangle^{-\ell} H^\ell(\mathbf{R}^m)} \le ||\varphi_{j,j}||_{\langle x \rangle^{-j} H^j(\mathbf{R}^m)} \le 2^{-j},$$

and so $\psi_j \to 0$ in $\mathcal{S}(\mathbf{R}^m)$.

Thus $A\psi_j \to 0$ in $\langle x \rangle^k H^{-k}(\mathbf{R}^n)$ for some k. Making k smaller, we may assume that $k \le 0$. But

$$||A\psi_j||_{\langle x \rangle^k H^{-k}(\mathbf{R}^d)} \ge ||A\psi_j||_{\langle x \rangle^j H^{-j}(\mathbf{R}^d)} \ge 1,$$

at least for large enough j, since we are assuming that $C_{-j,-k,m} = 1$. This is of course a contradiction. $\qquad\square$

Multiplication by $\langle x \rangle^s$ is an isomorphism of $\mathcal{S}$ for any s, and hence $\langle D \rangle^s$ is, too, and hence both are also of $\mathcal{S}'$ ($\langle D \rangle$ is defined by multiply the Fourier transform by $\langle \xi \rangle$ and then taking the inverse transform). Using the above statement, we see that in particular there is an operator B which is a composition of A with the above maps such that $B : H^{-m}(\mathbf{R}^m) \to H^n(\mathbf{R}^n)$. Now $\delta_y \in H^{-m}(\mathbf{R}^m)$, and hence, $B(\delta_y) \in H^n(\mathbf{R}^n)$. Moreover, $\delta.$ is a bounded continuous function into $H^{-m}(\mathbf{R}^m)$. In particular, the function $L(x, y) = B(\delta_y)(x)$ is well-defined and a bounded continuous function by Sobolev embedding, and so defines a distribution on $\mathcal{S}(\mathbf{R}^{n+m})$ by integrating. We next verify that $\int L(x, y)\varphi(y) \, dy = B(\varphi)(x)$, i.e. L is the kernel of B. L makes sense to evaluate pointwise in x by Sobolev embedding, and the integral makes sense since $L(x, \cdot)$ is bounded and continuous. We write the integral as a limit of Riemann sums which converge at least pointwise in x:

$$\int L(x,y)\varphi(y) \, dy = \lim_{\varepsilon \to 0} \varepsilon^m \sum_{\nu \in (\varepsilon\mathbf{Z})^m, \ |\nu| \le \varepsilon^{-1}} B(\delta_\nu)(x)\varphi(\nu)$$

$$= \lim_{\varepsilon \to 0} \varepsilon^m \sum B(\varphi(\nu)\delta_\nu)(x)$$

$$= \lim_{\varepsilon \to 0} B\left(\varepsilon^m \sum \varphi(\nu)\delta_\nu\right)(x).$$

Now, the argument of B converges in $H^{-m}(\mathbf{R}^m)$ to φ. Indeed, the Fourier transform of the arugment is

$$\varepsilon^m \sum_\nu e^{-i\nu\xi}\varphi(\nu),$$

which is a Riemann sum for the defining integral of $\hat{\varphi}(\xi)$, which certainly converges to $\hat{\varphi}$ in the appropriate weighted L^2 space by dominated convergence. Indeed,

$$\left| \varepsilon^m \sum_\nu e^{-i\nu\xi}\varphi(\nu) \right| \lesssim \varepsilon^m \sum_\varepsilon \langle\nu\rangle^{-N} \sim \int_{\mathbf{R}^m} \langle x\rangle^{-N} \lesssim 1,$$

and so

$$\left| \varepsilon^m \sum_\nu e^{-i\nu\xi}\varphi(\nu) - \varphi(\xi) \right|^2 \langle\xi\rangle^{-m} \lesssim \langle\xi\rangle^{-m}$$

is integrable. Thus,

$$\varepsilon^m \sum \varphi(\nu)\delta_\nu \to \varphi$$

in $H^{-m}(\mathbf{R}^d)$. Since B is continuous from $H^{-m}(\mathbf{R}^m)$ to $L^\infty(\mathbf{R}^n)$, we deduce from the above that

$$\int L(x,y)\varphi(y) \, dy = \lim_{\varepsilon\to 0} B\left(\varepsilon^m \sum \varphi(\nu)\delta_\nu \right)(x) = B(\varphi)(x),$$

which is what we wanted to show.

Finally, we deduce that for approrpriate i_k, $k = 1, 2, 3, 4$,

$$\begin{aligned}
\langle Au, v \rangle &= \langle \langle D_x \rangle^{i_1} \langle x \rangle^{i_2} B \langle y \rangle^{i_3} \langle D_y \rangle^{i_4}, v \rangle \\
&= \langle L, \langle y \rangle^{i_3} \langle D_y \rangle^{i_4} u \otimes \langle D_x \rangle^{i_1} \langle x \rangle^{i_2} v \rangle,
\end{aligned}$$

and so we may define the kernel K to A by

$$\langle K, \varphi \rangle = \langle L, \langle y \rangle^{i_3} \langle D_y \rangle^{i_4} \langle D_x \rangle^{i_1} \langle x \rangle^{i_2} \varphi \rangle,$$

which certainly defined a Schwartz-distribution.

Step 2: Now, returning to the original question. Suppose Y is an open subset of $\mathbf{R}^m$ and X is an open subset of $\mathbf{R}^n$, and let Y_j, X_j denote compact exhaustions of Y and X, respectively Then we may define maps $A_j : \mathcal{S}(\mathbf{R}^m) \to \mathcal{S}'(\mathbf{R}^n)$ by

$$\langle A_j u, v \rangle = \langle A\eta_j u, \chi_j v \rangle,$$

η_j, χ_j are cutoffs of Y_j and X_j, respectively. Notice that the associated kernels K_j of A_j by uniqueness must extend one another. The result follows. Explicitly, if $\varphi \in C_c^\infty(Y \times X)$, we may define

$$\langle K, \varphi \rangle = \langle K_j, (\eta_j \otimes \chi_j)\varphi \rangle$$

if $\operatorname{supp}\varphi \in Y_j \times X_j$. This is well-defined since if $k > j$, then

$$\langle K_k, (\eta_k \otimes \chi_k)\varphi \rangle = \langle K_k, (\eta_j \otimes \chi_j)\varphi \rangle,$$

3

but K_k and K_j must be the same when acting on functions in $C_c^\infty(Y_j \times X_j)$, by uniqueness (if $\varphi = \varphi_1 \otimes \varphi_2$, then both must be $\langle A\varphi_1, \varphi_2\rangle$). K is the kernel for A since if $u \in C_c^\infty(Y)$, $v \in C_c^\infty(X)$, then for large enough j,

$$\langle K, u \otimes v\rangle = \langle K_j, (\eta_j u) \otimes (\chi_j v)\rangle = \langle A_j \eta_j u, \chi_j v\rangle = \langle A\eta_j^2 u, \chi_j^2 v\rangle = \langle Au, v\rangle.$$

Lastly, we need to verify that K is continuous. But this is obvious since $\varphi_j \to \varphi$ means in particular that for large enough j, the supports of the φ_j and φ are uniformly compactly supported.

Step 3: On a manifold, we simply need to use partitions of unity. Suppose φ_j is a partition of unity subordinate to a collection of charts Y_j on Y, and ψ_j is subordinate to a collection of charts X_j of X. Set $A_{i,j} u = \psi_i A(\varphi_j u)$. This then has kernel $K_{i,j} \in \mathcal{D}'(Y_j \times X_i)$. Let φ_j' be supported in Y_j and 1 on the support of φ_j, and similarly for ψ_j'. Then $K = \sum(\psi_i' \otimes \varphi_j')K_{i,j}$ is a well-defined distribution (even if the sum is infinite, since it acts on compactly supported densities), and is the distribution kernel for A. Indeed,

$$\begin{aligned}
\langle K, u \otimes v\rangle &= \sum \langle K_{i,j}, (\varphi_j' u) \otimes (\psi_i' v)\rangle \\
&= \sum \langle A_{i,j}(\varphi_j' u), \psi_i' v\rangle \\
&= \sum \langle A(\varphi_j' \varphi_j u), \psi_i' \psi_i v\rangle = \sum \langle A(\varphi_j u), \psi_i v\rangle \\
&= \langle Au, v\rangle.
\end{aligned}$$

$\square$

Here we have supressed the importance that $\mathcal{D}'(M)$ is the dual to the compactly supported densities of a manifold on M. The implicit use of this was in the following: if U is a subset of M in the image of a coordinate patch with domain V, we identified $\mathcal{D}'(U)$ and $\mathcal{D}'(V)$ without further comment; to be rigorous, one must note the obvious correspondence between a smooth function f on V, and the "natural" density $f\,dx$.

A. Appendix

The purpose of this appendix is to (at least sketch) the proof of the following.

Theorem A.1. *If $k \geq j$ are integers, then we have a continuous inclusion*

$$\langle x\rangle^{-j} H^j(\mathbf{R}^d) \subseteq \langle x\rangle^{-k} H^k(\mathbf{R}^d).$$

Moreoever, the following are true:

(i) $\mathcal{S}(\mathbf{R}^d) = \bigcap_{k\in\mathbf{Z}} \langle x\rangle^{-k} H^k(\mathbf{R}^d)$

(ii) $\mathcal{S}'(\mathbf{R}^d) = \bigcup_{k\in\mathbf{Z}} \langle x\rangle^{k} H^{-k}(\mathbf{R}^d)$

4

Furthermore, if $\varphi_n \to \varphi$ in $\mathcal{S}(\mathbf{R}^d)$, then for all k $\varphi_n \to \varphi$ in the topology of $\langle x \rangle^{-k} H^k(\mathbf{R}^d)$. Similarly, if $u_n \to u$ weakly in $\mathcal{S}'(\mathbf{R}^d)$, then there is some k for which $u_n \to u$ in the topology of $\langle x \rangle^k H^{-k}(\mathbf{R}^d)$.

Remark A.2. Although the Theorem is true if we use real k instead of integer k, the technical estimates, which are implicit throughout the proof, are easier to prove for integer-indexed Sobolev spaces, since these have a characterization not using the Fourier transform, i.e. in terms of derivatives being in L^2 for $k \geq 0$, and their duals for $k \leq 0$. The required technical theorem to extend to arbitrary Sobolev spaces is the Sobolev multiplication lemma. In the present case, one can instead simply use the product rule.

Remark A.3. The version of this Theorem for compact manifolds is much simpler to prove, since there are no weights on the Sobolev spaces. It is possible to argue the Schwartz kernel theorem for compact manifolds directly, using the techniques in this note, without having to go through as many technical details.

Proof. Suppose $u \in \langle x \rangle^{-j} H^j(\mathbf{R}^n)$, i.e. $u = \langle x \rangle^{-j} v$, where $v \in H^j(\mathbf{R}^d)$. Then if $k \geq j$ are integers, then

$$u = \langle x \rangle^{-k} \left(\langle x \rangle^{-(j-k)} v \right),$$

and $v \in H^k(\mathbf{R}^d)$. Now $\langle x \rangle^{-(j+k)}$ is smooth and bounded together with all its derivatives. It follows that multiplication by $\langle x \rangle^{-(j+k)}$ is a bounded operator on $H^\ell(\mathbf{R}^d)$ for $\ell \geq 0$, and hence for all $\ell \in \mathbf{Z}$ by duality. In particular, $u \in \langle x \rangle^{-k} H^k(\mathbf{R}^d)$ with the estimate

$$||u||_{\langle x \rangle^{-k} H^k(\mathbf{R}^d)} = ||\langle x \rangle^{-(j-k)} v||_{H^k(\mathbf{R}^d)} \leq C ||v||_{H^k(\mathbf{R}^d)} \leq C ||v||_{H^j(\mathbf{R}^d)} = C ||u||_{\langle x \rangle^{-j} H^j(\mathbf{R}^d)}.$$

This proves the first statement. Observe that the proof actually shos that if $k \geq j$ and $k' \geq j'$ then we have the continuous inclusion

$$\langle x \rangle^{-j'} H^j(\mathbf{R}^d) \subseteq \langle x \rangle^{-k'} H^k(\mathbf{R}^d).$$

For the second, we define the Schwartz seminorms

$$||u||_{j,j'} = \sup_{x \in \mathbf{R}^d} |\langle x \rangle^{j'} \partial^j u|.$$

Here, $\partial^j u$ is interepted as a vector of all partials of order j, and $|\cdot|$ as the sup norm. Now by Sobolev embedding,

$$\sup_{j' \leq j - d - 1} |\partial^{j'} \langle x \rangle^j u| \leq C ||u||_{\langle x \rangle^{-j} H^j(\mathbf{R}^d)}.$$

It is not hard to see (using induction, for instance), that

$$\sup_{j \leq N, \, j' \leq j - d - a} ||u||_{j,j'} \leq C \left(\sup_{j \leq N, j' \leq j - d - 1} |\partial^{j'} \langle x \rangle^j u| \right).$$

The former is a family of norms on the Schwartz space. The latter is bounded by

$$C\left(\sup_{j\leq N}||u||_{\langle x\rangle^{-j}H^j(\mathbf{R}^d)}\right) \leq C||u||_{\langle x\rangle^{-N}H^N(\mathbf{R}^d)},$$

since the norm for $j = N$ controls the others. Since we obviously have a reverse inequality controlling the norms on $\langle x\rangle^{-N}H^N(\mathbf{R}^d)$ in terms of the Schwartz norms, we deduce that the families are equivalent, and so (i) of the Theorem follows.

For the converse, we take duals to see that

$$\mathcal{S}'(\mathbf{R}^d) = \bigcup_{k\in\mathbf{Z}}\langle x\rangle^k H^{-k}(\mathbf{R}^d).$$

Now for the convergence. $\mathcal{S}(\mathbf{R}^d)$ is a Fréchet space, and so by the Banach-Steinhaus theorem, if $u_n \to u$ weakly, the family u_n is equicontinuous. By (i) and the fact that the norms are increasing) we know that this means that there is some j for which

$$|\langle u_n, \varphi\rangle| \lesssim ||\varphi||_{\langle x\rangle^{-j}H^j(\mathbf{R}^d)}.$$

Since $\mathcal{S}(\mathbf{R}^d)$ is certainly dense in $\langle x\rangle^k H^\ell(\mathbf{R}^d)$ for any k, ℓ, this means that

$$||u_n||_{\langle x\rangle^j H^{-j}(\mathbf{R}^d)} \lesssim 1.$$

Next we show that there is some j such that every subsequence has a further subsequence which converges to u in $\langle x\rangle^{j'}H^{-j'}(\mathbf{R}^n)$, which is sufficient. We use the following lemma.

Lemma A.4. *Suppose $v_i \in \langle x\rangle^{k'}H^k(\mathbf{R}^d)$ are bounded. Suppose $\ell < k$. Then there is a subsequence v_{n_ℓ} such that v_{i_m} is convergent in $\langle x\rangle^{\ell'}H^\ell(\mathbf{R}^d)$, for some $\ell' \geq k'$.*

Given the lemma, we see that every subsequence has a further subsequence converging to something in some space $\langle x\rangle^{j'}H^{-j'}(\mathbf{R}^d)$. Since u is the weak limit of the subsequence, it must be the limit in $\langle x\rangle^{j'}H^{-j'}(\mathbf{R}^d)$, too (just test against $\mathcal{S}(\mathbf{R}^n)$). Since every subsequence has a further subsequence converging to the same limit in the same space, the entire sequence converges as well.

Proof of Lemma. The idea is to cutoff and use Reillich-Kondrashov to obtain convergence on each cutoff. Then the difference between ℓ' and k' will allow us to control the left-over tails.

Let φ_n be smooth cutoffs of $B(0,n)$ in $\mathbf{R}^d$ with support in $B(0, n+1/2)$. Then for each n fixed, the sequence $\varphi_n v_i$ is uniformly bounded in $H^k(\mathbf{R}^d)$, and so, by Reillich-Kondrashov compactness, has a subsequence converging to some limit w_n in $H^\ell(\mathbf{R}^d)$. Taking a diagonal subsequence, v_{i_m} we may assume that for each n fixed $\varphi_n v_{i_m} \to w_n$. Now $\varphi_n w_p = w_n$ whenever $p > n$, since $\varphi_n \varphi_p = \varphi_n$. Let $w \in H^\ell_{loc}(\mathbf{R}^d)$ be the distribution defined by $\varphi_n w = w_n$. Then by construction $v_{i_m} \to w$ in $H^\ell_{loc}(\mathbf{R}^d)$. We will show that in fact $w \in \langle x\rangle^{\ell'}H^\ell(\mathbf{R}^d)$ for large enough ℓ', and furthermore than v_{i_m} converges to it.

We first observe that
$$\langle x \rangle^{-\ell'} v_{k_m} = \langle x \rangle^{k'-\ell'} \langle x \rangle^{-k'} v_{i_m}.$$

Set $e = \ell' - k'$. Observe that of e is sufficiently large depending on d, s then

$$||(1 - \varphi_n)\langle x \rangle^{-e}\langle x \rangle^{-k'} v_{i_m}||_{H^\ell(\mathbf{R}^d)} \lesssim ||\langle x \rangle^{-k'} v_{m_n}||_{H^\ell(\mathbf{R}^d)}||||(1 - \varphi_n)\langle x \rangle^{-e}||_{H^q(\mathbf{R}^d)}, \qquad \text{(A.1)}$$

The first factor is uniformly bounded by assumption. The second factor goes to 0 as $n \to \infty$. We thus have that for $\ell' > k'$ large enough

$$||v_{i_m} - v_{i_{m'}}||_{\langle x \rangle^{\ell'} H^\ell(\mathbf{R}^d)} \le ||(1 - \varphi_n)\langle x \rangle^{-e}\langle x \rangle^{-k'}(v_{i_m} - v_{i_{m'}})||_{H^\ell(\mathbf{R}^d)} + ||\langle x \rangle^{-\ell'}\varphi_n(v_{i_m} - v_{i_{m'}})||_{H^\ell(\mathbf{R}^d)}.$$

We have determined that the first term can be made arbitrarily small for n large. After this, we notice that $\langle x \rangle^{-\ell'}\varphi_m \in C_c^\infty(\mathbf{R}^d)$ and $\langle x \rangle^{-\ell'}\varphi_p\varphi_n = \langle x \rangle^{-\ell'}\varphi_n$ for $p > n$. Since multiplication by a function in $C_c^\infty(\mathbf{R}^d)$ is continuous on $H^\ell(\mathbf{R}^d)$, and $\varphi_n v_{i_m}$ is Cauchy in $H^\ell(\mathbf{R}^d)$ by assumption, the second term can be made arbtrarily small if m, m' is sufficiently large. Hence v_{i_m} is Cauchy and thus convergent, and must converge to w. $\square$

With the Lemma proved, so is the Theorem. $\square$