We now show that the weak type $(1,1)$ property of the Hardy–Littlewood maximal function implies almost everywhere convergence for a variety of families of functions. We deduce this from the more general fact that a certain weak type property for the supremum of a family of linear operators implies almost everywhere convergence.

Here is our setup. Let $(X,\mu)$, $(Y,\nu)$ be measure spaces and let $0<p\le\infty$, $0<q<\infty$. Suppose that $D$ is a dense subspace of $L^p(X,\mu)$. This means that for all $f\in L^p$ and all $\delta>0$ there exists a $g\in D$ such that $\|f-g\|_{L^p}<\delta$. Suppose that for every $\varepsilon>0$, $T_\varepsilon$ is a linear operator that maps $L^p(X,\mu)$ into a subspace of measurable functions, which are defined everywhere on $Y$. For $y\in Y$, define a sublinear operator

$$T_*(f)(y)=\sup_{\varepsilon>0}|T_\varepsilon(f)(y)| \tag{2.1.14}$$

and assume that $T_*(f)$ is $\nu$- measurable for any $f\in L^p(X,\mu)$. We have the following.

**Theorem 2.1.14.** *Let $0<p<\infty$, $0<q<\infty$, and $T_\varepsilon$ and $T_*$ as previously. Suppose that for some $B>0$ and all $f\in L^p(X)$ we have*

$$\big\|T_*(f)\big\|_{L^{q,\infty}}\le B\big\|f\big\|_{L^p} \tag{2.1.15}$$

*and that for all $f\in D$,*

$$\lim_{\varepsilon\to0}T_\varepsilon(f)=T(f) \tag{2.1.16}$$

*exists and is finite $\nu$-a.e. (and defines a linear operator on $D$). Then for all functions $f$ in $L^p(X,\mu)$ the limit (2.1.16) exists and is finite $\nu$-a.e., and defines a linear operator $T$ on $L^p(X)$ (uniquely extending $T$ defined on $D$) that satisfies*

$$\big\|T(f)\big\|_{L^{q,\infty}}\le B\big\|f\big\|_{L^p} \tag{2.1.17}$$

*for all functions $f$ in $L^p(X)$.*[1]

*Proof.* Given $f$ in $L^p$, we define the *oscillation* of $f$:

$$O_f(y)=\limsup_{\varepsilon\to0}\limsup_{\theta\to0}|T_\varepsilon(f)(y)-T_\theta(f)(y)|\,.$$

We would like to show that for all $f\in L^p$ and $\delta>0$,

$$\nu(\{y\in Y:\ O_f(y)>\delta\})=0\,. \tag{2.1.18}$$

Once (2.1.18) is established, given $f\in L^p(X)$, we obtain that $O_f(y)=0$ for $\nu$-almost all $y$, which implies that $T_\varepsilon(f)(y)$ is Cauchy for $\nu$-almost all $y$, and it therefore converges $\nu$-a.e. to some $T(f)(y)$ as $\varepsilon\to0$. The operator $T$ defined this way on $L^p(X)$ is linear and extends $T$ defined on $D$.

To approximate $O_f$ we use density. Given $\eta>0$, find a function $g\in D$ such that $\|f-g\|_{L^p}<\eta$. Since $T_\varepsilon(g)\to T(g)$ $\nu$-a.e, it follows that $O_g=0$ $\nu$-a.e. Using this fact and the linearity of the $T_\varepsilon$'s, we conclude that

$$O_f(y)\le O_g(y)+O_{f-g}(y)=O_{f-g}(y)\qquad \nu\text{-a.e.}$$

[1] The same conclusion is valid if $T_\varepsilon(f)$ is defined $\nu$-a.e. and the convergence in (2.1.16) is taken along a *fixed* sequence $\varepsilon_j\to0$. In this case $T^{(*)}(f)=\sup_j|T_{\varepsilon_j}(f)|$.