View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Expositiones Mathematicae 28 (2010) 385–394

Contents lists available at ScienceDirect

Expositiones Mathematicae

journal homepage: www.elsevier.de/exmath

The Kolmogorov–Riesz compactness theorem ☆

Harald Hanche-Olsen [a], Helge Holden [a,b,*]

[a] Department of Mathematical Sciences, Norwegian University of Science and Technology, NO–7491 Trondheim, Norway
[b] Centre of Mathematics for Applications, University of Oslo, P.O. Box 1053, Blindern, NO–0316 Oslo, Norway

ARTICLE INFO

MSC:
primary, 46E30, 46E35;
secondary, 46N20

Keywords:
Kolmogorov–Riesz compactness theorem
Compactness in L^p

ABSTRACT

We show that the Arzelà–Ascoli theorem and Kolmogorov compactness theorem both are consequences of a simple lemma on compactness in metric spaces. Their relation to Helly's theorem is discussed. The paper contains a detailed discussion on the historical background of the Kolmogorov compactness theorem.

© 2010 Elsevier GmbH. All rights reserved.

1. Introduction

Compactness results in the spaces $L^p(\mathbb{R}^d)$ $(1 \le p < \infty)$ are often vital in existence proofs for nonlinear partial differential equations. A necessary and sufficient condition for a subset of $L^p(\mathbb{R}^d)$ to be compact is given in what is often called the Kolmogorov compactness theorem, or Fréchet–Kolmogorov compactness theorem. Proofs of this theorem are frequently based on the Arzelà–Ascoli theorem. We here show how one can deduce both the Kolmogorov compactness theorem and the Arzelà–Ascoli theorem from one common lemma on compactness in metric spaces, which again is based on the fact that a metric space is compact if and only if it is complete and totally bounded.

Furthermore, we trace out the historical roots of Kolmogorov's compactness theorem, which originated in Kolmogorov's classical paper [18] from 1931. However, there were several other approaches to the issue of describing compact subsets of $L^p(\mathbb{R}^d)$ prior to and after Kolmogorov, and

☆ Supported in part by the Research Council of Norway. This paper was written as part of the international research program on Nonlinear Partial Differential Equations at the Centre for Advanced Study at the Norwegian Academy of Science and Letters in Oslo during the academic year 2008–2009. The authors would like to thank Helge Kristian Jenssen for fruitful discussions.

* Corresponding author at: Department of Mathematical Sciences, Norwegian University of Science and Technology, NO–7491 Trondheim, Norway.
E-mail addresses: hanche@math.ntnu.no (H. Hanche-Olsen), holden@math.ntnu.no (H. Holden).
URLS: http://www.math.ntnu.no/~hanche/ (H. Hanche-Olsen), http://www.math.ntnu.no/~holden/ (H. Holden).

0723-0869/$ - see front matter © 2010 Elsevier GmbH. All rights reserved.
doi:10.1016/j.exmath.2010.03.001

several of these are described in Section 4. Furthermore, extensions to other spaces, say $L^p(\mathbb{R}^d)$ $(0 \le p < 1)$, Orlicz spaces, or compact groups, are described. Helly's theorem is often used as a replacement for Kolmogorov's compactness theorem, in particular in the context of nonlinear hyperbolic conservation laws, in spite of being more specialized (e.g., in the sense that its classical version requires one spatial dimension). For instance, Helly's theorem is an essential ingredient in Glimm's ground breaking existence proof for nonlinear hyperbolic systems [14]. We show below that Helly's theorem is an easy consequence of Kolmogorov's compactness theorem.

2. Preliminary results

An ε–*cover* of a metric space is a cover of the space consisting of sets of diameter at most ε. A metric space is called *totally bounded* if it admits a finite ε–cover for every $\varepsilon > 0$. It is well known that a metric space is compact if and only if it is complete and totally bounded (see, e.g., [34, p. 13]). Since we are interested in compactness results for subsets of Banach spaces, we may, and shall, concentrate our attention on total boundedness.

Here is the key lemma for many compactness results (in this lemma and its proof, every metric is named d):

Lemma 1. *Let X be a metric space. Assume that, for every $\varepsilon > 0$, there exists some $\delta > 0$, a metric space W, and a mapping $\Phi: X \to W$ so that $\Phi[X]$ is totally bounded, and whenever $x,y \in X$ are such that $d(\Phi(x),\Phi(y)) < \delta$, then $d(x,y) < \varepsilon$. Then X is totally bounded.*

Proof. For any $\varepsilon > 0$, pick δ, W and Φ as in the statement of the lemma. Since $\Phi[X]$ is totally bounded, there exists a finite δ–cover $\{V_1, \ldots, V_n\}$ of $\Phi[X]$. Then it immediately follows from the assumptions that $\{\Phi^{-1}(V_1), \ldots, \Phi^{-1}(V_n)\}$ is an ε- cover of X. Thus X is totally bounded. □

Lemma 1 embodies the main argument in the standard proof of the classical Arzelà–Ascoli theorem, as we now demonstrate.

Theorem 2 (*Arzelà–Ascoli*). *Let Ω be a compact topological space. Then a subset of $C(\Omega)$ is totally bounded in the supremum norm if, and only if,*

(i) *it is pointwise bounded, and*
(ii) *it is equicontinuous.*

Recall the definition of equicontinuity: Condition (ii) means that for every $x \in \Omega$ and every $\varepsilon > 0$ there is a neighborhood V of x so that $|f(y)-f(x)| < \varepsilon$ for all $y \in V$ and all f in the given set of functions.

Proof. Assume $\mathcal{F} \subset C(\Omega)$ is pointwise bounded and equicontinuous. Let $\varepsilon > 0$. Combining the equicontinuity of $\mathcal{F}$ and compactness of Ω, we can find a finite set of points $x_1, \ldots, x_n \in \Omega$ with neighborhoods $V_1, \ldots, V_n$ covering all of Ω so that $|f(x)-f(x_j)| < \varepsilon$ whenever $f \in \mathcal{F}$ and $x \in V_j$.

Define $\Phi : \mathcal{F} \to \mathbb{R}^n$ by

$$\Phi(f) = (f(x_1), \ldots, f(x_n)).$$

By the pointwise boundedness of $\mathcal{F}$, the image $\Phi[\mathcal{F}]$ is bounded, and hence totally bounded, in $\mathbb{R}^n$.

Furthermore, if $f,g \in \mathcal{F}$ with $\|\Phi(f)-\Phi(g)\|_\infty < \varepsilon$, then since any $x \in \Omega$ belongs to some V_j,

$$|f(x)-g(x)| \le |f(x)-f(x_j)| + |f(x_j)-g(x_j)| + |g(x_j)-g(x)| < 3\varepsilon,$$

and so $\|f-g\|_\infty \le 3\varepsilon$. By Lemma 1, $\mathcal{F}$ is totally bounded.

For the converse, assume that $\mathcal{F}$ is a totally bounded subset of $C(\Omega)$.

The existence of a finite ε–cover for $\mathcal{F}$, for any ε, clearly implies the boundedness of $\mathcal{F}$, thus establishing the uniform boundedness and hence also pointwise boundedness of $\mathcal{F}$.

To prove equicontinuity, let $x \in \Omega$ and $\varepsilon > 0$ be given. Pick an ε–cover $\{U_1, \ldots, U_n\}$ of $\mathcal{F}$, and choose $g_j \in U_j$ for $j = 1, \ldots, n$. Pick a neighborhood V_j of x so that $|g_j(y)-g_j(x)| < \varepsilon$ whenever $y \in V_j$, for

$j=1,\ldots,n$. Let $V=V_1\cap\cdots\cap V_n$. If $f\in U_j$ then $\|f-g_j\|_\infty\le\varepsilon$, and so when $y\in V$,

$$|f(y)-f(x)|\le|f(y)-g_j(y)|+|g_j(y)-g_j(x)|+|g_j(x)-f(x)|<3\varepsilon,$$

which proves equicontinuity. □

Remark 3. This theorem was first proved by Ascoli [3] for equi-Lipschitz functions and extended by Arzelà [2] to a general family of equicontinuous functions. See [4, p. 203].

We present the following theorem, first proved by Fréchet [12] for the case $p=2$, as a warm-up exercise, as the proof is short and nicely exposes some key ideas for the proof of Theorem 5.

Theorem 4. *A subset of l^p, where $1\le p<\infty$, is totally bounded if, and only if,*

(i) *it is pointwise bounded, and*
(ii) *for every $\varepsilon>0$ there is some n so that, for every x in the given subset,*

$$\sum_{k>n}|x_k|^p<\varepsilon^p.$$

Proof. Assume that $\mathcal{F}\subset l^p$ satisfies the two conditions. Given $\varepsilon>0$, pick n as in the second condition, and define a mapping $\Phi:\mathcal{F}\to\mathbb{R}^n$ by

$$\Phi(x)=(x_1,\ldots,x_n).$$

By the pointwise boundedness of $\mathcal{F}$, the image $\Phi(\mathcal{F})$ is totally bounded.

If $x,y\in\mathcal{F}$ with $|\Phi(x)-\Phi(y)|_p=(\sum_{k=1}^n|x_k-y_k|^p)^{1/p}<\varepsilon$, then

$$\|x-y\|_p\le\left(\sum_{k=1}^n|x_k-y_k|^p\right)^{1/p}+\left(\sum_{k>n}|x_k-y_k|^p\right)^{1/p}<\varepsilon+2\varepsilon=3\varepsilon.$$

By Lemma 1, $\mathcal{F}$ is totally bounded.

We will leave proving the converse as an exercise to the reader. The techniques from the proof of Theorem 2 are easily adapted. See also the proof of Theorem 5. □

3. The Kolmogorov–Riesz theorem

Theorem 5 (*Kolmogorov–Riesz*). *Let $1\le p<\infty$. A subset $\mathcal{F}$ of $L^p(\mathbb{R}^n)$ is totally bounded if, and only if,*

(i) *$\mathcal{F}$ is bounded,*
(ii) *for every $\varepsilon>0$ there is some R so that, for every $f\in\mathcal{F}$,*

$$\int_{|x|>R}|f(x)|^p dx<\varepsilon^p,$$

(iii) *for every $\varepsilon>0$ there is some $\rho>0$ so that, for every $f\in\mathcal{F}$ and $y\in\mathbb{R}^n$ with $|y|<\rho$,*

$$\int_{\mathbb{R}^n}|f(x+y)-f(x)|^p dx<\varepsilon^p.$$

Proof. Assume that $\mathcal{F}\subset L^p(\mathbb{R}^n)$ satisfies the three conditions. First, given $\varepsilon>0$, pick R as in the second condition, and ρ as in the third condition.

Let Q be an open cube centered at the origin so that $|y|<\frac{1}{2}\rho$ whenever $y\in Q$. Let $Q_1,\ldots,Q_N$ be mutually non-overlapping translates of Q so that the closure of $\bigcup_i Q_i$ contains the ball with radius R centered at the origin. Let P be the projection map of $L^p(\mathbb{R}^n)$ onto the linear span of the characteristic

functions of the cubes Q_i given by

$$Pf(x) = \begin{cases} \dfrac{1}{|Q_i|}\displaystyle\int_{Q_i} f(z)\,dz, & x \in Q_i, \quad i = 1,\ldots,N, \\ 0 & \text{otherwise.} \end{cases}$$

From (ii) and the definition of Pf we find, for $f \in \mathcal{F}$,

$$\|f-Pf\|_p^p < \varepsilon^p + \sum_{i=1}^{N} \int_{Q_i} |f(x)-Pf(x)|^p\,dx = \varepsilon^p + \sum_{i=1}^{N} \int_{Q_i} \left| \frac{1}{|Q_i|} \int_{Q_i} (f(x)-f(z))\,dz \right|^p dx.$$

Next we use Jensen's inequality and change a variable of integration, where we note that $x-z \in 2Q$ when $x,z \in Q_i$:

$$\begin{aligned} \|f-Pf\|_p^p &< \varepsilon^p + \sum_{i=1}^{N} \int_{Q_i} \frac{1}{|Q_i|} \int_{Q_i} |f(x)-f(z)|^p\,dz\,dx \le \varepsilon^p + \sum_{i=1}^{N} \int_{Q_i} \frac{1}{|Q_i|} \int_{2Q} |f(x)-f(x+y)|^p\,dy\,dx \\ &\le \varepsilon^p + \frac{1}{|Q|} \int_{2Q} \int_{\mathbb{R}^n} |f(x)-f(x+y)|^p\,dx\,dy < \varepsilon^p + \frac{1}{|Q|} \int_{2Q} \varepsilon^p\,dy = (2^n+1)\varepsilon^p \end{aligned}$$

by (iii). Thus $\|f-Pf\|_p < (2^n+1)^{1/p}\varepsilon$, and $\|f\|_p < (2^n+1)^{1/p}\varepsilon + \|Pf\|_p$. By the linearity of P, if $f,g \in \mathcal{F}$ and $\|Pf-Pg\|_p < \varepsilon$ then $\|f-g\|_p < ((2^n+1)^{1/p}+1)\varepsilon$. Moreover, since P is bounded (in fact $\|P\| = 1$) and $\mathcal{F}$ is bounded by (i), the image $P[\mathcal{F}]$ is bounded. Since the image of P is finite dimensional, $P[\mathcal{F}]$ is totally bounded. Thus $\mathcal{F}$ is totally bounded by Lemma 1.

For the converse, assume that $\mathcal{F}$ is totally bounded.

The existence of a finite ε–cover for $\mathcal{F}$, for any ε, clearly implies the boundedness of $\mathcal{F}$, thus establishing Condition (i).

To establish Condition (ii), let $\varepsilon > 0$ be given, let $\{U_1,\ldots,U_m\}$ be an ε–cover of $\mathcal{F}$, and chose $g_j \in U_j$ for $j = 1,\ldots,m$. Select R so that

$$\int_{x>R} |g_j(x)|^p dx < \varepsilon^p, \quad j = 1,\ldots,m.$$

If $f \in U_j$ then $\|f-g_j\|_p \le \varepsilon$, and so

$$\begin{aligned} \left(\int_{x>R} |f(x)|^p\,dx\right)^{1/p} &\le \left(\int_{x>R} |f(x)-g_j(x)|^p\,dx\right)^{1/p} + \left(\int_{x>R} |g_j(x)|^p\,dx\right)^{1/p} \\ &\le \|f-g_j\|_p + \left(\int_{x>R} |g_j(x)|^p\,dx\right)^{1/p} < 2\varepsilon, \end{aligned}$$

thus establishing Condition (ii).

Condition (iii) is established similarly, by noting that the inequality of the condition is easily established for any *single* function $f \in L^p(\mathbb{R}^n)$, for example using the fact that $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. Then, picking an ε–cover $\{U_1,\ldots,U_n\}$ and $g_j \in U_j$ for each j as in the previous paragraph, given $\varepsilon > 0$ we can find $\rho > 0$ with

$$\int_{\mathbb{R}^n} |g_j(x+y)-g_j(x)|^p\,dx < \varepsilon^p, \quad |y| < \rho, \quad j = 1,\ldots,m.$$

Again, if $f \in U_j$ we find

$$\begin{aligned} \left(\int_{\mathbb{R}^n} |f(x+y)-f(x)|^p\,dx\right)^{1/p} &\le \left(\int_{\mathbb{R}^n} |f(x+y)-g_j(x+y)|^p\,dx\right)^{1/p} + \left(\int_{\mathbb{R}^n} |g_j(x+y)-g_j(x)|^p\,dx\right)^{1/p} \\ &\quad + \left(\int_{\mathbb{R}^n} |g_j(x)-f(x)|^p\,dx\right)^{1/p} < 3\varepsilon, \end{aligned}$$

and the proof is complete. □

Remark 6. (I) A singleton set is clearly totally bounded, yet Condition (iii) is not obvious for a singleton set at first glance. However, it follows easily from the density of the space of smooth functions with compact support in L^p.

(II) In applications, one sometimes constructs a sequence $f_1, f_2, \ldots$ in L^p satisfying the first two conditions of Theorem 5 and the condition

$$\left(\int_{\mathbb{R}^n} |f_n(x+y)-f_n(x)|\,dx\right)^{1/p} < \alpha(y)+\beta(n), \quad \lim_{y\to 0}\alpha(y)=0, \quad \lim_{n\to\infty}\beta(n)=0.$$

Then for some N and $\delta > 0$, the right-hand side of the above inequality is less than ε for all $n > N$ and $|y|$ small enough. By the fact noted in the previous paragraph, we can choose a smaller upper bound for $|y|$ to make the integral smaller than ε for $n = 1,2,\ldots,N$. Thus $\{f_1, f_2, \ldots\}$ satisfies Condition (iii), and hence a convergent subsequence exists.

An interesting corollary to the Kolmogorov theorem is the following result, see [22], which also contains a variant using the uniform smoothness of the functions in $\mathcal{F}$ and their Fourier transforms. See also [7], which contains an alternate formulation based on the short-time Fourier transform, as well as one based on the wavelet transform.

Corollary 7. *Let $\mathcal{F} \subseteq L^2(\mathbb{R}^d)$ be such that $\sup_{f\in\mathcal{F}}\|f\|_2 \le M < \infty$. If*

$$\lim_{r\to\infty}\sup_{f\in\mathcal{F}}\int_{|x|\ge r}|f(x)|^2\,dx = 0 \quad \textit{and} \quad \lim_{\rho\to\infty}\sup_{f\in\mathcal{F}}\int_{|\xi|\ge\rho}|\hat{f}(\xi)|^2\,d\xi = 0,$$

then $\mathcal{F}$ is totally bounded in $L^2(\mathbb{R}^d)$.

Proof. We show that $\mathcal{F}$ satisfies the conditions of Theorem 5 for $p=2$. Clearly, Conditions (i) and (ii) are among our assumptions, so we only need to prove (iii). For $f \in \mathcal{F}$ we find:

$$\begin{aligned}\int_{\mathbb{R}^n}|f(x+y)-f(x)|^2\,dx &= \int_{\mathbb{R}^n}|(e^{i\xi\cdot y}-1)\hat{f}(\xi)|^2\,d\xi \le \int_{|\xi|<\rho}|(e^{i\xi\cdot y}-1)\hat{f}(\xi)|^2\,d\xi + 4\int_{|\xi|>\rho}|\hat{f}(\xi)|^2\,d\xi \\ &\le M^2\sup_{|\xi|<\rho}|e^{i\xi\cdot y}-1|^2+\varepsilon \quad \text{for } \rho \text{ big enough} \\ &< M^2\rho^2|y|^2+\varepsilon < 2\varepsilon\end{aligned}$$

if $|y| < \sqrt{\varepsilon}/(\rho M)$. Here ρ, and hence the upper bound on $|y|$, can be chosen independently of f. This shows Condition (iii) of Theorem 5 and finishes the proof. □

In the following result, $L^p_{\mathrm{loc}}(\Omega)$ is equipped with the topology of L^p convergence on compact subsets of Ω. Recall that Ω is the countable union of compacts, e.g., $\Omega = K_1 \cup K_2 \cup \ldots$ with $K_k = \{x \in \Omega : |x| \le k$ and $\mathrm{dist}(x, \mathbb{R}^n\backslash\Omega) \ge 1/k\}$. Moreover any compact subset of Ω is contained in some K_k, and so the topology on $L^p_{\mathrm{loc}}(\Omega)$ is given by the countable family of seminorms $\|f\|_k = \|f|_{K_k}\|_{L^p(K_k)}$. $L^p_{\mathrm{loc}}(\Omega)$ is complete with respect to the metric $(f,g) \mapsto \sum_{k=1}^{\infty}\min(2^{-k}, \|f-g\|_k)$.

Corollary 8. *Let $\Omega \subseteq \mathbb{R}^n$ be an open set. Write $f_K(x)=f(x)$ when $x \in K$, $f_K(x)=0$ otherwise. A subset $\mathcal{F} \subseteq L^p_{\mathrm{loc}}(\Omega)$ is totally bounded if, and only if, the following holds*:

(i) *For every compact $K \subset \Omega$ there is some M so that*

$$\int |f_K(x)|^p dx < M, \quad f \in \mathcal{F}.$$

(ii) *For every $\varepsilon > 0$ and every compact $K \subset \Omega$ there is some $\rho > 0$ so that*

$$\int |f_K(x+y)-f_K(x)|^p dx < \varepsilon^p, \quad f \in \mathcal{F}, \quad |y| < \rho.$$

Proof. Note that $\mathcal{F}$ is totally bounded in $L^p_{\mathrm{loc}}(\Omega)$ if and only if $\mathcal{F}_k = \{f_{K_k} : f \in \mathcal{F}\}$ is totally bounded for every k, with K_k as defined above. □

For the next result, recall that the *Sobolev space* $W^{k,p}(\mathbb{R}^n)$ is defined to consist of those measurable functions f which, together with all their distributional derivatives $D^\alpha f$ of order $|\alpha| \le k$, belong to $L^p(\mathbb{R}^n)$. Here $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index, i.e., each α_j is a nonnegative integer, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $D^\alpha = \partial^{|\alpha|}/(\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n})$. Finally, $W^{k,p}(\mathbb{R}^n)$ is equipped with the complete norm

$$\|f\|_{k,p} = \left(\int_{\mathbb{R}^n} \sum_{|\alpha| \le k} |D^\alpha f(x)|^p \, dx \right)^{1/p}.$$

Corollary 9. *A subset $\mathcal{F} \subseteq W^{k,p}(\mathbb{R}^n)$ is totally bounded if, and only if, the following holds*:

(i) *$\mathcal{F}$ is bounded, i.e., there is some M so that*

$$\int |D^\alpha f(x)|^p \, dx < M, \quad f \in \mathcal{F}, \quad |\alpha| \le k.$$

(ii) *For every $\varepsilon > 0$ there is some R so that*

$$\int_{|x| > R} |D^\alpha f(x)|^p \, dx < \varepsilon^p, \quad f \in \mathcal{F}, \quad |\alpha| \le k.$$

(iii) *For every $\varepsilon > 0$ there is some $\rho > 0$ so that*

$$\int_{\mathbb{R}^n} |D^\alpha f(x+y) - D^\alpha f(x)|^p \, dx < \varepsilon^p, \quad f \in \mathcal{F}, \quad |\alpha| \le k, \quad |y| < \rho.$$

Proof. Note that $\mathcal{F}$ is totally bounded in $W^{k,p}(\mathbb{R}^n)$ if and only if $D^\alpha[\mathcal{F}] = \{D^\alpha f : f \in \mathcal{F}\}$ is totally bounded in $L^p(\mathbb{R}^n)$ for every multi-index α with $|\alpha| \le k$. $\square$

4. A bit of history

In 1931, Kolmogorov [18] proved the first result in this direction. It characterizes compactness in $L^p(\mathbb{R}^n)$ for $1 < p < \infty$, in the case where all functions are supported in a common bounded set. Condition (iii) of Theorem 5 is replaced by the uniform convergence in L^p norm of spherical means of each function in the class to the function itself. (Clearly, our Condition (ii) is automatic in this case.)

Just a year later, Tamarkin [28] expanded this result to the case of unbounded supports by adding Condition (ii) of Theorem 5.

In 1933, Tulajkov [31] expanded the Kolmogorov–Tamarkin result to the case $p=1$.

In the same year, and probably independently, Riesz [25] proved the result for $1 \le p < \infty$, essentially in the form of our Theorem 5. Thus we feel somewhat justified in using the names Kolmogorov and Riesz in referring to the theorem, though we are perhaps being a bit unfair to Tamarkin and Tulajkov in doing so.

The compactness theorem has also seen generalizations in other directions.

Hanson [15] proved a necessary and sufficient condition for compactness of a family of measurable functions on a bounded measurable set, with respect to convergence in measure. (Here the measurable functions form a metric space in which the distance between two functions is the infimum of all $\varepsilon > 0$ so that the two functions differ by at most ε except on a set of measure $\le \varepsilon$.)

Fréchet [13] replaced Conditions (i) and (ii) of Theorem 5 with a single condition ("equisummability"), and generalized the theorem to arbitrary positive p.

Phillips [23, Theorem 3.7] proved a necessary and sufficient condition for compactness in L^p on a general measure space ($1 \le p \le \infty$), and indeed in *any* Banach space, which is however somewhat less suited to applications to PDEs. Nevertheless, our sufficiency proof for Theorem 5 is based on Phillips' criterion. (It is more common, albeit more involved, to use mollifiers in the proof.)

Weil [33] (see also [9, p. 269 ff]) extended the result to $L^p(G)$ where G is a locally compact group. Tsuji [30] considered the case of $L^p(\mathbb{R}^d)$ with $0 < p < 1$, and Takahashi [27] studied the same problem

in Orlicz spaces. A characterization of compact subsets of $L^p([0,T];\ B)$ (B a Banach space), which is very convenient in the context of time-dependent partial differential equations, is given by Simon [26] (see also [20]). A readable account of some of the historical development can be found in [8, p. 388]. Helly's theorem [16], which was published already in 1912, is easily seen to be a special case of Kolmogorov's compactness theorem in the one-dimensional case, see Section 6.

Further references include [32,17,5,6,11,21].

5. The Rellich–Kondrachov theorem

In this section we use Kolmogorov's theorem to prove a simple variant of the Rellich–Kondrachov theorem [24,19]. Our simplification consists in avoiding boundary regularity conditions by working on the entire space $\mathbb{R}^n$. The standard Rellich–Kondrachov theorem requires a bounded region. The present version replaces this by a uniform decay estimate, specially tailored to fit the framework of the present paper.

The Sobolev norm $\|f\|_{1,p}$ on $W^{1,p}(\mathbb{R}^n)$ is defined by

$$\|f\|_{1,p} = \left(\int_{\mathbb{R}^n} (|f(x)|^p + |\nabla f(x)|_p^p)\,dx\right)^{1/p}, \quad |\nabla f|_p = \left(\sum_j \left|\frac{\partial f}{\partial x_i}\right|^p\right)^{1/p}.$$

According to the Sobolev embedding theorem, if $p<n$ then $W^{1,p}(\mathbb{R}^n) \subset L^q(\mathbb{R}^n)$, and the inclusion map is bounded, for any q satisfying $p \le q \le p^*$, where p^* is the *conjugate Sobolev exponent*:

$$\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}.$$

To see where this exponent comes from, consider a function f and its scalings $f^\lambda(x) = f(x/\lambda)$ where $\lambda > 0$, and note that $\|f^\lambda\|_p = \lambda^{n/p}\|f\|_p$ and $\|\nabla f^\lambda\|_p = \lambda^{n/p-1}\|\nabla f\|_p$, so the inclusion map $W^{1,p} \to L^q$ can only be bounded if there exists a constant C with $\lambda^{n/q} \le C(\lambda^{n/p} + \lambda^{n/p-1})$ for all $\lambda > 0$. In the limits $\lambda \to \infty$ and $\lambda \to 0$ we conclude $n/q \le n/p$ and $n/q \ge n/p - 1$ respectively.

Theorem 10. *Assume $p<n$ and $p \le q < p^*$, and let $\mathcal{F}$ be a bounded subset of $W^{1,p}(\mathbb{R}^n)$. Assume that for every $\varepsilon > 0$ there exists some R so that, for every $f \in \mathcal{F}$,*

$$\int_{|x|>R} (|f(x)|^p + |\nabla f(x)|_p^p)\,dx < \varepsilon^p.$$

Then $\mathcal{F}$ is a totally bounded subset of $L^q(\mathbb{R}^n)$.

Proof. We shall show that $\mathcal{F}$ satisfies the hypotheses of Theorem 5, with p replaced by q. We shall use the Sobolev embedding inequality $\|f\|_q \le C\|f\|_{1,p}$, where the constant C depends only on p, q and n, and which is valid under the stated assumption, see [1, 4.30 (p. 101) and Theorem 4.12 I C (p. 85) with $j=0$, $k=n$, $m=1$]. Condition (i) of Theorem 5 follows immediately from the Sobolev embedding inequality. Condition (ii) is almost equally immediate, from applying the Sobolev embedding inequality to the function $x \mapsto f(x)\chi(|x|-R)$, where $\chi \in C^\infty(\mathbb{R})$, $0 \le \chi \le 1$, $\chi(x) = 0$ for $x<0$ and $\chi(x) = 1$ for $x>1$.

If we apply the Sobolev embedding inequality to the function $x \mapsto f(x/\lambda)$ where $\lambda > 0$ and change variables in the resulting integrals, we obtain

$$\lambda^{n/q}\|f\|_q \le C\left(\lambda^n \int_{\mathbb{R}^n} |f(x)|^p\,dx + \lambda^{n-p}\int_{\mathbb{R}^n} |\nabla f(x)|_p^p\,dx\right)^{1/p}. \tag{1}$$

We shall apply the above inequality not to f, but to $x \mapsto f(x+y) - f(x)$, where $f \in \mathcal{F}$.

Now let $\varepsilon > 0$ be given. By picking λ sufficiently large we can ensure that

$$C\left(\lambda^{n-p}\int_{\mathbb{R}^n} (|\nabla f(x+y) - \nabla f(x)|_p^p)\,dx\right)^{1/p} \le \varepsilon\lambda^{n/q} \tag{2}$$

for all $f \in \mathcal{F}$, since the integral in this expression is bounded uniformly for $f \in \mathcal{F}$.

Next, we find (using the Jensen and Hölder inequalities, then Fubini's theorem)

$$\int_{\mathbb{R}^n} |f(x+y)-f(x)|^p\,dx = \int_{\mathbb{R}^n} \left| \int_0^1 y\cdot\nabla f(x+ty)\,dt \right|^p dx$$
$$\le |y|_{p'}^p \int_0^1 \int_{\mathbb{R}^n} |\nabla f(x+ty)|_p^p\,dx\,dt = |y|_{p'}^p \int_{\mathbb{R}^n} |\nabla f(x)|_p^p\,dx,$$

(where p and p' are conjugate exponents) for any test function f, and hence for any $f \in W^{1,p}$. The integrals on the right-hand side of this inequality are uniformly bounded for $f \in \mathcal{F}$, and so we can find some $\delta > 0$ so that $|y| < \delta$ implies

$$C\left(\lambda^n \int_{\mathbb{R}^n} |f(x+y)-f(x)|^p\,dx\right)^{1/p} \le \varepsilon\lambda^{n/q}. \tag{3}$$

For such y and f, (1) applied to $x \mapsto f(x+y)-f(x)$ combined with (2) and (3) yield

$$\lambda^{n/q}\|f(\cdot+y)-f(\cdot)\|_q \le 2^{1/p}\varepsilon\lambda^{n/q},$$

and so assumption (iii) of Theorem 5 is satisfied. □

6. Helly's theorem

Helly's theorem is often referred to as *Helly's selection principle*, in order to avoid confusion with another theorem by Helly, stating that, given a collection of convex sets in $\mathbb{R}^n$ so that any $n+1$ of them have a point in common, then any finite subcollection has nonempty intersection. Helly's selection principle is essentially a corollary of the Kolmogorov–Riesz theorem, though historically it was not derived that way.

Recall that an integrable function f on the line is of bounded variation if it has finite essential or total variation, that is, if

$$\mathrm{TV}(f) = \sup \sum_{j=1}^{m} |f(x_{j+1})-f(x_j)| < \infty,$$

where the supremum is taken over all finite partitions $x_j < x_{j+1}$ such that each x_j is a point of approximate continuity of f (that is, $\delta^{-1}|\{x : |x-x_j| < \delta, |f(x)-f(x_j)| \ge \varepsilon\}| \to 0$ for every $\varepsilon > 0$ as $\delta \to 0$. See, e.g., [10, p. 47]). We need a lemma:

Lemma 11. *Let u be function of bounded variation on $\mathbb{R}$. Then*

$$\int_{-\infty}^{\infty} |u(x+y)-u(x)|\,dx \le |y|\mathrm{TV}(u)$$

for all $y \in \mathbb{R}$.

Proof. We may assume $y > 0$ without loss of generality. The calculation

$$\int_{-\infty}^{\infty} |u(x+y)-u(x)|\,dx = \sum_{j=-\infty}^{\infty} \int_0^y |u(x+jy+y)-u(x+jy)|\,dx$$
$$= \int_0^y \sum_{j=-\infty}^{\infty} |u(x+(j+1)y)-u(x+jy)|\,dx \le \int_0^y \mathrm{TV}(u)\,dx = y\,\mathrm{TV}(u),$$

finishes the proof. □

Theorem 12 (*Helly*). *Let (u_n) be a sequence of functions of bounded variation on the bounded real interval $[a,b]$. If there is a constant M so that $\mathrm{TV}(u_n) \le M$ and $\|u_n\|_\infty \le M$ for all n, then there is a subsequence of (u_n) which converges pointwise everywhere and in L^1 norm in $[a,b]$ to a function of bounded variation.*

Proof. Extend each function u_n to all of $\mathbb{R}$ by setting it to zero outside $[a,b]$. By Lemma 11, the set of all the functions u_n satisfy Condition (iii) of Theorem 5 (with $p=1$), while (i) holds by assumption and

(ii) is trivial. Hence there is a subsequence of (u_n) which converges in $L^1([a,b])$. Moreover, integration theory tells us that we also get pointwise convergence almost everywhere, possibly after passing to a subsequence once more. However, this is not quite enough.

Write instead $u_n = v_n - w_n$ where each v_n, w_n is an non-decreasing function: $v_n(x)$ is $u_n(a)$ plus the positive variation of u_n on the interval $[a,x]$, and $w_n(x)$ is the negative variation on the same interval. Then the sequences (v_n) and (w_n) both satisfy the conditions of the present theorem, and so, by the result of the previous paragraph, we may pass to a subsequence so that (v_n) and (w_n) both converge in $L^1([a,b])$, as well as pointwise almost everywhere.

Let v be the limit of the sequence (v_n). Clearly, v is non-decreasing on the set where pointwise convergence holds, and so we may assume that v is non-decreasing everywhere, after possibly redefining it on a set of measure zero.

Now it is clear that $v_n(x) \to v(x)$ for any point of continuity x for v: Given $\varepsilon > 0$, pick $\delta > 0$ so that $|y-x| < \delta$ implies $|v(y)-v(x)| < \varepsilon$, let $x-\delta < y < x < z < x+\delta$ with $v_n(y) \to v(y)$ and $v_n(z) \to v(z)$, and note that for n large enough we get $v(x)-2\varepsilon < v(y)-\varepsilon < v_n(y) \le v_n(x) \le v_n(z) < v(z)+\varepsilon < v(x)+2\varepsilon$, so that $|v_n(x)-v(x)| < 2\varepsilon$.

Since v has at most a countable number of discontinuities, a diagonal argument yields a further subsequence which converges at all the discontinuities of v as well, and so we have pointwise convergence everywhere.

In the same way we show that $w_n(x) \to w(x)$ for all x. Thus $u_n \to v-w$ pointwise, and v-w has bounded variation. □

Remark 13. The above proof is probably not the most natural one, but it does make clear the connection with the Kolmogorov–Riesz theorem. In a sense L^1 convergence is irrelevant: Pointwise convergence is the key, and L^1 convergence follows from the bounded convergence theorem.

It should be noted, however, that Helly's theorem, without pointwise convergence, is also true in higher dimensions [10, p. 176].

A recent generalization of Helly's selection principle (in one dimension) can be found in [29].

References

[1] A. Adams, J.F. Fournier, Sobolev Spaces, second ed., Academic Press, Amsterdam, 2003.
[2] C. Arzelà, Sulle funzione de linee, Mem. Accad. Sci. Bologna 5–5 (1894/95) 225–244 (in Italian).
[3] G. Ascoli, Le curve limiti di una varietá data di curve, Rend. Accad. Lincei 18 (1884) 521–586 (in Italian).
[4] E. DiBenedetto, Real Analysis, Birkhäuser, Boston, 2002.
[5] G. Bruno, R. Grande, A compactness criterion in B^q_{ap} spaces, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. 20 (1996) 95–121.
[6] G. Bruno, R. Grande, Compact embedding theorems for Sobolev–Besicovitch spaces of almost periodic functions, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. 20 (1996) 157–173.
[7] M. Dörfler, H.G. Feichtinger, K. Gröchenig, Compactness criteria in function spaces, Coll. Math. 94 (2002) 37–50.
[8] N. Dunford, J.T. Schwartz, Linear Operators. Part I: General Theory, Wiley, New York, 1988.
[9] R.E. Edwards, Functional Analysis, Holt, Rinehart and Winston, Chicago, 1965.
[10] L.C. Evans, R.F. Gariepy, Measure Theory and Fine Properties of Functions, CRC Press, Boca Raton, 1992.
[11] H.G. Feichtinger, Compactness in translation invariant Banach spaces of distributions and compact multipliers, J. Math. Anal. Appl. 102 (1984) 289–327.
[12] M. Fréchet, Essai de geometrie analytique, Nouv. Ann. Math. 4 (1908) 97–116 289–317 (in French).
[13] M. Fréchet, Sur les ensembles compacts de fonctions de carrés sommables, Acta Litt. Sci. Szeged. 8 (1937) 116–126 (in French).
[14] J. Glimm, Solutions in the large for nonlinear hyperbolic systems of equations, Comm. Pure Appl. Math. 18 (1965) 697–715.
[15] E.H. Hanson, A note on compactness, Bull. Am. Math. Soc. 39 (1933) 397–400.
[16] E. Helly, Über lineare Funktionaloperationen, Österreich. Akad. Wiss. Math.-Natur. Kl. S.-B. IIa 121 (1912) 265–297 (in German).
[17] S. Izumi, On the compactness of a class of functions, Proc. Imp. Acad. Tokyo 15 (1939) 111–113.
[18] A.N. Kolmogorov, Über Kompaktheit der Funktionenmengen bei der Konvergenz im Mittel, Nachr. Ges. Wiss. Göttingen 9 (1931) 60–63 (English translation: On the compactness of sets of functions in the case of convergence in the mean, in: V.M. Tikhomirov (Ed.), Selected Works of A.N. Kolmogorov, vol. I, Kluwer, Dordrecht, 1991, pp. 147–150)..
[19] W. Kondrachov, Sur certaines propriétés des fonctions dans l'espace, C. R. Acad. Sci. URSS., n. Sér 48 (1945) 535–538.
[20] E. Maitre, On a nonlinear compactness lemma in $L^p(0,T;B)$, Int. J. Math. Math. Sci. 27 (2003) 1725–1730.
[21] M. Nicolescu, On the criterion of compactness of A. Kolmogorov, Acad. Repub. Pop. Române. Bul. Şti. Ser. Math. Fiz. Chim. 2 (1950) 407–415 (in Romanian).

[22] R.L. Pego, Compactness in L^2 and the Fourier transform, Proc. Am. Math. Soc. 95 (1985) 252–254.
[23] R.S. Phillips, On linear transforms, Trans. Am. Math. Soc. 48 (1940) 516–541.
[24] F. Rellich, Ein Satz über mittlere Konvergenz, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1930) 30–35.
[25] M. Riesz, Sur les ensembles compacts de fonctions sommables, Acta Szeged Sect. Math. 6 (1933) 136–142 (in French). Also in L. Gårding, L. Hörmander (Eds.), Marcel Riesz—Collected Papers, Springer, Berlin (1988).
[26] J. Simon, Compact sets in the space $L^p(0,T;B)$, Ann. Mat. Pura Appl. 146 (1987) 65–96.
[27] T. Takahashi, On the compactness of the function-set by the convergence in mean of general type, Studio Math. 5 (1934) 141–150.
[28] J.D. Tamarkin, On the compactness of the space L_p, Bull. Am. Math. Soc. 32 (1932) 79–84.
[29] Yu.V. Tret'yachenko, V.V. Chistyakov, Selection principle for pointwise bounded sequences of functions, Math. Notes 84 (2008) 396–406.
[30] M. Tsuji, On the compactness of space L^p $(p > 0)$ and its application to integral operators, Kodai Math. J. (1951) 33–36.
[31] A. Tulajkov, Zur Kompaktheit im Raum L_p für $p=1$., Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl. I 1933, nr. 39, 167–170 (in German).
[32] P. Veress, Über Functionenmengen, Acta Sci. Math. (Szeged) 3 (1927) 177–192 (in German).
[33] A. Weil, L'intégration dans les groupes topologiques et ses applications, Hermann et Cie., Paris, 1940.
[34] K. Yosida, Functional Analysis, Springer, Berlin, 1980.