

Weak parallelogram laws on Banach spaces and applications to prediction

R. Cheng · W. T. Ross

Published online: 6 January 2015
© Akadémiai Kiadó, Budapest, Hungary 2015

Abstract This paper concerns a family of weak parallelogram laws for Banach spaces. It is shown that the familiar Lebesgue spaces satisfy a range of these inequalities. Connections are made to basic geometric ideas, such as smoothness, convexity, and Pythagorean-type theorems. The results are applied to the linear prediction of random processes spanning a Banach space. In particular, the weak parallelogram laws furnish coefficient growth estimates, Baxter-type inequalities, and criteria for regularity.

Keywords Parallelogram law · Pythagorean theorem · Uniform convexity · Best predictor · Baxter's inequality · Purely nondeterministic

Mathematics Subject Classification 46B20 · 46B25 · 60G25

1 Introduction

The familiar parallelogram law states that for any vectors x and y in a Hilbert space \mathcal{H} , we have

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad (1.1)$$

If this condition is imposed on a normed space, then in fact the polarization identity

$$\langle x, y \rangle = \frac{\|x + y\|^2 - \|x - y\|^2}{4} + i \frac{\|ix - y\|^2 - \|ix + y\|^2}{4}$$

R. Cheng (✉)
Department of Mathematics and Statistics, Old Dominion University,
Engineering and Computational Sciences Building, Norfolk, VA 23529, USA
e-mail: rcheng@odu.edu

W. T. Ross
Department of Mathematics and Computer Science, University of Richmond,
Richmond, VA 23173, USA
e-mail: wross@richmond.edu

(assuming complex scalars) determines an inner product, with $\|x\|^2 = \langle x, x \rangle$ for all vectors x . Thus, the parallelogram law would have to be altered or weakened in some way in order to apply to a more general normed space.

One example of this is furnished by Clarkson’s inequalities, which constitute parallelogram laws of sorts for the spaces $L^p = L^p(\Omega, \Sigma, \mu)$, where $1 < p < \infty$ and (Ω, Σ, μ) is any measure space (see, for instance, [7, p. 119]). If $1 < p \leq 2$, then

$$\|x + y\|_p^p + \|x - y\|_p^p \geq 2^{p-1} (\|x\|_p^p + \|y\|_p^p) \tag{1.2}$$

for all x and y in L^p ; and if $2 \leq p < \infty$, then

$$\|x + y\|_p^p + \|x - y\|_p^p \leq 2^{p-1} (\|x\|_p^p + \|y\|_p^p) \tag{1.3}$$

for all x and y in L^p . Another example comes from Bynum and Drew [5] and Bynum [4]. They discovered what they call weak parallelogram laws for L^p :

If $1 < p \leq 2$, then

$$\|x + y\|_p^2 + (p - 1)\|x - y\|_p^2 \leq 2 (\|x\|_p^2 + \|y\|_p^2) \tag{1.4}$$

for all x and y in L^p ; and if $2 \leq p < \infty$, then

$$\|x + y\|_p^2 + (p - 1)\|x - y\|_p^2 \geq 2 (\|x\|_p^2 + \|y\|_p^2) \tag{1.5}$$

for all x and y in L^p . That is, they are able to impose a version of condition (1.1) on the space L^p , at the cost of introducing a constant factor $(p - 1)$, and weakening the equation to an inequality. In both examples, the resulting inequalities tell us something about the geometry of the space, such as smoothness and convexity properties.

Guided by these two examples, and in the interest of pursuing a parallelogram law for general normed linear spaces, let us adopt the following terminology.

Definition 1.1 Let $C > 0$, and $1 < p < \infty$. A Banach Space \mathcal{X} satisfies a p -lower weak parallelogram law with constant C if

$$\|x + y\|^p + C\|x - y\|^p \leq 2^{p-1} (\|x\|^p + \|y\|^p) \tag{1.6}$$

for all x and y in \mathcal{X} . In this case let us say that \mathcal{X} is p -LWP(C).

Similarly, a Banach Space \mathcal{X} satisfies a p -upper weak parallelogram law with constant C if

$$\|x + y\|^p + C\|x - y\|^p \geq 2^{p-1} (\|x\|^p + \|y\|^p) \tag{1.7}$$

for all x and y in \mathcal{X} . In this case let us say that \mathcal{X} is p -UWP(C).

It may be convenient during usage to suppress the parameter p or the constant C . Naturally, we speak of Banach spaces satisfying LWP or UWP as weak parallelogram spaces. By this terminology, the Clarkson inequalities say that L^p is p -UWP(1) when $1 < p \leq 2$, and L^p is p -LWP(1) when $2 \leq p < \infty$. Similarly, Bynum’s inequalities assert that L^p is 2-LWP($p - 1$) when $1 < p \leq 2$, and L^p is 2-UWP($p - 1$) when $2 \leq p < \infty$.

In this paper, our objectives are to obtain some properties of weak parallelogram spaces, and to apply those ideas toward the prediction of certain processes. First, a complete description is obtained for the values of r and p for which the space L^r satisfies p -LWP(C) for some $C > 0$, and respectively p -UWP(C).

The geometry of weak parallelogram spaces is then explored. We note that LWP spaces are uniformly convex, and UWP spaces are uniformly smooth. We also derive Pythagorean-type

theorems for weak parallelogram spaces. In this context, orthogonality is in the Birkhoff–James sense: two vectors x and y in \mathcal{X} satisfy $x \perp y$ if

$$\|x + ay\| \geq \|x\|$$

for all scalars a . (See [1] for a recent review of Birkhoff–James orthogonality in normed linear spaces).

Weak parallelogram spaces are shown to have the following properties. If \mathcal{X} is p -LWP, then there exists a constant $K > 0$ such that

$$\|x + y\|^p \geq \|x\|^p + K\|y\|^p$$

whenever $x \perp y$; similarly, if \mathcal{X} is p -UWP, then there exists a constant $K > 0$ such that

$$\|x + y\|^p \leq \|x\|^p + K\|y\|^p$$

Indeed, if $p = 2$ and $K = 1$, and equality holds, then we have the usual Pythagorean Theorem.

These ideas are used to study certain sequences of vectors in a Banach space. Specifically, we are interested in sequences $\{X_n\}_{n=0}^\infty$ of unit vectors for which $m < n$ implies that $X_m \perp X_n$ in the Birkhoff–James sense. This structure is motivated by prediction problems for p -stationary processes, which were likewise extensions of prediction theory for stationary Gaussian processes (i.e., the $p = 2$ case). For examples of processes of this type, see [6, 9–11, 14]. For processes in the broader class of weak parallelogram spaces, bounds are obtained for the coefficients $\{a_n\}$ whenever $\sum_{n=0}^\infty a_n X_n$ converges in norm.

In addition, a Baxter-type inequality is derived for LWP spaces. This gives us a practical bound for the average error incurred when the best estimate of X_0 in the linear span of $\{X_1, X_2, X_3, \dots\}$, truncated at the n th term, is used in place of the best estimate of X_0 in the linear span of $\{X_1, X_2, X_3, \dots, X_n\}$. The original Hilbert space version due to Baxter [2] has proved very useful in time series analysis [12, 16]. Our work extends this notion to a broader class of processes.

Finally, a number of criteria are given for such sequences $\{X_n\}_{n=0}^\infty$ to be regular (in the sense of being purely linearly nondeterministic), if the underlying space is LWP. For instance, regularity of the sequence is shown to be equivalent to $\{X_n\}_{n=0}^\infty$ being a conditional basis for its span.

We conclude this section with some notational and technical preliminaries. All Banach spaces being considered here are assumed to be of dimension 2 or greater, so as to avoid trivialities. The function space $L^p(\Omega, \Sigma, \mu)$ is simply referred to as L^p , there being no need to refer to the underlying measure space. The functions may be real or complex valued. Typically, it is assumed that $1 < p < \infty$, so that L^p is a reflexive Banach space. As usual, the norm in L^p is written $\|\cdot\|_p$. In this setting, let p' denote the conjugate index to p : thus $(1/p) + (1/p') = 1$.

If x is a nonzero vector in a Banach space \mathcal{X} , then a *norming functional* for x is a norm one functional T in \mathcal{X}^* such that $T(x) = \|x\|$. The Hahn–Banach theorem assures the existence of such a functional. If the norming functional for x is unique, then let us call it T_x . If every nonzero vector x in \mathcal{X} has a unique norming functional, then \mathcal{X} is said to be *smooth*. For example, the space L^p is smooth whenever $1 < p < \infty$, but fails to be smooth when $p = 1$ or $p = \infty$. Here are some well established facts about smoothness.

Proposition 1.2 *The following statements about a Banach space \mathcal{X} are equivalent: (i) \mathcal{X} is smooth; (ii) the norm in \mathcal{X} has directional derivatives at every nonzero point; and (iii)*

Birkhoff–James orthogonality \perp in \mathcal{X} is linear in its second argument. Furthermore, if \mathcal{X} is smooth, then $x \perp y$ in \mathcal{X} if and only if $T_x(y) = 0$.

These properties are particularly relevant to the linear prediction problems of the later sections. For the proofs, see [3] for the equivalence of (i) and (ii); and [13] for the equivalence of (i) and (iii).

Other natural and interesting questions arise. One might consider, for example, properties of and relationships between the parameters p and C in the condition p -LWP(C) or p -UWP(C). In addition, the weak parallelogram properties exhibited by the L^p spaces suggest a more general duality relationship. Furthermore, connections could be established between the weak parallelogram laws and other ways of expressing a parallelogram law on a general normed space, for example, the Rademacher type and co-type properties. In order to maintain our present focus on applications, we defer these issues to another paper [8].

2 Weak parallelogram laws for L^p

The inequalities (1.2), (1.3), (1.4) and (1.5) show that for $1 < p < \infty$, the L^p spaces satisfy certain weak parallelogram laws. Our aim in the this section is to identify all of the weak parallelogram laws satisfied by the L^p spaces.

Theorem 2.1 *If $1 < p \leq 2$, and p' is the conjugate index, then L^p is:*

$$\begin{aligned} & r - \text{UWP}(1) \quad \text{when } 1 < r \leq p; \\ & r - \text{LWP}((p - 1)^{r/2}) \quad \text{when } 2 \leq r \leq p'; \text{ and} \\ & r - \text{LWP}(1) \quad \text{when } r \geq p'. \end{aligned}$$

If $2 \leq p < \infty$, and p' is the conjugate index, then L^p is:

$$\begin{aligned} & r - \text{LWP}(1) \quad \text{when } p \leq r < \infty; \\ & r - \text{UWP}((p - 1)^{r/2}) \quad \text{when } p' \leq r \leq 2; \text{ and} \\ & r - \text{UWP}(1) \quad \text{when } r \leq p'. \end{aligned}$$

To prove this, we begin by extending Clarkson’s inequalities. There are eight cases, presented in two lemmas.

Lemma 2.2 *Suppose that $1 < p < \infty$ and $1 < r < \infty$.*

(i) *If $r \geq \max\{p, p'\}$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + \|x - y\|_p^r \leq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right) \tag{2.1}$$

(ii) *If $r \leq \min\{p, p'\}$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + \|x - y\|_p^r \geq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right) \tag{2.2}$$

(iii) *If $1 < p \leq r \leq p' < \infty$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + \|x - y\|_p^r \leq 2^{r/p} \left(\|x\|_p^r + \|y\|_p^r \right) \tag{2.3}$$

(iv) *If $1 < p' \leq r \leq p < \infty$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + \|x - y\|_p^r \geq 2^{r/p} \left(\|x\|_p^r + \|y\|_p^r \right) \tag{2.4}$$

Proof In the case $2 \leq r \leq p < \infty$, we use the inequality (1.3) to get

$$\begin{aligned} & \left(\|x + y\|_p^r + \|x - y\|_p^r \right)^{p/r} \\ & \leq \|x + y\|_p^p + \|x - y\|_p^p \\ & \leq 2^{p-1} (\|x\|_p^p + \|y\|_p^p) \\ & = 2^{p-1} (\|x\|_p^p \cdot 1 + \|y\|_p^p \cdot 1) \\ & \leq 2^{p-1} \left(\|x\|_p^r + \|y\|_p^r \right)^{p/r} \left(1^{r/(r-p)} + 1^{r/(r-p)} \right)^{(r-p)/r} \end{aligned}$$

The last step comes from applying the Hölder inequality to two-dimensional ℓ^1 , using the exponent r/p and its conjugate. Now raise both sides to the power r/p , noting that on the right hand side, 2 is raised to the power

$$\frac{r(p-1)}{p} + \frac{r(r-p)}{rp} = r-1$$

This proves part of case (i). If $2 \leq p' \leq r < \infty$, then

$$\begin{aligned} & 2 \left(\|x\|_p^r + \|y\|_p^r \right)^{p'(p-1)/r} \\ & \leq 2 \left(\|x\|_p^{p'} + \|y\|_p^{p'} \right)^{p-1} \\ & \leq \|x + y\|_p^p + \|x - y\|_p^p \\ & \leq \left(\|x + y\|_p^r + \|x - y\|_p^r \right)^{p/r} 2^{1-p/r} \end{aligned} \quad (2.5)$$

where the step from line (2.5) to the next is yet another of Clarkson's inequalities [7, p. 119].

It follows that

$$2 \left(\|x\|_p^r + \|y\|_p^r \right) \leq \|x + y\|_p^r + \|x - y\|_p^r$$

Now replace x and y with $x + y$ and $x - y$, respectively, to see that this is equivalent to (2.1). This completes the verification of (i). The rest is similar. \square

Indeed, when $p = r$, these are the original Clarkson's Inequalities, and the special case $p = r = 2$ is the parallelogram law proper. However, only cases (i) and (ii) above are weak parallelogram laws as defined here in (1.6) and (1.7). The other two cases are near-misses since the multiplicative constant $2^{r/p}$ is of the wrong form. To obtain further results, we need the following extension of the Bynum and Drew inequalities.

Lemma 2.3 *Suppose that $1 < r < \infty$ and $1 < p < \infty$.*

(v) *If $1 < p \leq r \leq 2$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + (p-1)^{r/2} \|x - y\|_p^r \leq 2 \left(\|x\|_p^r + \|y\|_p^r \right) \quad (2.6)$$

(vi) *If $1 < p \leq 2 \leq r \leq p'$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + (p-1)^{r/2} \|x - y\|_p^r \leq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right) \quad (2.7)$$

(vii) *If $2 \leq r \leq p < \infty$, then for all x and y in L^p ,*

$$\|x + y\|_p^r + (p-1)^{r/2} \|x - y\|_p^r \geq 2 \left(\|x\|_p^r + \|y\|_p^r \right) \quad (2.8)$$

(viii) If $p' \leq r \leq 2 \leq p < \infty$, then for all x and y in L^p ,

$$\|x + y\|_p^r + (p - 1)^{r/2} \|x - y\|_p^r \geq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right) \tag{2.9}$$

Proof For (v), use Bynum’s weak parallelogram law (1.4) for L^p , along with Hölder’s Inequality on two-dimensional $\ell^{2/r}$, as follows:

$$\begin{aligned} & \|x + y\|_p^r + (p - 1)^{r/2} \|x - y\|_p^r \\ & \leq \left(\|x + y\|_p^2 + (p - 1) \|x - y\|_p^2 \right)^{r/2} (1 + 1)^{1-r/2} \\ & \leq \left(2[\|x\|_p^2 + \|y\|_p^2] \right)^{r/2} 2^{1-r/2} \\ & \leq 2 \left(\|x\|_p^r + \|y\|_p^r \right) \end{aligned}$$

The other assertions are similarly handled. □

Note that when $r = 2$ we indeed get the inequalities of Bynum [4] and Bynum and Drew [5].

Cases (vi) and (viii) are weak parallelogram laws. It will turn out that for the remaining cases (v) and (vii), which are again near-misses, we can rule out any r -weak parallelogram laws in principle. (See Proposition 3.6.) This exhausts all of the possibilities for p and r , and affirms Theorem 2.1. Note that L^r is p -LWP if and only if $L^{r'}$ is p' -UWP; a more general duality relationship is established in [8]. This completes the description of p -weak parallelogram laws for the L^r spaces.

3 Geometry of weak parallelogram spaces

Let us explore some of the the geometric consequences of the weak parallelogram laws. First, we note the following convexity and smoothness properties. The proofs follow those of [4] for the $p = 2$ case, mutatis mutandis.

Proposition 3.1 *Let $1 < p < \infty$. If a Banach space \mathcal{X} satisfies p -LWP(C), then \mathcal{X} is uniformly convex. If a Banach space \mathcal{X} satisfies p -UWP(C), then \mathcal{X} is uniformly smooth.*

Thus, a LWP space is reflexive, and it enjoys the unique nearest point property by the Milman-Pettis Theorem [15].

Next, suppose that \mathcal{X} is a smooth Banach space, so that for each nonzero x in \mathcal{X} , there is a unique norm one functional T_x satisfying $T_x(x) = \|x\|$. We establish below that weak parallelogram laws on smooth Banach spaces are expressible as bounding conditions on the norming functionals. Here $\Re z$ stands for the real part of a complex number z .

Lemma 3.2 *Let $1 < p < \infty$.*

A smooth Banach space \mathcal{X} is p -LWP if and only if for some positive constant K , and for all $x \neq 0$ and y in \mathcal{X} ,

$$\|x + y\|^p \geq \|x\|^p + K \|y\|^p + p \|x\|^{p-1} \Re(T_x(y)) \tag{3.1}$$

A smooth Banach space \mathcal{X} is p -UWP if and only if for some positive constant K , and for all $x \neq 0$ and y in \mathcal{X} ,

$$\|x + y\|^p \leq \|x\|^p + K \|y\|^p + p \|x\|^{p-1} \Re(T_x(y)) \tag{3.2}$$

Proof Suppose that (3.1) holds. Then

$$\begin{aligned}\|x\|^p + K\|y\|^p + p\|x\|^{p-1}\mathfrak{R}(T_x(y)) &\leq \|x + y\|^p \\ \|x\|^p + K\| - y\|^p + p\|x\|^{p-1}\mathfrak{R}(T_x(-y)) &\leq \|x - y\|^p \\ 2\|x\|^p + 2K\|y\|^p &\leq \|x + y\|^p + \|x - y\|^p\end{aligned}$$

Replace x with $x + y$, and replace y with $x - y$, to find that

$$\|x + y\|^p + K\|x - y\|^p \leq 2^{p-1} (\|x\|^p + \|y\|^p) \quad (3.3)$$

Thus \mathfrak{X} is p -LWP(K).

Conversely, assume that \mathfrak{X} is p -LWP(C) for some $C > 0$, and apply

$$\|u + v\|^p + C\|u - v\|^p \leq 2^{p-1} (\|u\|^p + \|v\|^p)$$

to the pair of vectors $u = x$ and $v = x + 2^{-n}y$, for $n = 0, 1, 2, \dots$. The first step, with $n = 0$, provides that

$$\|x + \frac{1}{2}y\|^p + C\|\frac{1}{2}y\|^p \leq \frac{1}{2}\|x\|^p + \frac{1}{2}\|x + y\|^p$$

and subsequent steps may take the form

$$2\|x + 2^{-(n+1)}y\|^p + 2C\|2^{-(n+1)}y\|^p - \|x\|^p \leq \|x + 2^{-n}y\|^p \quad (3.4)$$

We then apply (3.4) repeatedly, substituting the last $\|x + 2^{-n}y\|^p$ term with the smaller quantity on the left. This yields the estimate

$$\begin{aligned}\|x + y\|^p &\geq -(1 + 2 + 2^2 + \dots + 2^{n-1})\|x\|^p \\ &\quad + \left[\frac{1}{2^{p-1}} + \frac{1}{(2^{p-1})^2} + \frac{1}{(2^{p-1})^3} + \dots + \frac{1}{(2^{p-1})^n} \right] C\|y\|^p \\ &\quad + 2^n \|x + 2^{-n}y\|^p \\ &= -(2^n - 1)\|x\|^p + \frac{1 - 2^{-(p-1)n}}{2^{p-1} - 1} C\|y\|^p + 2^n \|x + 2^{-n}y\|^p \\ &= \|x\|^p + \frac{1 - 2^{-(p-1)n}}{2^{p-1} - 1} C\|y\|^p + \frac{\|x + 2^{-n}y\|^p - \|x\|^p}{2^{-n}}\end{aligned}$$

Now take the limit as n tends toward infinity. Smoothness ensures that the directional derivative in the final term exists. To obtain its value, apply McShane's Lemma [7, Lemma 11.17, p. 120] to get

$$\lim_{a \rightarrow 0^+} \frac{\|x + ay\|^p - \|x\|^p}{a} = p\|x\|^{p-1}\mathfrak{R}(T_x(y))$$

The conclusion is

$$\|x + y\|^p \geq \|x\|^p + \frac{C}{2^{p-1} - 1} \|y\|^p + p\|x\|^{p-1}\mathfrak{R}(T_x(y))$$

This affirms (3.1) with $K = C/(2^{p-1} - 1)$, and an analogous argument establishes (3.2). \square

This, in turn, gives rise to a Pythagorean Theorem for weak parallelogram spaces. Let $1 < p < \infty$.

Theorem 3.3 *If a smooth Banach space \mathcal{X} is p -LWP(C), then there exists a positive constant K such that whenever $x \perp y$ in \mathcal{X} ,*

$$\|x\|^p + K\|y\|^p \leq \|x + y\|^p \tag{3.5}$$

If \mathcal{X} is p -UWP(C), then there exists a positive constant K such that whenever $x \perp y$ in \mathcal{X} ,

$$\|x\|^p + K\|y\|^p \geq \|x + y\|^p \tag{3.6}$$

In either case, the constant K can be chosen to be $C/(2^{p-1} - 1)$.

Proof The inequalities follow from the fact that $T_x(y) = 0$ if and only if $x \perp y$, as noted in Proposition 1.2. The proof of Lemma 3.2 assures that the constant $K = C/(2^{p-1} - 1)$ suffices. □

Let us refer to the constant K as a Pythagorean Constant for the space \mathcal{X} .

The L^p spaces certainly satisfy Pythagorean inequalities. These follow directly from the weak parallelogram laws for L^p .

Corollary 3.4 *Let $1 < r < \infty$, and $1 < p < \infty$.*

If $1 < p \leq 2 \leq r < \infty$ or $2 \leq p \leq r < \infty$, then there exists $K > 0$ such that

$$\|x\|_p^r + K\|y\|_p^r \leq \|x + y\|_p^r$$

whenever $x \perp y$ in L^p .

If $1 < r \leq p \leq 2$ or $1 < r \leq 2 \leq p < \infty$, then there exists $K > 0$ such that

$$\|x\|_p^r + K\|y\|_p^r \geq \|x + y\|_p^r$$

whenever $x \perp y$ in L^p .

Our first application to prediction theory is a bound on the coefficient growth for certain moving average processes. Let $\{X_n\}_{n=0}^\infty$ be a sequence of vectors in a Banach Space \mathcal{X} . Let us say that the sequence is an *innovation sequence* if each X_n is nonzero, and $X_m \perp X_n$ whenever $m < n$, where \perp is Birkhoff-James orthogonality. (This is a time reversal from the usual terminology.) In the special case $\{Z_n\}_{n=0}^\infty$ is an orthonormal sequence in a Hilbert space, then of course we have $\|\sum_{n=0}^\infty a_n Z_n\|^2 = \sum_{n=0}^\infty |a_n|^2$. This extends to Banach spaces in the following way.

Corollary 3.5 *Let $\{X_n\}_{n=0}^\infty$ be an innovation sequence of unit vectors in a smooth Banach Space \mathcal{X} , and suppose that the series $\sum_{n=0}^\infty a_n X_n$ converges in norm.*

If \mathcal{X} satisfies p -LWP, then

$$\left\| \sum_{n=0}^\infty a_n X_n \right\|^p \geq \sum_{n=0}^\infty K^n |a_n|^p \tag{3.7}$$

where K is the associated Pythagorean constant.

If \mathcal{X} satisfies p -UWP, then

$$\left\| \sum_{n=0}^\infty a_n X_n \right\|^p \leq \sum_{n=0}^\infty K^n |a_n|^p$$

where K is the associated Pythagorean constant.

This follows from applying the Pythagorean inequalities repeatedly.

The above norm estimates furnish a sense of the growth or decay of the coefficients a_n . For example, in the LWP case, (3.7) gives rise to the crude bound

$$|a_n| \leq K^{-n/p} \left\| \sum_{m=0}^{\infty} a_m X_m \right\|$$

where once again K is the associated Pythagorean constant.

Finally, let us return to the spaces L^p and use the Pythagorean inequalities to affirm that Theorem 2.1 indeed describes all of the r -weak parallelogram laws for L^p . That is, for r and p outside of its conditions, a weak parallelogram law fails to exist.

Proposition 3.6 *Let $1 < r < \infty$, and $1 < p < \infty$. If $r > p$ or $r > 2$, then there does not exist a positive constant C such that for all x and y in L^p ,*

$$\|x + y\|_p^r + C\|x - y\|_p^r \geq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right)$$

If $r < p$ or $r < 2$, then there does not exist a positive constant C such that for all x and y in L^p ,

$$\|x + y\|_p^r + C\|x - y\|_p^r \leq 2^{r-1} \left(\|x\|_p^r + \|y\|_p^r \right)$$

Proof Our strategy is to show that the respective Pythagorean inequalities fail, and hence the corresponding weak parallelogram laws fail. Consider the example of 2-dimensional ℓ^p with vectors $x = (a, a)$ and $y = (1, -1)$. Note that $x \perp_p y$. Assuming a is positive and large, we find that

$$\begin{aligned} & \frac{\|x + y\|_p^r - \|x\|_p^r}{\|y\|_p^r} \\ &= \frac{\left(|a + 1|^p + |a - 1|^p \right)^{r/p} - \left(|a|^p + |a|^p \right)^{r/p}}{(1 + 1)^{r/p}} \\ &= a^r \left[1 + \frac{p(p-1)}{2a^2} + O(1/a^3) \right]^{r/p} - a^r \\ &= \frac{1}{2} a^{r-2} r(p-1) + O(a^{r-3}) \end{aligned}$$

where the estimates come from the binomial series:

$$\left(1 + \frac{1}{a} \right)^p = 1 + p \left(\frac{1}{a} \right) + \frac{p(p-1)}{2} \left(\frac{1}{a} \right)^2 + O\left(\left[\frac{1}{a} \right]^3 \right)$$

As $a \rightarrow \infty$ this tends to the limit 0 if $1 < r < 2$; it diverges to ∞ if $2 < r < \infty$. It follows that ℓ^p fails to be r -LWP if $1 < r < 2$, and that ℓ^p fails to be r -UWP if $2 < r < \infty$.

Next, consider 2-dimensional ℓ^p with vectors $x = (1, 0)$ and $y = (0, a)$. As before, $x \perp_p y$. Assume that a is positive and small. Then

$$\begin{aligned} & \frac{\|x + y\|_p^r - \|x\|_p^r}{\|y\|_p^r} \\ &= \frac{\left(1 + a^p \right)^{r/p} - 1}{a^r} \\ &= \frac{r}{p} a^{p-r} + O(a^{2p-r}) \end{aligned}$$

As $a \rightarrow 0+$, this tends to 0 if $p > r$; it diverges to ∞ if $r > p$. We may conclude that ℓ^p fails to be r -LWP if $p > r$, and it fails to be r -UWP if $r > p$.

Now, every L^p (of dimension at least 2) contains a possibly weighted copy of 2-dimensional ℓ^p . It is a simple matter to account for the weights, and thus extend the above conclusions to L^p . Indeed, if the weights are w_1 and w_2 , then the mapping $(x_1, x_2) \mapsto (x_1/w_1^{1/p}, x_2/w_2^{1/p})$ is an isometry from ℓ^p to weighted ℓ^p that preserves orthogonality. \square

The results of this section amplify those of [4], where $r = 2$, as well as [10], where $p = r$, and apply to a larger class of Banach spaces.

4 Baxter’s inequality

Our next application is an extension of Baxter’s inequality from the stationary Gaussian case to a much broader class of processes. Suppose that $\{Z_k\}_{k=-\infty}^{\infty}$ is a centered stationary Gaussian process. Let us write \hat{Z} for the best linear least-squares estimate of Z_0 based on the entire past

$$\{\dots, Z_{-3}, Z_{-2}, Z_{-1}\}$$

and \hat{Z}_n for the best linear estimate of Z_0 based on the finite past

$$\{Z_{-n}, \dots, Z_{-3}, Z_{-2}, Z_{-1}\}$$

Let \hat{Z} and \hat{Z}_n have series representations

$$\begin{aligned} \hat{Z} &= \sum_{k=1}^{\infty} a_k Z_{-k} \\ \hat{Z}_n &= \sum_{k=1}^n a_{k,n} Z_{-k} \end{aligned}$$

The original Baxter’s inequality [2] states that if the process has a positive and continuous spectral density function, then there exist constants B and N such that

$$\sum_{k=1}^n |a_{k,n} - a_k| \leq B \sum_{k=n+1}^{\infty} |a_k|$$

whenever $n \geq N$. The importance of Baxter’s inequality is that it gives us an estimate of the error in using the truncated infinite best predictor in place of the finite best predictor, expressed in terms of the tail of the best predictor. This is useful in applications, and has been extended in a number of directions (see, for example, [12, 16]).

Extending this idea to arbitrary normed spaces is challenging since there doesn’t exist any notion of a spectrum in such generality. Here, however, is a version of Baxter’s inequality for LWP spaces, which evidently retain just enough Hilbert space character. Suppose that the sequence of vectors $\{X_n\}_{n=0}^{\infty}$ spans the Banach space \mathcal{X} . We adopt the notation \hat{X} for the best linear estimate of X_0 based on $\{X_1, X_2, X_3, \dots\}$, and \hat{X}_n for the best linear estimate of X_0 based on $\{X_1, X_2, X_3, \dots, X_n\}$. Let us write $\hat{X}_{(n)}$ for the n th partial sum of \hat{X} .

Theorem 4.1 *Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of unit vectors spanning a smooth Banach space \mathcal{X} . If \mathcal{X} satisfies p -LWP(C), and the series representation*

$$\hat{X} = \sum_{n=1}^{\infty} a_n X_n$$

converges in norm to \hat{X} , then there exist constants B and N such that

$$\|\hat{X}_n - \hat{X}_{(n)}\|^p \leq B \|\hat{X} - \hat{X}_{(n)}\| \tag{4.1}$$

whenever $n \geq N$.

Proof By assumption, $X_0 - \hat{X}_n$ is orthogonal to any vector in closed linear span of $\bigvee \{X_k\}_{k=1}^n$. In particular,

$$X_0 - \hat{X}_n \perp \hat{X}_n - \hat{X}_{(n)}$$

It follows that

$$\|X_0 - \hat{X}_n\|^p + C \|\hat{X}_n - \hat{X}_{(n)}\|^p \leq \|X_0 - \hat{X}_{(n)}\|^p$$

from the Pythagorean inequality (3.5), and therefore

$$\begin{aligned} C \|\hat{X}_n - \hat{X}_{(n)}\|^p &\leq \|X_0 - \hat{X}_{(n)}\|^p - \|X_0 - \hat{X}_n\|^p \\ &\leq \|X_0 - \hat{X}_{(n)}\|^p - \|X_0 - \hat{X}\|^p \end{aligned} \tag{4.2}$$

The Mean Value Theorem, applied to the function $f(x) = x^p$, assures that for any positive numbers $a > b$ there exists a number c , with $a > c > b$, such that

$$\frac{a^p - b^p}{a - b} = pc^{p-1}$$

Use $c^{p-1} \leq a^{p-1} + b^{p-1}$, $a = \|X_0 - \hat{X}_{(n)}\|$, $b = \|X_0 - \hat{X}\|$, and the Triangle inequality

$$\|X_0 - \hat{X}_{(n)}\| - \|X_0 - \hat{X}\| \leq \|\hat{X} - \hat{X}_{(n)}\|$$

and then the chain of estimates continues from (4.2) with

$$\begin{aligned} &\leq p \|\hat{X} - \hat{X}_{(n)}\| \left(\|X_0 - \hat{X}\|^{p-1} + \|X_0 - \hat{X}_{(n)}\|^{p-1} \right) \\ &\leq 2p \|\hat{X} - \hat{X}_{(n)}\| \end{aligned}$$

for n sufficiently large. □

Once again, what we have in (4.1) is an error estimate for using the truncated best predictor in place of the best finite predictor, in terms of the tail of the best predictor. This is a new result even for L^p , with $p \neq 2$. Thus it is applicable, for example, to a random process $\{X_n\}_{n=-\infty}^{\infty}$ that satisfies $E|X_n|^p < \infty$ for all n , and

$$E \left| \sum_{n=1}^N c_n X_{k_n+K} \right|^p = E \left| \sum_{n=1}^N c_n X_{k_n} \right|^p$$

for all integers K , positive integers n , scalars c_1, c_2, \dots, c_n , and indices k_1, k_2, \dots, k_n . Such a process is said to be p -stationary. For $1 < p \leq 2$, the p -stationary processes include weakly stationary processes, harmonizable stable processes, certain SoS stable moving averages, and p -th order strictly stationary processes (see [6,9–11,14]).

We emphasize that the Banach space norm $\|\cdot\|$ in Theorem 4.1 need not be the L^p norm; rather, the requirement is for the space to satisfy $p\text{-LWP}(C)$, which occurs for a large class of spaces.

5 Regularity

Let $\{X_n\}_{n=0}^\infty$ be a sequence of nonzero vectors spanning a Banach space \mathcal{X} . Define for all $n = 0, 1, 2, \dots$,

$$\begin{aligned} \mathcal{X}_n &= \bigvee \{X_n, X_{n+1}, X_{n+2}, \dots\} \\ \mathcal{X}_\infty &= \bigcap_n \mathcal{X}_n \end{aligned}$$

where \bigvee means the closed span under the norm topology. We say that the sequence $\{X_n\}_{n=0}^\infty$ is *regular* if $\mathcal{X}_\infty = (0)$. This is a linear version of the stronger condition that the sequence $\{X_n\}_{n=0}^\infty$ has trivial tail σ -field. It is also analogous to a Gaussian process being purely non-deterministic. In the stationary Gaussian case, complete spectral criteria are well known for regularity (see, for example, [17]). However, in the present more general setting, there is no comparable notion of spectrum, and even an innovation sequence can fail to be regular. Our aim is to characterize the regular innovation sequences in LWP spaces.

We can attempt to define the coordinate functionals k_n on finite linear combinations in the obvious way:

$$k_n \left(\sum a_k X_k \right) = a_n$$

It needs to be established that each k_n is well-defined, and extends continuously to all of \mathcal{X} . We find that a sufficient condition is for \mathcal{X} to satisfy LWP.

Lemma 5.1 *Let $\{X_n\}_{n=0}^\infty$ be an innovation sequence of unit vectors spanning a smooth Banach space \mathcal{X} . If \mathcal{X} satisfies p -LWP, then the coordinate functionals k_n , for $n = 0, 1, 2, \dots$, extend continuously to \mathcal{X} .*

Proof If LWP holds, then \mathcal{X} satisfies a lower Pythagorean inequality. Thus by Corollary 3.5 there is a constant C such that

$$\left\| \sum_{n=0}^\infty a_n X_n \right\|^p \geq \sum_{n=0}^\infty C^n |a_n|^p$$

holds when all but finitely many of the coefficients a_n are zero. From this it follows that if two such sums on the left are equal, their corresponding coefficients must coincide. It also follows that k_n is a bounded linear functional, with $\|k_n\| \leq C^{-n/p}$. □

Let us keep the symbol k_n for the extended functional.

These functionals connect to our objective by way of the following lemmas.

Lemma 5.2 *Let $\{X_n\}_{n=0}^\infty$ be an innovation sequence of unit vectors spanning a smooth Banach space \mathcal{X} , and assume that \mathcal{X} satisfies LWP. Then $X \in \mathcal{X}_\infty$ if and only if $k_n(X) = 0$ for all $n = 0, 1, 2, \dots$*

Proof Suppose that $X \in \mathcal{X}_\infty$. Then for each nonnegative integer m , there exists Y_m belonging to the span of $\{X_m, X_{m+1}, X_{m+2}, \dots\}$ such that $\|X - Y_m\| < 1/m$. Thus by the continuity of the functional k_n , we have $k_n(X) = \lim_{m \rightarrow \infty} k_n(Y_m) = 0$ for each n .

Conversely, suppose that $k_n(X) = 0$ for all $n = 0, 1, 2, \dots$. Consider finite sums $Y_m = \sum_{n=0}^\infty a_{m,n} X_n$ converging in norm to X . For any fixed positive integer N , and any $\epsilon > 0$, it happens that $\lim_{m \rightarrow \infty} a_{m,n} = 0$ for all $0 \leq n \leq N$. Thus we can choose m sufficiently large that Y_N differs from a vector in the span of $\{X_m, X_{m+1}, X_{m+2}, \dots\}$ by no more than ϵ in

norm. This shows that X belongs to \mathfrak{X}_N , and since N was arbitrary, it follows that X belongs to \mathfrak{X}_∞ . \square

Let us write T_n for the norming functional of X_n . Here is the way the norming functionals $\{T_n\}$ and coordinate functionals $\{k_n\}$ are related.

Lemma 5.3 *Let $\{X_n\}_{n=0}^\infty$ be an innovation sequence of unit vectors spanning a smooth Banach space \mathfrak{X} . If \mathfrak{X} satisfies LWP, then the spans of $\{k_n\}_{n=0}^\infty$ and $\{T_n\}_{n=0}^\infty$ in \mathfrak{X}^* coincide.*

Proof We verify that

$$\begin{aligned} k_0 &= T_0 \\ k_n &= T_n - \sum_{m=0}^{n-1} T_n(X_m)k_m \end{aligned}$$

for all $n = 1, 2, 3, \dots$. Indeed, the equations are true when the respective functionals are evaluated at any X_j . From this we deduce that each T_N belongs to the span of $\{k_n\}_{n=0}^N$, and similarly each k_N belongs to the span of $\{T_n\}_{n=0}^N$.

Thus $\{k_n\}_{n=0}^N$ and $\{T_n\}_{n=0}^N$ span the same subspace for all N . The assertion follows. \square

Lemma 5.4 *Let $\{X_k\}_{k=0}^\infty$ be an innovation sequence of unit vectors spanning a smooth p -LWP space \mathfrak{X} , and let T be a nonzero functional on \mathfrak{X} . The following are equivalent:*

- (i) $T \perp k_n$ for all $n = 0, 1, 2, \dots$
- (ii) $T/\|T\|$ is the norming functional for some nonzero $X \in \mathfrak{X}_\infty$.

Proof In any case, the p -LWP condition implies that \mathfrak{X} is uniformly convex, and hence reflexive. Furthermore, \mathfrak{X}^* must then be smooth.

If condition (i) holds, then by [13, Theorem 2.1], there exists a nonzero vector X in $\mathfrak{X} = \mathfrak{X}^{**}$ such that $k_n(X) = 0$ for all indexes n , and $T(X) = \|T\|\|X\|$. By Lemma 5.2, this implies that X belongs to \mathfrak{X}_∞ . It is readily seen that $T/\|T\|$ is the norming functional for X , and hence (ii) holds.

Conversely, assume (ii). Then for any constant a and any index n , we have

$$\begin{aligned} \|T + ak_n\| &\geq |T(X/\|X\|) + ak_n(X/\|X\|)| \\ &= |T(X/\|X\|)| \\ &= \|T\| \end{aligned}$$

and (i) follows. \square

Here is our characterization of regular innovation sequences in LWP spaces, in terms of the coordinate functionals, the norming functionals for the sequence, and a basis property. (See [17] for the definition of conditional basis).

Theorem 5.5 *Let $\{X_n\}_{n=0}^\infty$ be an innovation sequence of unit vectors spanning a smooth p -LWP space \mathfrak{X} . The following are equivalent.*

- (i) *The sequence $\{X_n\}_{n=0}^\infty$ is regular.*
- (ii) *The associated coordinate functionals $\{k_n\}_{n=0}^\infty$ span \mathfrak{X}^* .*
- (iii) *The sequence $\{X_n\}_{n=0}^\infty$ is a conditional basis for \mathfrak{X} .*
- (iv) *The associated norming functionals $\{T_n\}_{n=0}^\infty$ span \mathfrak{X}^* .*

Proof The equivalence of (i) and (ii) is immediate from Lemma 5.4.

Suppose that (i) holds, and for X in \mathcal{X} we have $k_n(X) = a_n$ for all n . If the series $\sum_{n=0}^{\infty} a_n X_n$ converges in norm to Y , then $k_n(X - Y) = 0$ for all n . By assumption this forces $X = Y$, and (iii) is affirmed.

Conversely, suppose (iii) holds. If X belongs to \mathcal{X}_{∞} , then $k_n(X) = 0$ for all n . By assumption, the series $\sum_{n=0}^{\infty} k_n(X)X_n$ converges to X . But the left hand side is the zero vector, and we may conclude that (i) holds.

The equivalence of (ii) and (iv) follows from Lemma 5.3. \square

These results show that weak parallelogram spaces serve as a natural environment in which to pursue prediction theory, extending previous work on Gaussian and p -stationary processes.

Acknowledgments The authors are grateful to the Referee for his or her helpful comments and suggestions.

References

1. J. Alonso, H. Martini, On Birkhoff orthogonality and isosceles orthogonality in normed linear spaces. *Aequat. Math.* **83**, 153–189 (2012)
2. G. Baxter, A norm inequality for a “finite section” Wiener–Hopf equation. *Illinois J. Math.* **7**, 97–103 (1963)
3. B. Beauzamy, *Introduction to Banach Spaces and their Geometry* (North-Holland, New York, 1982)
4. W.L. Bynum, Weak parallelogram laws for Banach spaces. *Can. Math. Bull.* **19**, 269–275 (1976)
5. W.L. Bynum, J.H. Drew, A weak parallelogram law for ℓ_p . *Am. Math. Mon.* **79**, 1012–1015 (1972)
6. S. Cambanis, C.D. Hardin Jr, A. Weron, Innovations and Wold decompositions of stable processes. *Probab. Theory Relat. Fields* **79**, 1–27 (1988)
7. N.L. Carothers, *A Short Course on Banach Space Theory* (Cambridge University Press, New York, 2005)
8. R. Cheng, C.B. Harris, Duality of the weak parallelogram laws on Banach spaces. *J. Math. Anal. Appl.* **404**, 64–70 (2013)
9. R. Cheng, Ch. Houdré, On the linear prediction of some L^p random fields. *J. Aust. Math. Soc. (Series A)* **67**, 31–50 (1999)
10. R. Cheng, A.G. Miamee, M. Pourahmadi, On the geometry of $L^p(\mu)$ with applications to infinite variance processes. *J. Aust. Math. Soc.* **74**, 35–42 (2003)
11. R. Cheng, A.G. Miamee, M. Pourahmadi, Regularity and minimality of infinite variance processes. *J. Theor. Probab.* **13**, 1115–1122 (2000)
12. R. Cheng, M. Pourahmadi, Baxter’s inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Relat. Fields* **95**, 115–124 (1993)
13. R.C. James, Orthogonality and linear functionals in normed linear spaces. *Trans. Am. Math. Soc.* **61**, 265–292 (1947)
14. A.G. Miamee, M. Pourahmadi, Wold decomposition, prediction and parametrization of stationary processes with infinite variance. *Probab. Theory Relat. Fields* **79**, 145–164 (1988)
15. B.J. Pettis, A proof that every uniformly convex space is reflexive. *Duke Math. J.* **5**, 249–253 (1939)
16. M. Pourahmadi, *Foundations of Time Series Analysis and Prediction Theory* (Wiley, New York, 2001)
17. Y.A. Rozanov, *Stationary Random Processes* (Holden-Day, San Francisco, 1967)