Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 162 (2010) 1717–1738

Journal of Approximation Theory

www.elsevier.com/locate/jat

Slow convergence of sequences of linear operators II: Arbitrarily slow convergence

Frank Deutsch[a,*], Hein Hundal[b]

[a] *Department of Mathematics, Penn State University, University Park, PA 16802, United States*
[b] *146 Cedar Ridge Drive, Port Matilda, PA 16870, United States*

Received 28 September 2009; received in revised form 9 April 2010; accepted 5 May 2010
Available online 27 May 2010

Communicated by Heinz Bauschke

Abstract

We study the *rate of convergence* of a sequence of linear operators that converges *pointwise* to a linear operator. Our main interest is in characterizing the slowest type of pointwise convergence possible. This is a continuation of the paper Deutsch and Hundal (2010) [14]. The main result is a "lethargy" theorem (Theorem 3.3) which gives useful conditions that guarantee arbitrarily slow convergence. In the particular case when the sequence of linear operators is generated by the powers of a single linear operator, we obtain a "dichotomy" theorem, which states the surprising result that either there is linear (fast) convergence or arbitrarily slow convergence; no other type of convergence is possible. The dichotomy theorem is applied to generalize and sharpen: (1) the von Neumann–Halperin cyclic projections theorem, (2) the rate of convergence for intermittently (i.e., "almost" randomly) ordered projections, and (3) a theorem of Xu and Zikatanov.

Keywords: Arbitrarily slow convergence; Higher powers of linear operators; Cyclic projections; Alternating projections; Randomly ordered projections; Intermittently ordered projections; Subspace corrections; Finite elements; Domain decomposition; Multigrid method; Rate of convergence

1. Introduction

There are some important algorithms in analysis that are all special cases of the following type. Let (L_n) be sequence of bounded linear operators from a Banach space X to a normed

* Corresponding author.
E-mail addresses: deutsch@math.psu.edu (F. Deutsch), hundalhh@yahoo.com (H. Hundal).

0021-9045/$ - see front matter
doi:10.1016/j.jat.2010.05.002

linear space Y, and suppose that the sequence converges *pointwise* to a bounded linear operator L, that is,

$$L(x) := \lim_{n\to\infty} L_n(x) \quad \text{for each } x \in X.$$

A natural and practical question that arises then is: What can be said about the *rate* or speed of this convergence? This is an interesting and important question that does not seem to have been studied in a general systematic way before. In the paper [14] we began such a theoretical study with the main emphasis on convergence that is "extremely slow", along with numerous applications. It is the object of this paper to continue this study by giving conditions that imply the *slowest* possible type of convergence, namely, "arbitrarily slow convergence", and to give more applications.

Recall that arbitrarily slow convergence was defined in [14]. (The sequence (L_n) is said to converge to L *arbitrarily slowly* provided it converges pointwise and, for each sequence of real numbers $(\phi(n))$ with $\lim_n \phi(n) = 0$, there exists $x = x_\phi \in X$ such that $\|L_n(x) - L(x)\| \geq \phi(n)$ for each n.) The phrase "arbitrarily slow convergence" has appeared in several papers. But in many of these, no precise definition was given. But even the precise definitions differed in a significant way. However, Schock [29] did give such a definition for a special class of methods for obtaining approximate solutions to a particular linear operator equation. In [14] we extended Schock's definition to a more general setting and showed that his definition was equivalent to what we called there "almost arbitrarily slow convergence" (see [14, Lemma 2.11]).

Section 3 contains the main result of the paper, a "lethargy" theorem (Theorem 3.3). It provides essential sufficient conditions guaranteeing that the sequence (L_n) converges to L arbitrarily slowly. Furthermore, Theorem 3.3 is the basis for all the main results and applications in Sections 4 and 6–8.

In Section 4 we consider the important special case when the sequence (L_n) is generated by the powers of a *single* linear operator T, i.e., $L_n = T^n$ for each n. The main result here is a "dichotomy" theorem (Theorem 4.4): If $\|T\| \leq 1$ and (T^n) converges pointwise to 0, then either $\|T^{n_1}\| < 1$ for some n_1 (in which case (T^n) converges to 0 linearly), or $\|T^n\| = 1$ for all n (in which case (T^n) converges to 0 arbitrarily slowly). This shows that, in the case of powers, there are exactly two different types of convergence possible: either linear (possibly finite) or arbitrarily slow. There are no intermediate types of pointwise convergence possible.

In Section 5 we compare our lethargy theorem with a classical "lethargy theorem" of Bernstein.

In Section 6 we apply Theorem 4.4 to sharpen and improve the von Neumann–Halperin theorem. Briefly, (Theorem 6.4) if $M_1, M_2, \ldots, M_r$ are closed subspaces in a Hilbert space X, then exactly one of two statements holds: either (1) $\sum_1^r M_i^\perp$ is closed (in which case $((P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n)$ converges to $P_{\cap_1^r M_i}$ linearly), or (2) $\sum_1^r M_i^\perp$ is not closed (in which case $((P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n)$ converges to $P_{\cap_1^r M_i}$ arbitrarily slowly). (Here P_M denotes the orthogonal projection onto M.) This generalizes a result established for the special case of two subspaces, where it was proved by a combination of the results of Bauschke et al. [5] and Bauschke et al. [7].

In Section 7 we obtain a generalization of Theorem 6.4 to the situation where the projections are *intermittently ordered*, which is "almost" randomly ordered (and not necessarily cyclically ordered).

In Section 8 we apply Theorem 4.4 to obtain a rate of convergence result that sharpens and improves one of the main results of Xu and Zikatanov [33].

We conclude this Introduction by recalling some common notation. If H is a Hilbert space and M is a closed (linear) subspace, we denote the *orthogonal projection* onto M by P_M. It is well-known that P_M is linear and has norm 1 (unless $M = \{0\}$), and $P_M(x)$ is the unique point in M closest to x:

$$\|x - P_M(x)\| = d(x, M) := \inf_{y \in M} \|x - y\|.$$

The *orthogonal complement* of M is the set

$$M^\perp := \{x \in H \mid \langle x, m\rangle = 0 \text{ for all } m \in M\}.$$

Further, if T is any bounded linear mapping from one normed linear space X into another Y, then the *kernel* or *null space* of T is the set

$$\ker T := \mathcal{N}(T) := \{x \in X \mid T(x) = 0\}.$$

All other undefined notation is standard and can be found, e.g., in [13].

2. Types of convergence

In this section, we assume that X and Y are normed linear spaces over the same scalar field and let $\mathcal{B}(X, Y)$ denote the normed linear space of all bounded linear operators L from X to Y with the usual norm (the operator norm)

$$\|L\| := \sup_{x \neq 0} \frac{\|L(x)\|}{\|x\|}.$$

Let the sequence (L_n) and L be in $\mathcal{B}(X, Y)$.

First it is convenient to recall various types of convergence.

Definition 2.1. The sequence (L_n) is said to converge to L in *norm* (resp., *pointwise*) provided that $\lim_n \|L_n - L\| = 0$ (resp., $\lim_n \|L_n(x) - L(x)\| = 0$ for each $x \in X$).

The following definition is somewhat more general than the usual notion of "linear convergence", but is a consequence of the usual (stronger) notion.

Definition 2.2. The sequence (L_n) is said to converge to L *linearly* if there exist constants $\alpha \in [0, 1)$ and $c \in \mathbb{R}$ such that $\|L_n - L\| \leq c\,\alpha^n$ for each n.

In "big O" notation, this can be rephrased as (L_n) converges to L linearly provided $\|L_n - L\| = O(\alpha^n)$ for some $\alpha \in [0, 1)$. (We should note that some authors call this "geometric" convergence.)

Let $\mathcal{O}$ denote the collection of all real-valued functions on the positive integers $\mathbb{N}$ that converge to 0. That is,

$$\mathcal{O} := \{\phi \mid \phi : \mathbb{N} \to \mathbb{R}, \lim_n \phi(n) = 0\}. \tag{2.1}$$

Next we define two types of very slow pointwise convergence. Indeed, arbitrarily slow convergence is the *slowest possible* type of pointwise convergence!

Definition 2.3. The sequence (L_n) converges to L *arbitrarily slowly* (resp., *almost arbitrarily slowly*) if the following two conditions are satisfied:

(1) $L_n(x) \to L(x)$ for each $x \in X$.

(2) For each $\phi \in \mathcal{O}$, there exists $x = x_\phi \in X$ such that

$$\|L_n(x) - L(x)\| \geq \phi(n) \quad \text{for each } n \in \mathbb{N} \text{ (resp., for infinitely many } n \in \mathbb{N}\text{)}.$$

These definitions were given in [14], where the definition of arbitrarily slow convergence was also shown to be equivalent to one that was first given by Bauschke et al. [5] (see also [7]).

Note that arbitrarily slow convergence of (L_n) to L is just pointwise convergence that can be made slowest possible. Clearly, if (L_n) converges arbitrarily slowly to L, then it must also converge almost arbitrarily slowly, but the converse is false in general [14, Example 3.4].

Remark 2.4. As noted in [14, Theorem 2.9], a definition *equivalent* to Definition 2.3 is obtained if the set $\mathcal{O}$ is replaced by the (more restrictive) set

$$\widetilde{\mathcal{O}} := \{\phi \mid \phi : \mathbb{N} \to (0, \infty), \phi(n+1) \leq \phi(n), \lim_n \phi(n) = 0\}. \tag{2.2}$$

That is, unlike $\mathcal{O}$, the functions in $\widetilde{\mathcal{O}}$ are also positive and decreasing.

Definition 2.5 (*[14]*)**.** Let $\phi \in \mathcal{O}$. The sequence (L_n) is said to converge to L *pointwise with order* ϕ provided that for each $x \in X$ there exists a constant $c_x > 0$ such that $\|L_n(x) - L(x)\| \leq c_x \phi(n)$ for each $n \in \mathbb{N}$.

The following result characterizes almost arbitrarily slow convergence.

Fact 2.6 (*[14, Theorem 3.1]*)**.** *Let X be a Banach space, Y a normed linear space, and let (L_n) and L be in $\mathcal{B}(X, Y)$. Then the following statements are equivalent:*

(1) *The sequence (L_n) converges to L almost arbitrarily slowly.*
(2) *(L_n) converges to L pointwise, but not pointwise with order ϕ for any $\phi \in \mathcal{O}$.*
(3) *(L_n) converges to L pointwise, but not in norm.*

Finally, it is important to note that almost arbitrarily slow convergence and arbitrarily slow convergence are *not possible* in a finite-dimensional space.

Fact 2.7 (*[14, Theorem 2.14]*)**.** *Suppose that X is finite-dimensional and L_n, L are linear operators from X to Y. Then (L_n) converges to L pointwise if and only if it converges in norm.*

In particular, arbitrarily slow convergence or almost arbitrarily slow convergence is never possible when X is finite-dimensional.

3. A lethargy theorem

The main result of this section is a *lethargy theorem* (Theorem 3.3) which states that, under hypotheses that are essential, (L_n) converges to L arbitrarily slowly. This will be the basis for all the main results in the later sections.

It is convenient to first record a simple observation.

Lemma 3.1. *There exist positive real numbers $\alpha, \beta, \gamma, \varepsilon$, and ν that satisfy the following conditions:*

(1) $\beta > 1 > \nu$,
(2) $\gamma < 1 - \beta^{-1}$, *and*
(3) $\frac{\alpha[\gamma(1-\nu)-\varepsilon-\beta^{-1}(1+\nu)]}{\beta(1-\beta^{-1})} > 1$.

Proof. Note that $\alpha = 50$, $\beta = 5$, $\gamma = 3/5$, $\varepsilon = 1/5$, $\nu = 1/50$ work. $\square$

The main technical step in the Lethargy Theorem 3.3 is the following result which characterizes almost arbitrarily slow convergence as something formally stronger. It is this strengthening that is needed to prove Theorem 3.3. The definition of almost arbitrarily slow convergence can be rephrased as follows: The sequence (L_n) converges to L almost arbitrarily slowly if and only if for each $\phi \in \widetilde{\mathcal{O}}$, there exists $x = x_\phi \in X$ and a subsequence (p_i) of (n) such that $\|L_{p_i}(x)-L(x)\| \geq \phi(p_i)$ for all i. The next theorem shows that one can obtain the following stronger "index shifted" conclusion for almost arbitrarily slow convergence: For each $\phi \in \widetilde{\mathcal{O}}$, there exists $x = x_\phi \in X$ and a subsequence (p_i) of (n) such that $\|L_{p_i}(x) - L(x)\| \geq \phi(p_{i-1})$ for all i. Since ϕ is a decreasing function, $\phi(p_{i-1}) \geq \phi(p_i)$ and so the latter property implies the former. However, we know of no "simple" proof of the former implying the latter other than the one given below.

Theorem 3.2 (*Another Characterization of Almost Arbitrarily Slow Convergence*)**.** *Let X be a Banach space, Y a normed linear space, and let $L, L_1, L_2, \ldots$ be linear operators in $\mathcal{B}(X, Y)$. Then the following statements are equivalent:*

(1) *(L_n) converges to L almost arbitrarily slowly.*
(2) *(L_n) converges to L pointwise and $\rho := \limsup_n \|L_n - L\| > 0$.*
(3) *(L_n) converges to L pointwise and for each $\phi \in \widetilde{\mathcal{O}}$, there exist $z = z_\phi \in X$ and a subsequence (p_i) of (n) such that*

$$\|L_{p_i}(z) - L(z)\| \geq \phi(p_i) \quad \textit{for each } i. \tag{3.1}$$

(4) *(L_n) converges to L pointwise and for each $\phi \in \widetilde{\mathcal{O}}$, there exist $z = z_\phi \in X$ and a subsequence (p_i) of (n) such that*

$$\|L_{p_i}(z) - L(z)\| > \phi(p_{i-1}) \quad \textit{for each } i. \tag{3.2}$$

Proof. First note that the equivalence of (1) and (3) is just a rewording of the definition using Remark 2.4, while the equivalence of statements (1) and (2) is from the equivalence of (1) and (3) in Fact 2.6. Also, if (4) holds, then since ϕ is monotonically decreasing, it follows from (3.2) that

$$\|L_{p_i}(z) - L(z)\| > \phi(p_{i-1}) \geq \phi(p_i) \quad \text{for each } i. \tag{3.3}$$

In particular, $\|L_n(z)-L(z)\| > \phi(n)$ for infinitely many n, and hence (L_n) converges to L almost arbitrarily slowly, i.e., (1) holds. Thus to complete the proof it suffices to verify that (2) implies (4). Thus assume that (2) holds. By the Uniform Boundedness Theorem, we have $\rho < \infty$.

Choose positive numbers $\alpha, \beta, \gamma, \varepsilon$, and ν to satisfy the conclusion of Lemma 3.1. In addition, fix any $\phi \in \widetilde{\mathcal{O}}$ and define $\psi := \frac{1}{\rho}\phi$. Note that $\psi \in \widetilde{\mathcal{O}}$ since $\phi \in \widetilde{\mathcal{O}}$.

Choose a subsequence (n_k) of (n) such that $\lim_k \|L_{n_k} - L\| = \rho$. By passing to a further subsequence if necessary, we may assume that

$$\rho(1 - \nu) < \|L_{n_k} - L\| < \rho(1 + \nu) \quad \text{for each } k. \tag{3.4}$$

Define $T_k := \frac{1}{\rho}(L_{n_k} - L)$ for each $k \in \mathbb{N}$. Then (T_k) converges to 0 pointwise, and

$$1 - \nu < \|T_k\| < 1 + \nu \quad \text{for each } k. \tag{3.5}$$

Claim. *There exist a sequence (x_n) in X and a strictly increasing sequence of positive integers (m_n) such that, for each $n \in \mathbb{N}$,*

(i) $\|x_n\| = \alpha/\beta^n$,
(ii) $\|T_{m_n}(x_k)\| < \frac{\alpha\varepsilon}{n\beta^n(1-\beta^{-1})}$ *for each $k < n$,*

(iii) $\|T_{m_n}(x_n)\| > \frac{\alpha\gamma}{\beta^n(1-\beta^{-1})}(1-\nu)$, *and*
(iv) $\psi(m_n) < 1/\beta^n$.

Suppose for the moment that the Claim is true. Set

$$z = \sum_1^\infty x_k.$$

Then z is a well-defined point in X since X is complete and

$$\sum_1^\infty \|x_k\| = \sum_1^\infty \frac{\alpha}{\beta^k} = \frac{\alpha\beta^{-1}}{1-\beta^{-1}} < \infty.$$

Fix any $i \in \mathbb{N}$. Then, using the Claim, we deduce that

$$\begin{aligned}
\|T_{m_i}(z)\| &= \left\|\sum_k T_{m_i}(x_k)\right\| = \left\|T_{m_i}(x_i) + \sum_{k\neq i} T_{m_i}(x_k)\right\| \\
&\geq \|T_{m_i}(x_i)\| - \left\|\sum_{k\neq i} T_{m_i}(x_k)\right\| \geq \|T_{m_i}(x_i)\| - \sum_{k\neq i}\|T_{m_i}(x_k)\| \\
&= \|T_{m_i}(x_i)\| - \sum_{k<i}\|T_{m_i}(x_k)\| - \sum_{k>i}\|T_{m_i}(x_k)\| \\
&> \frac{\alpha\gamma(1-\nu)}{\beta^i(1-\beta^{-1})} - \sum_{k<i}\frac{\alpha\varepsilon}{i\beta^i(1-\beta^{-1})} - \sum_{k>i}(1+\nu)\|x_k\| \\
&= \frac{\alpha\gamma(1-\nu)}{\beta^i(1-\beta^{-1})} - \frac{(i-1)\alpha\varepsilon}{i\beta^i(1-\beta^{-1})} - \sum_{k>i}\frac{\alpha}{\beta^k}(1+\nu) \\
&> \frac{\alpha\gamma(1-\nu)}{\beta^i(1-\beta^{-1})} - \frac{\alpha\varepsilon}{\beta^i(1-\beta^{-1})} - \frac{\alpha(1+\nu)}{\beta^{i+1}(1-\beta^{-1})} \\
&= \frac{\alpha}{\beta^i(1-\beta^{-1})}\left[\gamma(1-\nu) - \varepsilon - (1+\nu)\beta^{-1}\right] \\
&> \frac{1}{\beta^{i-1}} \quad \text{(by Lemma 3.1(3))}, \\
&> \psi(m_{i-1}) \quad \text{(by (iv) of the Claim)}.
\end{aligned}$$

It follows from this inequality that

$$\|L_{n_{m_i}}(z) - L(z)\| = \rho\|T_{m_i}(z)\| > \rho\psi(m_{i-1}) = \phi(m_{i-1}) \geq \phi(n_{m_{i-1}}) \tag{3.6}$$

for each i. Let $p_i = n_{m_i}$ for each i. Then (p_i) is a subsequence of (n) and, by (3.6),

$$\|L_{p_i}(z) - L(z)\| > \phi(p_{i-1}) \quad \text{for each } i. \tag{3.7}$$

Thus (4) holds. This shows that if the Claim holds, then (4) holds.

It remains to prove the Claim. We construct the sequences (x_n) and (m_n) inductively and simultaneously. For $n = 1$, choose an integer m_1 so large that $\psi(m_1) < 1/\beta$. Since $\|T_{m_1}\| > 1-\nu$, we see that

$$\sup_{\|x\|=\alpha/\beta} \|T_{m_1}(x)\| > (1-\nu)\alpha/\beta. \tag{3.8}$$

Using Lemma 3.1(2), it follows that

$$\frac{\alpha\gamma(1-\nu)}{\beta(1-\beta^{-1})} < \frac{\alpha(1-\nu)}{\beta}. \tag{3.9}$$

Clearly, using (3.8) and (3.9), we can choose $x_1 \in X$ such that

$$\|x_1\| = \frac{\alpha}{\beta} \quad \text{and} \quad \|T_{m_1}(x_1)\| > \frac{\alpha\gamma(1-\nu)}{\beta(1-\beta^{-1})}.$$

Thus x_1 and m_1 satisfy (i), (iii), and (iv) of the Claim when $n = 1$. Since (ii) of the Claim is vacuously satisfied when $n = 1$, x_1 and m_1 satisfy (i)–(iv) of the Claim.

Next suppose that $x_1, x_2, \dots, x_n$ in X and $m_1 < m_2 < \cdots < m_n$ in $\mathbb{N}$ have been chosen to satisfy properties (i)–(iv) of the Claim. Thus, for each $i \le n$,

$$\|x_i\| = \frac{\alpha}{\beta^i}, \tag{3.10}$$

$$\|T_{m_i}(x_k)\| < \frac{\alpha\varepsilon}{i\beta^i(1-\beta^{-1})} \quad \text{for each } k < i, \tag{3.11}$$

$$\|T_{m_i}(x_i)\| > \frac{\alpha\gamma(1-\nu)}{\beta^i(1-\beta^{-1})}, \quad \text{and} \tag{3.12}$$

$$\psi(m_i) < \frac{1}{\beta^i}. \tag{3.13}$$

We construct $x_{n+1} \in X$ and an integer $m_{n+1} > m_n$ as follows. Since (T_k) converges to 0 pointwise, we see that

$$\lim_{k\to\infty} \left(\max_{1\le j\le n} \|T_k(x_j)\| \right) = 0.$$

Hence there exists an integer $\widetilde{m_n} \ge m_n$ such that

$$\max_{1\le j\le n} \|T_k(x_j)\| < \frac{\alpha\varepsilon}{(n+1)\beta^{n+1}(1-\beta^{-1})} \quad \text{for all } k \ge \widetilde{m_n}. \tag{3.14}$$

Choose an integer $m_{n+1} > \widetilde{m_n}$ such that $\psi(m_{n+1}) < 1/\beta^{n+1}$. Then (3.13) holds when i is replaced by $n+1$. Let $k < n+1$. Then (3.14) implies that

$$\|T_{m_{n+1}}(x_k)\| < \frac{\alpha\varepsilon}{(n+1)\beta^{n+1}(1-\beta^{-1})}.$$

Thus (3.11) holds for $i = n+1$ and $k < n+1$.

Since $\|T_{m_{n+1}}\| > 1-\nu$, Lemma 3.1(2) implies that

$$\sup_{\|x\|=\alpha/\beta^{n+1}} \|T_{m_{n+1}}(x)\| > \frac{\alpha(1-\nu)}{\beta^{n+1}} > \frac{\alpha\gamma(1-\nu)}{\beta^{n+1}(1-\beta^{-1})}.$$

Hence there exists $x_{n+1} \in X$ such that $\|x_{n+1}\| = \alpha/\beta^{n+1}$ and

$$\|T_{m_{n+1}}(x_{n+1})\| > \frac{\alpha\gamma(1-\nu)}{\beta^{n+1}(1-\beta^{-1})}.$$

The last two paragraphs show that (3.10)–(3.13) hold when $i = n+1$. This construction proves the Claim by induction, and hence that (2) $\Rightarrow$ (4). $\square$

The main lethargy theorem of this paper can now be stated. It gives a useful sufficient condition that insures arbitrarily slow convergence. Of course, since arbitrarily slow convergence implies almost arbitrarily slow convergence, what is the additional condition that is necessary to add to almost arbitrarily slow convergence to guarantee arbitrarily slow convergence?

Theorem 3.3 (*Lethargy Theorem*)**.** *Let X be a Banach space, Y a normed linear space, and let $L, L_1, L_2, \ldots$ be bounded linear operators in $\mathcal{B}(X, Y)$. Suppose that (L_n) converges to L almost arbitrarily slowly and satisfies the following monotonicity condition:*

$$\|L_{n+1}(x) - L(x)\| \leq \|L_n(x) - L(x)\| \quad \textit{for each } n \in \mathbb{N} \textit{ and } x \in X. \tag{3.15}$$

Then (L_n) converges to L arbitrarily slowly.

Proof. Let $\phi \in \widetilde{\mathcal{O}}$. By Theorem 3.2, we have that there exist $z = z_\phi \in X$ and a subsequence (m_i) of (n) such that

$$\|L_{m_i}(z) - L(z)\| > \phi(m_{i-1}) \quad \text{for each } i. \tag{3.16}$$

Fix any $k \in \mathbb{N}$. If $m_i \leq k \leq m_{i+1}$ for some i, then, using (3.15) and the fact that ϕ is decreasing, we obtain

$$\frac{\|L_k(z) - L(z)\|}{\phi(k)} \geq \frac{\|L_{m_{i+1}}(z) - L(z)\|}{\phi(m_i)} > 1.$$

This proves that

$$\|L_k(z) - L(z)\| > \phi(k) \quad \text{for each } k \geq m_1. \tag{3.17}$$

Define

$$\rho := \max_{1 \leq k \leq m_1} \left\{ \frac{\phi(k)}{\|L_k(z) - L(z)\|}, 1 \right\} \quad \text{and} \quad x := \rho z.$$

Then, for all $k \geq m_1$, we obtain from (3.17) that

$$\|L_k(x) - L(x)\| = \rho \|L_k(z) - L(z)\| \geq \|L_k(z) - L(z)\| > \phi(k). \tag{3.18}$$

If $1 \leq k \leq m_1$, the definition of ρ implies that

$$\begin{aligned} \|L_k(x) - L(x)\| &= \rho \|L_k(z) - L(z)\| \\ &\geq \frac{\phi(k)}{\|L_k(z) - L(z)\|} \|L_k(z) - L(z)\| = \phi(k). \end{aligned} \tag{3.19}$$

Hence (L_n) converges to L arbitrarily slowly. □

Remark 3.4. The Lethargy Theorem is *best possible* in the sense that *none* of the hypotheses is superfluous. More precisely, the theorem is false in general if either of the following hypotheses is omitted: the almost arbitrarily slow convergence of (L_n) to L and the monotonicity condition (3.15).

To see that the "almost arbitrarily slow convergence" hypothesis cannot be dropped, see [14, Example 2.5]. To see that the monotonicity condition (3.15) cannot be dropped, see [14, Example 3.4].

Remark 3.5. It is perhaps worth observing that the monotonicity condition (3.15) is related to the *Fejér monotonicity condition* which has been shown to be useful in convexity and optimization (see, e.g., [4,9]). (Recall that if C is a closed convex set in X, then a sequence

(x_n) in X is said to be *Fejér monotone with respect to* C if $\|x_{n+1} - c\| \le \|x_n - c\|$ for each $c \in C$.) Indeed, using this terminology, the condition (3.15) may be restated as: For each $x \in X$, the sequence $(L_n(x))$ is Fejér monotone with respect to $L(x)$.

4. Trichotomy for powers of an operator

The simplest application of the Lethargy Theorem 3.3, and the most useful for some of our later applications, occurs when the sequence (L_n) is generated by the *powers of a single operator.* This will follow as a consequence of the Lethargy Theorem and the next lemma.

Lemma 4.1. *Let T be a linear operator from a Banach space X into itself with $\|T\| \le 1$. Then*
(1) $\|T^{n+1}(x)\| \le \|T^n(x)\|$ *for each $x \in X$ and $n \in \mathbb{N}$.*
(2) $\|T^{n+1}\| \le \|T^n\| \le 1$ *for each $n \in \mathbb{N}$.*
(3) $\|T^{mn}\| \le \|T^m\|^n$ *for each $m, n \in \mathbb{N}$.*

Proof. (1) and (2) are well-known and easy to verify using the simple fact that $\|T(z)\| \le \|z\|$ for any $z \in X$. Since $T^{mn} = (T^m)^n$, we see that $\|T^{mn}\| = \|(T^m)^n\| \le \|T^m\|^n$, which proves (3). □

The main result of this section is the following *trichotomy* theorem for powers of a linear operator.

Theorem 4.2 (*Trichotomy*)**.** *Let X be a Banach space and $T : X \to X$ be a linear operator with $\|T\| \le 1$. Then exactly one of the following three statements holds:*
(1) $\|T^{n_1}\| < 1$ *for some n_1; in this case, (T^n) converges to 0 linearly.*
(2) $\|T^n\| = 1$ *for all n and $T^n(x) \to 0$ for each $x \in X$; in this case, (T^n) converges to 0 arbitrarily slowly.*
(3) $\|T^n\| = 1$ *for all n and $T^n(x) \not\to 0$ for some $x \in X$.*

Proof. Clearly, there are three mutually exclusive possibilities: (a) $\|T^{n_1}\| < 1$ for some n_1, or (b) $\|T^n\| = 1$ for all n and $T^n(x) \to 0$ for all $x \in X$, or (c) $\|T^n\| = 1$ for all n and $T^n(x) \not\to 0$ for some x. Thus it remains to verify that if (a) (resp., (b)) holds, then (T^n) converges to 0 linearly (resp., arbitrarily slowly).

Suppose that (a) holds: $\rho_1 := \|T^{n_1}\| < 1$ for some n_1. If $\rho_1 > 0$, set $\rho := \rho_1^{1/n_1}$ so that $0 < \rho < 1$. Let $n \in \mathbb{N}$. Then $n = kn_1 + i$ for some integer $k \ge 0$ and $i \in \{0, 1, \dots, n_1 - 1\}$. Then, using Lemma 4.1, we deduce

$$\begin{aligned}\|T^n\| &= \|T^{kn_1+i}\| \le \|T^{kn_1}\| \le \|T^{n_1}\|^k = \rho_1^k = \rho^{n_1 k}\\ &= \frac{\rho^{n_1 k+i}}{\rho^i} = \frac{\rho^n}{\rho^i} \le \frac{\rho^n}{\rho^{n_1-1}} = \mu\rho^n,\end{aligned}$$

where $\mu := 1/\rho^{n_1-1}$.

On the other hand, if $\rho_1 = 0$, it is clear that $\|T^n\| \le 2^{n_1}(1/2)^n$ for each n. In either case, we obtain that there exist constants $\beta \in \mathbb{R}$ and $\alpha \in [0, 1)$ such that

$$\|T^n\| \le \beta\alpha^n \quad \text{for each } n \in \mathbb{N}.$$

Thus (T^n) converges to 0 linearly.

Next suppose that (b) holds: $\|T^n\| = 1$ for all n and $T^n(x) \to 0$ for all $x \in X$. Using Lemma 4.1 and Theorems 3.2 and 3.3 with $L_n = T^n$, it follows that (T^n) converges to 0 arbitrarily slowly. □

In contrast to the case for [14, Example 3.4], it turns out that for powers of an operator, almost arbitrarily slow convergence and arbitrarily slow convergence are the same.

Corollary 4.3. *Let X be complete and $T \in \mathcal{B}(X, X)$ with $\|T\| \leq 1$. Then (T^n) converges to 0 arbitrarily slowly if and only if (T^n) converges to 0 almost arbitrarily slowly.*

Proof. Clearly, arbitrarily slow convergence implies almost arbitrarily slow convergence. Conversely, if (T^n) converges to 0 almost arbitrarily slowly, then the Trichotomy Theorem implies that (T^n) converges either linearly to 0 or arbitrarily slowly. But if (T^n) converges linearly to 0, then it converges to 0 in norm. By Fact 2.6, it follows that (T^n) converges to 0 pointwise, but not almost arbitrarily slowly. This contradiction shows that (T^n) must converge to 0 arbitrarily slowly. □

In all the applications of the Trichotomy Theorem that we make, the condition $T^n(x) \to 0$ for all $x \in X$ is known (or can be shown) to hold. For ease of reference, we state the trichotomy in this particular case, and we obtain the following *dichotomy theorem*.

Theorem 4.4 (*Dichotomy*)**.** *Let X be a Banach space and $T : X \to X$ be a linear operator with $\|T\| \leq 1$ and $T^n(x) \to 0$ for each $x \in X$. Then exactly one of the following two statements holds:*

(1) *There exists $n_1 \in \mathbb{N}$ such that $\|T^{n_1}\| < 1$, and (T^n) converges to 0 linearly.*
(2) *$\|T^n\| = 1$ for each $n \in \mathbb{N}$, and (T^n) converges to 0 arbitrarily slowly.*

Example 4.5. Let $L : \ell_2 \to \ell_2$ denote the *left-shift* operator. That is, for each $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i \in \ell_2$,

$$L(x) = \sum_{i=2}^{\infty} \langle x, e_i \rangle e_{i-1},$$

where $\{e_i \mid i = 1, 2, \ldots\}$ is the canonical orthonormal basis in ℓ_2: $e_i(j) = \delta_{ij}$, Kronecker's delta. Then (L^n) converges to 0 arbitrarily slowly.

To see this, first note that for each $x \in \ell_2$ it is easy to check that

$$L^n(x) = \sum_{i=n+1}^{\infty} \langle x, e_i \rangle e_{i-n} \quad \text{for each } n \in \mathbb{N}.$$

It follows from this that $\|L^n\| = 1$ for each $n \in \mathbb{N}$ and, for each $x \in \ell_2$,

$$\|L^n(x)\| = \|x\|^2 - \sum_{i=1}^{n} |x(i)|^2 \to 0.$$

The result now follows from the Dichotomy Theorem 4.4.

Corollary 4.6. *Let X be a Banach space and $T : X \to X$ a linear operator with $\|T\| \leq 1$. Then (T^n) converges to 0 linearly if and only if $\|T^{n_1}\| < 1$ for some $n_1 \in \mathbb{N}$.*

Proof. This follows immediately from the Trichotomy Theorem 4.2. (Alternatively, a direct proof may be given from first principles.) □

Corollary 4.7. *Let X be a Banach space and $T : X \to X$ a linear operator with $\|T\| \leq 1$. Then the following statements are equivalent:*

(1) *(T^n) converges to 0 arbitrarily slowly;*

(2) (T^n) *converges to* 0 *almost arbitrarily slowly;*
(3) (T^n) *converges to* 0 *pointwise and* $\|T^n\| = 1$ *for each* n*;*
(4) (T^n) *converges to* 0 *pointwise, but* (T^n) *does not converge to* 0 *pointwise with order* ϕ *for any* $\phi \in \mathcal{O}$.

Proof. The equivalence of statements (1) and (2) follows from Corollary 4.3, while that of (1) and (3) follows from the Dichotomy Theorem 4.4. By Fact 2.6, (4) is equivalent to (2). □

As we noted in [14, Theorem 2.14], when X is finite-dimensional, arbitrarily slow convergence or almost arbitrarily slow convergence is *never possible*. Thus we have the following easy consequence of the Dichotomy Theorem 4.4 and Corollary 4.6.

Corollary 4.8. *Let* X *be finite-dimensional and let* $T \in \mathcal{B}(X, X)$ *satisfy* $\|T\| \le 1$*. Then the following statements are equivalent:*

(1) (T^n) *converges to* 0 *pointwise.*
(2) *There exists an integer* n_1 *such that* $\|T^{n_1}\| < 1$.
(3) (T^n) *converges to* 0 *linearly.*

5. The Bernstein Lethargy Theorem

In this section we compare the Lethargy Theorem 3.3 with the classical lethargy theorem of Bernstein. Let $\{x_1, x_2, x_3, \ldots\}$ be a set of linearly independent elements in a normed linear space X with the property that each $x \in X$ can be approximated arbitrarily well by elements in the linear space spanned by the x_n's. That is, for each $x \in X$ and each $\varepsilon > 0$, there exist scalars α_i for $i = 1, 2, \ldots, n$ such that $y = \sum_1^n \alpha_i x_i$ satisfies $\|x - y\| < \varepsilon$. Note that such a space must be *separable*, that is, it must contain a countable dense set (viz., all linear combinations with rational coefficients). For such a sequence, define the linear subspaces M_n by

$$M_n := \operatorname{span}\{x_1, x_2, \ldots, x_n\} \quad \text{for } n = 1, 2, \ldots.$$

In particular, for each n, $\dim M_n = n$, $M_n \subset M_{n+1}$, and $X = \overline{\cup_1^\infty M_n}$.

(It is perhaps of some interest to note that a converse of this is also valid. More precisely, if X is an infinite-dimensional separable Banach space, then there exists a sequence of subspaces (M_n) such that $M_n \subset M_{n+1}$, $\dim M_n = n$, and $X = \overline{\cup_1^\infty M_n}$. To see this, let $\{x_1, x_2, \ldots\}$ be a countable dense set in X and, for each $n \in \mathbb{N}$, set $m_n := \min\{m \mid \dim(\operatorname{span}\{x_1, x_2, \ldots, x_m\}) = n\}$ and $M_n := \operatorname{span}\{x_1, x_2, \ldots, x_{m_n}\}$. Then $\dim M_n = n$, $M_n \subset M_{n+1}$, and $X = \overline{\cup_1^\infty M_n}$.)

Denote the distance from any $x \in X$ to M_n by

$$d(x, M_n) := \inf_{y \in M_n} \|x - y\|.$$

Then we can state the following lethargy theorem.

Theorem 5.1 (*Bernstein Lethargy Theorem*)**.** *Let* X *be a Banach space and assume* (M_n) *is an increasing sequence of subspaces with* $\dim M_n = n$ *and* $X = \overline{\cup_1^\infty M_n}$*. For each* $\phi \in \widetilde{\mathcal{O}}$*, there exists* $x = x_\phi \in X$ *such that*

$$d(x, M_n) = \phi(n) \text{ for each } n \in \mathbb{N}.$$

Bernstein [8] actually proved this result for the case when $X = C[a, b]$ in 1938, but Timan [31, pp. 41–43] observed that it holds in the more general case as stated above. (See also Davis [10, p. 322].)

How does the Bernstein Lethargy Theorem compare with the Lethargy Theorem 3.3? In general, a direct comparison is not possible since the latter is phrased in terms of linear operators, while the former is phrased in terms of distances to finite-dimensional subspaces. There is one case, however, where a reasonable comparison is possible. This is when X is a (separable) Hilbert space since then $d(x, M_n) = \|x - P_{M_n}(x)\|$. In this case, Bernstein's Lethargy Theorem can be stated in the following form.

Theorem 5.2 (*Bernstein Lethargy Theorem: Hilbert Space Case*)**.** *Let H be a Hilbert space and let (M_n) be an increasing sequence of subspaces such that* $\dim M_n = n$ *and $H = \overline{\cup_1^\infty M_n}$. Then for each $\phi \in \widetilde{\mathcal{O}}$ there exists $x = x_\phi \in H$ such that*

$$\|x - P_{M_n}(x)\| = \phi(n) \quad \textit{for all } n \in \mathbb{N}. \tag{5.1}$$

It should be mentioned that in Theorem 5.2, it is not hard to also show that (P_{M_n}) converges pointwise to the identity operator I and, in particular, that (P_{M_n}) converges to I arbitrarily slowly. Before comparing this with our results, let us first note that there is an even stronger version of Theorem 5.2 that we can establish.

Theorem 5.3. *Let H be a Hilbert space and let (M_n) be a sequence of closed (not necessarily finite-dimensional) subspaces in H having the property that $\{0\} \neq M_n \subset M_{n+1}$, $M_n \neq M_{n+1}$, and let $M := \overline{\cup_1^\infty M_n}$. Then (P_{M_n}) converges pointwise to P_M, and for each $\phi \in \widetilde{\mathcal{O}}$ there exists $x = x_\phi \in H$ such that*

$$\|P_{M_n}(x) - P_M(x)\| = \phi(n) \quad \textit{for each } n \in \mathbb{N}. \tag{5.2}$$

In particular, (P_{M_n}) converges arbitrarily slowly to P_M.

Proof. Fix any $x \in H$. Since $M_n \subset M_{n+1} \subset M$ for each n, the projections P_{M_n} and P_M must commute (see, e.g., [13, Lemma 9.2, p. 194]) and so

$$P_{M_n}x = P_M P_{M_n}x = P_{M_n} P_M x \quad \text{for each } n. \tag{5.3}$$

Since $P_M x \in M$ and $\cup_1^\infty M_n$ is dense in M, for each $\varepsilon > 0$ there exists $y \in \cup_1^\infty M_n$ such that

$$\|P_M x - y\| < \varepsilon. \tag{5.4}$$

Now $y \in M_n$ for some n, and hence for all n sufficiently large. Use (5.3) and (5.4) to obtain that, for all n sufficiently large,

$$\|P_{M_n}x - P_M x\| = \|P_{M_n} P_M x - P_M x\| = d(P_M x, M_n) \leq \|P_M x - y\| < \varepsilon. \tag{5.5}$$

This proves that $(P_{M_n}x)$ converges to $P_M x$. Since x was arbitrary, (P_{M_n}) converges to P_M pointwise.

Since $M_n \subset M_{n+1}$ and $M_n \neq M_{n+1}$ for each n, we may choose a sequence (e_n) with

$$e_1 \in M_1, \quad \|e_1\| = 1, \quad \text{and} \quad e_n \in M_n \cap M_{n-1}^\perp \quad \text{for all } n \geq 2. \tag{5.6}$$

To see this, it suffices to note that $M_n \cap M_{n-1}^\perp \neq \{0\}$ for all $n \geq 2$. And to verify this, we choose any $y \in M_n \setminus M_{n-1}$ and let $z := y - P_{M_{n-1}}y$. Then $z \in M_n \cap M_{n-1}^\perp \setminus \{0\}$. Note that by the way it was chosen, the sequence (e_n) is orthonormal.

Next we define subspaces E_n and E by

$$E_n := \operatorname{span}\{e_1, e_2, \ldots, e_n\} \quad \text{for all } n \in \mathbb{N} \text{ and } E = \overline{\cup_1^\infty E_n}. \tag{5.7}$$

Clearly, $E_n \subset E_{n+1} \subset E$, $\dim E_n = n$, and $E_n \subset M_n$ for all $n \in \mathbb{N}$. By Bernstein's Theorem 5.2 (applied in the Hilbert space E), for each $\phi \in \widetilde{\mathcal{O}}$ there exists $x = x_\phi \in E$ such that

$$\|P_{E_n}(x) - x\| = \phi(n) \quad \text{for each } n \in \mathbb{N}. \tag{5.8}$$

To verify (5.2) and hence complete the proof, it suffices by (5.8) to show that $P_M(x) = x$ and $P_{M_n}(x) = P_{E_n}(x)$. But as noted above, $x \in E \subset M$ and so $P_M(x) = x$. Next observe that $P_{E_n}(x) \in E_n \subset M_n$ and by the way the e_n were chosen in (5.6), we see that

$$x - P_{E_n}(x) = \sum_{n+1}^{\infty} \alpha_k e_k \in M_n^\perp$$

since $M_{n+i}^\perp \subset M_n^\perp$ and $e_{n+i} \in M_{n+i-1}^\perp$ for all $i \in \mathbb{N}$. By the well-known characterization of best approximations (see, e.g., [13, Theorem 4.9]), it follows that $P_{E_n}(x) = P_{M_n}(x)$ and the proof is complete. $\square$

Remark 5.4. (1) Comparing Theorems 5.3 and 5.2, we see that in Theorem 5.3, the closed subspaces are not necessarily increasing by one dimension at each step as in Theorem 5.2, and they can even be infinite-dimensional. Secondly, the closure of the union of the subspaces in Theorem 5.3 does not have to be the whole space as in Theorem 5.2.

(2) It is worth noting that Theorem 5.3 is no longer valid if the hypothesis that $M_n \neq M_{n+1}$ for all n is dropped. For if $M_n = M_{n+1}$ for some n, say $n = n_1$, then we see that *every* $x \in X$ must satisfy

$$\|P_{M_{n_1}}(x) - P_M(x)\| = \|P_{M_{n_1+1}}(x) - P_M(x)\|. \tag{5.9}$$

Hence if $\phi \in \widetilde{\mathcal{O}}$ is chosen so that $\phi(n_1) > \phi(n_1 + 1)$, then because of (5.9), we see that (5.2) is impossible for both n_1 and $n_1 + 1$ no matter which x is chosen.

However, if we are only interested in concluding arbitrarily slow convergence, then we can further weaken the hypothesis of Theorem 5.3.

Theorem 5.5. *Let H be a Hilbert space and let (M_n) be any nondecreasing sequence of closed subspaces (not necessarily finite-dimensional) such that the closed subspace $M := \overline{\cup_1^\infty M_n}$ is infinite-dimensional and $M \neq M_n$ for every n. Then (P_{M_n}) converges to P_M arbitrarily slowly.*

Proof. By the same proof as in Theorem 5.3, we can show that (P_{M_n}) converges pointwise to P_M.

Next we show that

$$M_n^\perp \cap M \neq \{0\} \quad \text{for each } n. \tag{5.10}$$

Fix any $n \in \mathbb{N}$. Since $M_n \neq M$ by hypothesis, choose $x \in M \setminus M_n$. Then $z := x - P_{M_n}x \in M \cap M_n^\perp \setminus \{0\}$, which verifies (5.10).

From (5.10) and the fact that $\ker P_{M_n} = M_n^\perp$ (see, e.g., [13, p. 76]), it follows that

$$\ker P_{M_n} \cap \{x \in H \mid \|P_M x\| \geq \|x\|\} = M_n^\perp \cap M \neq \{0\} \text{ for each } n. \tag{5.11}$$

By [14, Lemma 3.3] (P_{M_n}) converges to P_M almost arbitrarily slowly.

Finally, we claim that

$$\|P_{M_{n+1}}x - P_M x\| \leq \|P_{M_n}x - P_M x\| \quad \text{for each } x \in X, n \in \mathbb{N}. \tag{5.12}$$

To see this, use (5.3) and the fact that $M_n \subset M_{n+1}$ to deduce that

$$\begin{aligned}\|P_{M_{n+1}}x - P_M x\| &= \|P_{M_{n+1}}P_M x - P_M x\| \\ &= d(P_M x, M_{n+1}) \leq d(P_M x, M_n) \\ &= \|P_{M_n}P_M x - P_M x\| = \|P_{M_n}x - P_M x\|,\end{aligned}$$

which proves (5.12).

The result now follows by an application of Theorem 3.3. □

Remark 5.6. (1) Note that the main difference in the hypotheses of Theorems 5.3 and 5.5 is that in the latter, it is *not* assumed that $M_n \neq M_{n+1}$ for each n. However, it was observed in Remark 5.4(2) that this hypothesis was essential for obtaining Eq. (5.2) in Theorem 5.3.

(2) Theorem 5.5 is best possible in the sense that if either of the two hypotheses (M is infinite-dimensional, and $M \neq M_n$ for all n) is dropped, then the conclusion fails. For if M were finite-dimensional, then by Fact 2.7 (taking $X = M$), the sequence of projections could not converge arbitrarily slowly, while if $M = M_n$ for some n, then $M = M_n$ for all n sufficiently large. Hence it follows that $P_{M_n} = P_M$ for all n sufficiently large, and so for any $\phi \in \widetilde{\mathcal{O}}$ and any $x \in H$, we have $\|P_{M_n}x - P_M x\| = 0 < \phi(n)$ for all n large, so arbitrarily slow convergence is not possible.

6. Application to cyclic projections

In this section, we give an application of the Dichotomy Theorem 4.4 to cyclic projections in Hilbert space, or, more precisely, to the von Neumann–Halperin theorem. The von Neumann–Halperin theorem has had many far-reaching applications in at least a dozen different areas of mathematics including solving linear equations, linear prediction theory, image restoration, and computed tomography (see the survey [11], or the book [13, Chapter 9] for more details and references).

Theorem 6.1 (*von Neumann–Halperin*). *Let $M_1, M_2, \ldots, M_r$ be closed subspaces of the Hilbert space H and $M = \cap_1^r M_i$. Then, for each $x \in H$,*

$$\lim_{n\to\infty} \|(P_{M_r}P_{M_{r-1}}\cdots P_{M_1})^n(x) - P_M(x)\| = 0.$$

In the two-subspace case ($r = 2$), this result was first proved by von Neumann in 1933 (but wasn't published until 1949–50 [25,26]). Halperin [21] extended the von Neumann theorem to any number $r \geq 2$ of subspaces. The von Neumann theorem (i.e., the $r = 2$ case) was discovered independently by several authors including Aronszajn [2], Nakano [24], Wiener [32], Powell [27], Gordon et al. [20] and Hounsfield [22]—the Nobel Prize winning inventor of the EMI scanner.

Given any $x \in H$, let $x_n := (P_{M_r}P_{M_{r-1}}\cdots P_{M_1})^n(x)$ for each $n \in \mathbb{N}$. We note that the von Neumann–Halperin theorem shows that the sequence (x_n) always converges to $P_M(x)$. However, the theorem says nothing about the *rate* of convergence. To say something about this, we need the following fact.

Fact 6.2. *Let $M_1, M_2, \ldots, M_r$ be closed subspaces of the Hilbert space H and $M := \cap_1^r M_i$. Then the following statements are equivalent:*

(1) $\sum_1^r M_i^\perp$ *is closed.*
(2) $\|P_{M_r\cap M^\perp}P_{M_{r-1}\cap M^\perp}\cdots P_{M_1\cap M^\perp}\| < 1.$

(3) *There exists* $\alpha \in [0, 1)$ *such that*

$$\|(P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp})^n\| \le \alpha^n \quad \textit{for each } n \in \mathbb{N}.$$

Proof. Bauschke et al. [5, Theorem 3.7.4] proved the equivalence of statements (1) and (2). Clearly, (3) implies (2) (take $n = 1$). Finally, if (2) holds, then letting $\alpha := \|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\| < 1$, it follows that (3) holds. □

Remark 6.3. It is worth mentioning that the statement (1) of Fact 6.2 has an equivalent formulation because: $\sum_1^r M_i^\perp$ *is closed if and only if the collection of subspaces* $\{M_1, M_2, \ldots, M_r\}$ *has the "strong CHIP" property.* (This is an immediate consequence of the proof of Example 10.5 in [13]). The strong CHIP property was shown to be a fundamental property that arose in, for example, constrained interpolation [15,16], convex optimization [12], various types of "regularity" and Jameson's property (G) [6]. For more detail and references, see the historical notes on pp. 283–285 of [13].

The *cyclic projections algorithm* for the subspaces $\{M_1, M_2, \ldots, M_r\}$ is the algorithm that generates, starting with any $x \in H$, the sequence

$$x_0 := x, \quad \text{and} \quad x_n := P_{M_{[n]}}(x_{n-1}) \quad \text{for each } n \in \mathbb{N},$$

where $[n]$ is the function "mod r" with values in $\{1, 2, \ldots, r\}$. That is,

$$[n] = \{1, 2, \ldots, r\} \cap \{n - kr \mid k = 1, 2, \ldots\}.$$

In particular, $x_{nr} = (P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n(x)$. In this terminology, the von Neumann–Halperin theorem shows that, for each $x \in H$, the cyclic projections algorithm generates a sequence that converges to $P_M(x)$.

One important corollary of the Dichotomy Theorem 4.4 is what we call the *von Neumann–Halperin dichotomy*.

Theorem 6.4 (*von Neumann–Halperin Dichotomy*)**.** *Let* $M_1, M_2, \ldots, M_r$ *be closed subspaces of the Hilbert space* H *and* $M := \cap_1^r M_i$. *Then exactly one of the following two statements holds.*

(1) $\sum_1^r M_i^\perp$ *is closed. Then* $((P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n)$ *converges to* P_M *linearly.*
(2) $\sum_1^r M_i^\perp$ *is not closed. Then* $((P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n)$ *converges to* P_M *arbitrarily slowly.*

Proof. If $\sum_1^r M_i^\perp$ is closed, then Fact 6.2 implies that there exists $\alpha \in [0, 1)$ such that

$$\|(P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp})^n\| \le \alpha^n \quad \text{for each } n \in \mathbb{N}. \tag{6.1}$$

However, by [13, Lemma 9.30, p. 218], we have the identity

$$(P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp})^n = (P_{M_r} \cdots P_{M_1})^n - P_M \quad \text{for each } n \in \mathbb{N}. \tag{6.2}$$

Using (6.1) and (6.2) we deduce that

$$\|(P_{M_r} \cdots P_{M_1})^n - P_M\| \le \alpha^n \quad \text{for each } n \in \mathbb{N}.$$

This proves statement (1).

If $\sum_1^r M_i^\perp$ is not closed, then from Fact 6.2, it follows that

$$\|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\| = 1. \tag{6.3}$$

From this we can conclude that $\|(P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp})^n\| = 1$ for each $n \in \mathbb{N}$. (For otherwise, $\|(P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp})^{n_1}\| < 1$ for some $n_1 \in \mathbb{N}$. Then Fact 6.2 would imply that the sum of the n_1 sums $\sum_1^r M_i^\perp + \sum_1^r M_i^\perp + \cdots + \sum_1^r M_i^\perp$ is closed. But this sum of subspaces reduces to the single sum $\sum_1^r M_i^\perp$, which is non-closed by assumption, and this is a contradiction.) Let

$$T = P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp}.$$

Then $T : H \to H$ is linear, $\|T^n\| = 1$ for each $n \in \mathbb{N}$ and, using the von Neumann–Halperin theorem along with (6.2), it follows that $T^n(x) \to 0$ for each $x \in H$. Now apply the Dichotomy Theorem 4.4 to obtain statement (2). □

Remark 6.5. (1) For the special case of two subspaces ($r = 2$), this result was stated by Bauschke et al. [5]. Bauschke et al. [7] found an error in the proof of [5] that invalidated the proof of the theorem, but they showed that the theorem was nevertheless true by providing an alternate proof. Briefly, the case of $r = 2$ in Theorem 6.4 was due to [5,7], established by a substantially different and more involved proof than that given here.

(2) We suspect that when $\sum_1^r M_i^\perp$ is not closed and $\phi \in \widetilde{\mathcal{O}}$, then the $x = x_\phi$ that satisfies $\|((P_{M_r} \cdots P_{M_1})^n(x) - P_M(x))\| \geq \phi(n)$ for all n must in general be chosen from $M^\perp \setminus \sum_1^r M_i^\perp$.

By a translation argument, we can easily generalize Theorem 6.4 to the case of affine sets. Recall that a set V is *affine* if $\alpha x + (1-\alpha)y \in V$ whenever $x, y \in V$ and $\alpha \in \mathbb{R}$. Equivalently, V is the translate of a (unique) subspace M: $V = M + v$ for any $v \in V$. Indeed, $M = V - V$ (see, e.g., [13, Theorem 10.17, p. 247].)

Theorem 6.6 (*Affine Sets Dichotomy*)**.** *Let $V_1, V_2, \ldots, V_r$ be closed affine sets in the Hilbert space H with $V := \cap_1^r V_i \neq \emptyset$. Then exactly one of the following two statements holds.*

(1) *$\sum_1^r (V_i - V_i)^\perp$ is closed. Then $((P_{V_r} P_{V_{r-1}} \cdots P_{V_1})^n)$ converges to P_V linearly.*
(2) *$\sum_1^r (V_i - V_i)^\perp$ is not closed. Then $((P_{V_r} P_{V_{r-1}} \cdots P_{V_1})^n)$ converges to P_V arbitrarily slowly.*

Proof. Let $M_i := V_i - V_i$ $(i = 1, 2, \ldots, r)$, $M = \cap_1^r M_i$, and fix any $v \in V$. Then $V_i = M_i + v$ $(i = 1, 2, \ldots, r)$. Recall (see, e.g., [13, Theorem 2.7(ii), p. 25]) that

$$P_S(x) = P_{S-y}(x - y) + y$$

for any set S in H and any points $x, y \in H$. By repeated application of this fact, we can readily deduce that

$$(P_{V_r} \cdots P_{V_1})^n(x) - P_V(x) = (P_{M_r} \cdots P_{M_1})^n(x - v) - P_M(x - v) \tag{6.4}$$

for each $x \in H, n \in \mathbb{N}$. It follows from this that $((P_{V_r} \cdots P_{V_1})^n)$ converges linearly (respectively, arbitrarily slowly) to P_V if and only if $((P_{M_r} \cdots P_{M_1})^n)$ converges linearly (respectively, arbitrarily slowly) to P_M. But by the von Neumann–Halperin dichotomy (Theorem 6.4), the latter happens if and only if $\sum_1^r M_i^\perp$ is closed. Since $M_i = V_i - V_i$, the theorem is now an immediate consequence of Theorem 6.4. □

The following two corollaries will be useful for comparison to the results of Xu and Zikatanov in Section 8. If in the von Neumann–Halperin dichotomy theorem one replaces each subspace M_i by its orthogonal complement $M_i^\perp$ and recalls the well-known facts that $P_{M_i^\perp} = I - P_{M_i}$, $M_i^{\perp\perp} = M_i$, and $\cap_1^r M_i^\perp = (\sum_1^r M_i)^\perp$ (see, e.g., [13]), then one obtains:

Corollary 6.7. *Let $M_1, M_2, \ldots, M_r$ be closed subspaces in the Hilbert space H and let $M := \overline{\sum_1^r M_i}$. Then exactly one of the following two statements holds.*

(1) *$\sum_1^r M_i$ is closed. Then $([(I-P_{M_r})(I-P_{M_{r-1}})\cdots(I-P_{M_1})]^n)$ converges to $I-P_M$ linearly.*
(2) *$\sum_1^r M_i$ is not closed. Then $([(I-P_{M_r})(I-P_{M_{r-1}})\cdots(I-P_{M_1})]^n)$ converges to $I-P_M$ arbitrarily slowly.*

By a proof analogous to that of Theorem 6.6, we obtain the following consequence of Corollary 6.7 that is actually more general than Corollary 6.7.

Theorem 6.8. *Let $V_1, V_2, \ldots, V_r$ be closed affine sets in H with $V := \cap_1^r V_i \neq \emptyset$. Then exactly one of the following two statements holds.*

(1) *$\sum_1^r (V_i - V_i)$ is closed. Then $[(I-P_{V_r})(I-P_{V_{r-1}})\cdots(I-P_{V_1})]^n$ converges to $I-P_V$ linearly.*
(2) *$\sum_1^r (V_i - V_i)$ is not closed. Then $[(I-P_{V_r})(I-P_{V_{r-1}})\cdots(I-P_{V_1})]^n$ converges to $I-P_V$ arbitrarily slowly.*

7. Application to intermittent projections

In this section we give an application of the Dichotomy Theorem 4.4 to intermittent or "almost" randomly ordered projections. Throughout this section H will always denote a Hilbert space and $M_1, \ldots, M_r$ will be a collection of r closed subspaces in H with $M := \cap_1^r M_i$.

Definition 7.1. A function $\sigma : \mathbb{N} \to \{1, 2, \ldots, r\}$ is called a *random selection* for $\{1, 2, \ldots, r\}$ if for each $n \in \mathbb{N}$, there exists $N(n) \in \mathbb{N}$, $N(n) > n$, such that

$$\{\sigma(n), \sigma(n+1), \ldots, \sigma(N(n))\} = \{1, 2, \ldots, r\}. \tag{7.1}$$

The following is easy to verify.

Lemma 7.2. *Let $\sigma : \mathbb{N} \to \{1, 2, \ldots, r\}$. Then the following statements are equivalent:*

(1) *σ is a random selection for $\{1, 2, \ldots, r\}$.*
(2) *The range of σ is $\{1, 2, \ldots, r\}$ and σ assumes each value in its range infinitely often.*

A *random product* of the projections $P_{M_1}, P_{M_2}, \ldots, P_{M_r}$ is the sequence (S_n), where

$$S_n := P_{M_{\sigma(n)}} P_{M_{\sigma(n-1)}} \cdots P_{M_{\sigma(1)}} \quad (n = 1, 2, \ldots), \tag{7.2}$$

and where σ is a random selection for $\{1, 2, \ldots, r\}$.

Recall that a sequence (x_n) in H is said to converge *weakly* to $x \in H$ provided that

$$\lim_{n\to\infty} \langle x_n, z\rangle = \langle x, z\rangle \quad \text{for each } z \in H.$$

Fact 7.3 (*[1]*). *If (S_n) is the random product of projections (7.2), then for each $x \in H$ the sequence $(S_n(x))$ converges weakly to $P_M(x)$.*

For some far-reaching generalizations of Fact 7.3, see Dye et al. [18].

Apparently, it is still unknown whether or not the convergence in Fact 7.3 must be in norm. However, when certain additional conditions are imposed on either the subspaces M_i or the function σ, then norm convergence in Fact 7.3 is indeed guaranteed.

One result along these lines is the following.

Fact 7.4 (*Bauschke [3, Example 3.8]*). *Let* (S_n) *be the random product of projections* (7.2). *If* $\sum_{i\in J} M_i^\perp$ *is closed for each nonempty subset* J *of* $\{1, 2, \ldots, r\}$, *then*

$$\lim_n \|S_n(x) - P_M(x)\| = 0 \quad \textit{for each } x \in H.$$

We do not know whether Bauschke's result 7.4 is valid under the weaker condition that only $\sum_1^r M_i^\perp$ be closed. However, with an additional condition on the function σ, then the answer is affirmative.

Definition 7.5 (*[4, Definition 3.18]*). A random selection function $\sigma : \mathbb{N} \to \{1, 2, \ldots, r\}$ is called an *intermittent selection* for $\{1, 2, \ldots, r\}$ if there exists $N_1 \in \mathbb{N}$ such that, for each $n \in \mathbb{N}$,

$$\{\sigma(n), \sigma(n+1), \ldots, \sigma(n+N_1)\} = \{1, 2, \ldots, r\}. \tag{7.3}$$

Note that an intermittent selection is a random selection with the property that for each $n \in \mathbb{N}$, the $N(n)$ that works in the definition of random selection does *not* depend on n, but is some fixed N_1 that works for all n.

An *intermittent product* of the projections $P_{M_1}, P_{M_2}, \ldots, P_{M_n}$ is the sequence (S_n), where

$$S_n := P_{M_{\sigma(n)}} P_{M_{\sigma(n-1)}} \cdots P_{M_{\sigma(1)}} \quad (n = 1, 2, \ldots), \tag{7.4}$$

and where $\sigma : \mathbb{N} \to \{1, 2, \ldots, r\}$ is an intermittent selection for $\{1, 2, \ldots, r\}$.

Note that if we define $\sigma(n) := [n]$, where $[\cdot]$ is the function "mod r", i.e.,

$$[n] := \{n - kr \mid k = 0, 1, 2, \ldots\} \cap \{1, 2, \ldots, r\},$$

then the function σ satisfies the above hypothesis (with $N_1 = r - 1$) and

$$S_{rn} = (P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n \tag{7.5}$$

is just the sequence of "cyclically" ordered projections that appeared in the von Neumann–Halperin theorem of the preceding section.

We need the following fact, which generalizes the von Neumann–Halperin cyclic projections theorem to intermittently ordered projections.

Fact 7.6 (*Hundal and Deutsch [23, Subspace Case of Theorem 3.1]*). *Let* $\sigma : \mathbb{N} \to \{1, 2, \ldots, r\}$ *be an intermittent selection function, and let* S_n *be the intermittent product* (7.4). *Then*

$$\lim_n \|S_n(x) - P_M(x)\| = 0 \quad \textit{for each } x \in H.$$

Theorem 7.7. *Assume the hypothesis of* Fact 7.6 *with* σ *an intermittent selection. Then exactly one of the following two statements holds:*

(1) $\sum_1^r M_i^\perp$ *is closed; then* (S_n) *converges to* P_M *linearly.*
(2) $\sum_1^r M_i^\perp$ *is not closed; then* (S_n) *converges to* P_M *arbitrarily slowly.*

Proof. Note that, using property (7.3), we deduce

$$\sum_1^r M_i^\perp = \sum_{(n-1)N+1}^{nN} M_{\sigma(i)}^\perp = \sum_{i=1}^{nN} M_{\sigma(i)}^\perp \quad \text{for each } n \in \mathbb{N}. \tag{7.6}$$

Let $L_n := S_n - P_M$ for each $n \in \mathbb{N}$. By Fact 7.6, (L_n) converges to 0 pointwise. Also, using some basic properties of projections (see, e.g., [13, Lemma 9.30, p. 218], we obtain

$$\begin{aligned} L_n &= P_{M_{\sigma(nN)}} P_{M_{\sigma(nN-1)}} \cdots P_{M_{\sigma(1)}} - P_M \\ &= P_{M_{\sigma(nN)}} P_{M_{\sigma(nN-1)}} \cdots P_{M_{\sigma(1)}} (I - P_M) \\ &= P_{M_{\sigma(nN)}} P_{M_{\sigma(nN-1)}} \cdots P_{M_{\sigma(1)}} P_{M^\perp} \\ &= (P_{M_{\sigma(nN)}} P_{M^\perp})(P_{M_{\sigma(nN-1)}} P_{M^\perp}) \cdots (P_{M_{\sigma(1)}} P_{M^\perp}) \\ &= P_{M_{\sigma(nN)} \cap M^\perp} P_{M_{\sigma(nN-1)} \cap M^\perp} \cdots P_{M_{\sigma(1)} \cap M^\perp}. \end{aligned}$$

In addition, since L_n is a product of projections, it is easy to see that

$$\|L_n\| \le 1 \quad \text{and} \quad \|L_{n+1}(x)\| \le \|L_n(x)\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}. \tag{7.7}$$

We consider the following two cases.

Case i: $\sum_1^r M_i^\perp$ is not closed.

By (7.6), we see that $\sum_{i=1}^{nN} M_{\sigma(i)}^\perp$ is not closed for each $n \in \mathbb{N}$. By Fact 6.2, it follows that $\|L_n\| = 1$ for each $n \in \mathbb{N}$. By the Lethargy Theorem 3.3, (L_n) must converge to 0 arbitrarily slowly. Hence (S_n) converges to P_M arbitrarily slowly.

Case ii: $\sum_1^r M_i^\perp$ is closed.

By (7.6), $\sum_{i=(n-1)N+1}^{nN} M_{\sigma(i)}$ is closed for each $n \in \mathbb{N}$. For each $m \in \mathbb{N}$, set

$$\alpha_m := \|P_{M_{\sigma(mN)} \cap M^\perp} P_{M_{\sigma(mN-1)} \cap M^\perp} \cdots P_{M_{\sigma((m-1)N+1)} \cap M^\perp}\|. \tag{7.8}$$

By Fact 6.2 it follows that $\alpha_m < 1$ for each m, and

$$\|L_n\| \le \alpha_n \alpha_{n-1} \cdots \alpha_1. \tag{7.9}$$

Note that α_m only depends on the ordering of the N subspaces $M_{\sigma(mN)} \cap M^\perp$, $M_{\sigma(mN-1)} \cap M^\perp, \ldots, M_{\sigma((m-1)N)} \cap M^\perp$. Thus the set $\{\alpha_m \mid m = 1, 2, \ldots\}$ is finite. Letting $\alpha := \max_m \alpha_m$, it follows that $\alpha < 1$ and $\|L_n\| \le \alpha^n$ for each n. This shows that (L_n) converges to 0 linearly, which implies that (S_n) converges to P_M linearly. □

Remark 7.8. (1) It should be mentioned that statement (1) of Theorem 7.7 is also a consequence of a result of Bauschke and Borwein [4, Theorem 5.7].

(2) Note that Theorem 7.7 is a generalization of the Von Neumann–Halperin Dichotomy (Theorem 6.4).

8. Application to a Xu–Zikatanov Theorem

In their fundamental paper [33], Xu and Zikatanov wove a beautiful tapestry connecting the method of alternating projections with the method of subspace corrections in a Hilbert space. The method of subspace corrections is applied in the area of finite-element analysis and is also referred to as domain decomposition or the multigrid method. We shall sharpen and refine one of their two main results (viz., [33, Theorem 4.7]) by using the Dichotomy Theorem 4.4.

Let H be a Hilbert space, let $V_1, V_2, \ldots, V_r$ be closed subspaces of H, and let $T_i : H \to V_i$ be bounded linear mappings satisfying the following two assumptions for each $i = 1, 2, \ldots, r$:

(A1) The range of T_i is V_i and $T_i|_{V_i} : V_i \to V_i$ is an isomorphism.
(A2) $\|T_i(x)\|^2 \le \omega \langle T_i(x), x \rangle$ for each $x \in H$ and some constant $\omega \in (0, 2)$.

In particular, assumption (A2) guarantees that $I - T_i$ is nonexpansive: $\|I - T_i\| \le 1$. (In the applications that are made in [33], the T_i may be regarded as *approximations* to the projections P_{V_i}.

Typically, the T_i correspond to damped Jacobi, Gauss–Seidel, or successive overrelaxation methods applied at different mesh resolutions.)

Let

$$E := (I - T_r)(I - T_{r-1}) \cdots (I - T_1), \tag{8.1}$$

$$\mathrm{Fix}(E) := \{x \in H \mid E(x) = x\}, \quad \text{and} \tag{8.2}$$

$$\mathcal{N}(T_i) := \{x \in H \mid T_i(x) = 0\}. \tag{8.3}$$

Fact 8.1 (*Xu–Zikatanov [33, Lemma 4.4]*).

$$M := \mathrm{Fix}(E) = \bigcap_1^r \mathcal{N}(T_i) = \bigcap_1^r V_i^{\perp}, \quad \textit{and} \tag{8.4}$$

$$V := M^{\perp} = \overline{\sum_1^r V_i}. \tag{8.5}$$

Fact 8.2 (*Xu–Zikatanov [33, Theorem 4.6]*). *The following two statements are equivalent:*
(1) $\sum_1^r V_i$ *is closed;*
(2) $\|EP_V\| < 1$.

Next note the identities

$$\begin{aligned} E^n - (I - P_V) &= E^n - P_M = (E - P_M)^n = [E(I - P_M)]^n \\ &= (EP_{M^{\perp}})^n = (EP_V)^n. \end{aligned} \tag{8.6}$$

The second equality $E^n - P_M = (E - P_M)^n$ follows from the fact that $P_M E = P_M = EP_M$, which in turn is a consequence of the (not obvious) fact that $T_i = T_i P_{V_i} = P_{V_i} T_i$ (see [33, Eq. (2.10)]), and hence that $P_M T_i = P_M P_{V_i} T_i = 0$ and $T_i P_M = P_i P_{V_i} P_M = 0$ for all i since $P_{V_i} P_M = 0 = P_M P_{V_i}$.

Fact 8.3 (*Xu–Zikatanov [33, Theorem 4.7]*).

$$\lim_{n\to\infty} \|E^n(x) - (I - P_V)(x)\| = 0 \quad \textit{for each } x \in H, \tag{8.7}$$

or equivalently,

$$\lim_{n\to\infty} \|(EP_V)^n(x)\| = 0 \quad \textit{for each } x \in H. \tag{8.8}$$

Since $\|EP_V\| \le 1$, it follows from the Dichotomy Theorem 4.4 and Facts 8.2 and 8.3 that we obtain the following dichotomy pertaining to the Xu–Zikatanov theory.

Theorem 8.4 (*Xu–Zikatanov Dichotomy*). *Exactly one of the following two statements holds.*
(1) $\sum_1^r V_i$ *is closed. Then* (E^n) *converges to* $(I - P_V)$ *linearly.*
(2) $\sum_1^r V_i$ *is not closed. Then* (E^n) *converges to* $(I - P_V)$ *arbitrarily slowly.*

We should note that this result is somewhat more general than Theorem 6.8 since here the T_i need not be equal to P_{V_i}, but need only be a certain kind of approximation to P_{V_i}.

9. Further applications

A Google search for the term "arbitrarily slow convergence" brings up hits in the areas of probability and statistics (density estimation [17], the central limit theorem [28], and Gibbs sampling [19]) and numerical methods for linear inverse problems [30]. The Lethargy Theorem and arguments similar to the proof of the Lethargy Theorem can be used to reproduce many of the arbitrarily slow or almost arbitrarily slow convergence results for these applications. Unfortunately, our proofs are lengthy so they are not included here.

Acknowledgment

We are greatly indebted to two referees for their comments and questions which helped us to make the final version of this paper more complete and definitive.

References

[1] I. Amemiya, T. Ando, Convergence of random products of contractions in Hilbert space, Acta Sci. Math. (Szeged) 26 (1965) 239–244.
[2] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–403.
[3] H.H. Bauschke, A norm convergence result on random products of relaxed projections in Hilbert space, Trans. Amer. Math. Soc. 347 (1995) 1365–1373.
[4] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (1996) 367–426.
[5] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, Contemp. Math. 204 (1997) 1–38.
[6] H.H. Bauschke, J.M. Borwein, W. Li, The strong conical hull intersection property, bounded linear regularity, Jameson's property(G), and error bounds in convex optimization, Math. Program. 86 (1) (1999) 135–160.
[7] H.H. Bauschke, F. Deutsch, H. Hundal, Characterizing arbitrarily slow convergence in the method of alternating projections, Int. Trans. Oper. Res. 16 (2009) 413–435.
[8] S.N. Bernstein, On the inverse problem of the theory of the best approximation of continuous functions, Sochineniya II (1938) 292–294.
[9] P.L. Combettes, Fejér-monotonicity in convex optimization, in: C.A. Floudas, P.M. Pardalos (Eds.), Encyclopedia of Optimization, Kluwer Acad. Pub., 2000.
[10] P.J. Davis, Interpolation and Approximation, Blaisdell, New York, 1963.
[11] F. Deutsch, The method of alternating orthogonal projections, in: S.P. Singh (Ed.), Approximation Theory, Spline Functions and Applications, Kluwer Acad. Pub., the Netherlands, 1992, pp. 105–121.
[12] F. Deutsch, The role of the strong conical hull intersection property in convex optimization and approximation, in: C.K. Chui, L.L. Schumaker (Eds.), Approximation Theory IX, Vanderbilt University Press, Nashville, TN, 1998, pp. 143–150.
[13] F. Deutsch, Best Approximation in Inner Product Spaces, Springer-Verlag, New York, 2001.
[14] F. Deutsch, H. Hundal, Slow convergence of sequences of linear operators I: almost arbitrarily slow convergence, J. Approx. Theory 162 (2010) 1701–1716.
[15] F. Deutsch, W. Li, J.D. Ward, A dual approach to constrained interpolation from a convex subset of Hilbert space, J. Approx. Theory 80 (1997) 381–405.
[16] F. Deutsch, V.A. Ubhaya, J.D. Ward, Y. Xu, Constrained best approximation in Hilbert space III. Applications to n-convex functions, Constr. Approx. 12 (1996) 361–384.
[17] L. Devroye, On arbitrarily slow rates of global convergence in density estimation, Probab. Theory Related Fields 62 (1983) 475–483.
[18] J. Dye, M.A. Khamsi, S. Reich, Random products of contractions in Banach spaces, Trans. Amer. Math. Soc. 325 (1991) 87–99.
[19] A. Golightly, D.J. Wilkinson, Bayesian inference for nonlinear multivariate diffusion models observed with error, Comput. Statist. Data Anal. 52 (2008) 1674–1693.
[20] R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, J. Theoret. Biol. 29 (1970) 471–481.
[21] I. Halperin, The product of projection operators, Acta Sci. Math. (Szeged) 23 (1962) 96–99.
[22] G.N. Hounsfield, Computerized transverse axial scanning (tomography); Part I description of system, Br. J. Radiol. 46 (1973) 1016–1022.
[23] H. Hundal, F. Deutsch, Two generalizations of Dykstra's cyclic projections algorithm, Math. Program. 77 (1997) 335–355.

[24] H. Nakano, Spectral theory in the Hilbert space, Japan Soc. Promotion Sc., 1953, Tokyo.
[25] J. von Neumann, On rings of operators. Reduction theory, Ann. of Math. 50 (1949) 401–485.
[26] J. von Neumann, Functional Operators—Vol. II. The Geometry of Orthogonal Spaces, in: Annals of Math. Studies, #22, Princeton University Press, Princeton, NJ, 1950 [This is a reprint of mimeographed lecture notes first distributed in 1933.].
[27] M.J.D. Powell, A new algorithm for unconstrained optimization, in: J.B. Rosen, O.L. Mangasarian, K. Ritter (Eds.), Nonlinear Programming, Academic Press, New York, 1970.
[28] W. Rhee, M. Talagrand, Bad rates of convergence for the central limit theorem in Hilbert space, Ann. Probab. 12 (1984) 843–850.
[29] E. Schock, Arbitrarily slow convergence, uniform convergence and superconvergence of Galerkin-like methods, IMA J. Numer. Anal. 5 (1985) 153–160.
[30] E. Schock, Semi-iterative methods for the approximated solutions of ill-posed problems, Numer. Math. 50 (1987) 263–271.
[31] A.F. Timan, Theory of Approximation of Functions of a Real Variable, MacMillan, New York, 1963.
[32] N. Wiener, On the factorization of matrices, Comment. Math. Helv. 29 (1955) 97–111.
[33] J. Xu, L. Zikatanov, The method of alternating projections and the method of subspace corrections in Hilbert space, J. Amer. Math. Soc. 15 (2002) 573–597.