1

Greedy approximation with regard to bases

1.1 Introduction

It is well known that in many problems it is very convenient to represent a function by a series with regard to a given system of functions. For example, in 1807 Fourier suggested representing a 2π-periodic function by its series (known as the Fourier series) with respect to the trigonometric system. A very important feature of the trigonometric system that made it attractive for the representation of periodic functions is orthogonality. For an orthonormal system $\mathcal{B} := \{b_n\}_{n=1}^{\infty}$ of a Hilbert space H with an inner product $\langle \cdot, \cdot \rangle$, one can construct a Fourier series of an element f in the following way:

$$f \sim \sum_{n=1}^{\infty} \langle f, b_n \rangle b_n. \tag{1.1}$$

If the system $\mathcal{B}$ is a basis for H, then the series in (1.1) converges to f in H and (1.1) provides a unique representation

$$f = \sum_{n=1}^{\infty} \langle f, b_n \rangle b_n \tag{1.2}$$

of f with respect to $\mathcal{B}$. This representation has nice approximative properties. By Parseval's identity,

$$\|f\|^2 = \sum_{n=1}^{\infty} |\langle f, b_n \rangle|^2, \tag{1.3}$$

we obtain a convenient way to calculate, or estimate, the norm $\|f\|$.

It is known that the partial sums

$$S_m(f, \mathcal{B}) := \sum_{n=1}^{m} \langle f, b_n \rangle b_n \tag{1.4}$$

provide the best approximation; that is, defining

$$E_m(f,\mathcal{B}) := \inf_{\{c_n\}} \|f - \sum_{n=1}^{m} c_n b_n\| \tag{1.5}$$

to be the distance of f from the span$\{b_1,\dots,b_m\}$, we have

$$\|f - S_m(f,\mathcal{B})\| = E_m(f,\mathcal{B}). \tag{1.6}$$

Identities (1.3) and (1.6) are fundamental properties of Hilbert spaces and their orthonormal bases. These properties make the theory of approximation in H from the span$\{b_1,\dots,b_m\}$, or linear approximation theory, simple and convenient.

The situation becomes more complicated when we replace a Hilbert space H by a Banach space X. In a Banach space X we consider a Schauder basis Ψ instead of an orthonormal basis $\mathcal{B}$ in H. In Section 1.2 we discuss Schauder bases in detail. If $\Psi := \{\psi_n\}_{n=1}^{\infty}$ is a Schauder basis for X, then for any $f \in X$ there exists a unique representation

$$f = \sum_{n=1}^{\infty} c_n(f,\Psi)\psi_n$$

that converges in X.

Theorem 1.3 from Section 1.2 states that the partial sum operators S_m, defined by

$$S_m(f,\Psi) := \sum_{n=1}^{m} c_n(f,\Psi)\psi_n,$$

are uniformly bounded operators from X to X. In other words, there exists a constant B such that, for any $f \in X$ and any m, we have

$$\|S_m(f,\Psi)\| \le B\|f\|.$$

This inequality implies an analog of (1.6): for any $f \in X$,

$$\|f - S_m(f,\Psi)\| \le (B+1)E_m(f,\Psi), \tag{1.7}$$

where

$$E_m(f,\Psi) := \inf_{\{c_n\}} \|f - \sum_{n=1}^{m} c_n\psi_n\|.$$

Inequality (1.7) shows that the $S_m(f,\Psi)$ provides near-best approximation from span$\{\psi_1,\dots,\psi_m\}$. Thus, if we are satisfied with near-best approximation instead of best approximation, then the linear approximation theory with respect to Schauder bases becomes simple and convenient. The partial sums $S_m(\cdot, \Psi)$ provide near-best approximation for any individual element of X.

Motivated by computational issues, researchers became interested in nonlinear approximation with regard to a given system instead of linear approximation. For example, in the case of representation (1.2) in a Hilbert space, one can take an approximant of the form

$$S_\Lambda(f, \mathcal{B}) := \sum_{n\in\Lambda} \langle f, b_n\rangle b_n, \quad |\Lambda| = m,$$

instead of an approximant $S_m(f, \mathcal{B})$ from an m-dimensional linear subspace. Then the two approximants $S_m(f, \mathcal{B})$ and $S_\Lambda(f, \mathcal{B})$ have the same sparsity: both are linear combinations of m basis elements. However, we can achieve a better approximation error with $S_\Lambda(f, \mathcal{B})$ than with $S_m(f, \mathcal{B})$ if we choose Λ correctly. In the case of a Hilbert space and an orthonormal basis $\mathcal{B}$, an optimal choice Λ_m of Λ is obvious: Λ_m is a set of m indices with the biggest (in absolute value) coefficients $\langle f, b_n\rangle$. Then, by Parseval's identity (1.3), we obtain

$$\|f - S_{\Lambda_m}(f, \mathcal{B})\| \le \|f - S_m(f, \mathcal{B})\|.$$

Also, it is clear that the $S_{\Lambda_m}(f, \mathcal{B})$ realizes the best m-term approximation of f with regard to $\mathcal{B}$,

$$\|f - S_{\Lambda_m}(f, \mathcal{B})\| = \sigma_m(f, \mathcal{B}) := \inf_{\Lambda:|\Lambda|=m} \inf_{\{c_n\}} \|f - \sum_{n\in\Lambda} c_n b_n\|. \tag{1.8}$$

The approximant $S_{\Lambda_m}(f, \mathcal{B})$ can be obtained as a realization of m iterations of the *greedy approximation step*. For a given $f \in H$ we choose at a *greedy step* an index n_1 with the biggest $|\langle f, b_{n_1}\rangle|$. At a greedy approximation step we build a new element $f_1 := f - \langle f, b_{n_1}\rangle b_{n_1}$.

The identity (1.8) shows that the greedy approximation works perfectly in nonlinear approximation in a Hilbert space with regard to orthonormal basis $\mathcal{B}$.

This chapter is devoted to a systematic study of greedy approximation in Banach spaces. In Section 1.2 we discuss the following natural question. Equation (1.8) proves the existence of the best m-term approximant in a Hilbert space with respect to an orthonormal basis. Further, we discuss the existence of the best m-term approximant in a Banach space with respect to a Schauder basis. That discussion illustrates that the situation regarding existence theorems is much more complex in Banach spaces than in Hilbert spaces. We also give some sufficient conditions on a Schauder basis that guarantee the existence of the best m-term approximant. However, the problem is far from being completely solved.

The central issue of this chapter is the question: Which bases are suitable for greedy approximation? Greedy approximation with regard to a Schauder basis is defined in a similar way to the greedy approximation with regard to an orthonormal basis (see above). The greedy algorithm picks the terms with the biggest (in absolute value) coefficients from the expansion

$$f = \sum_{n=1}^{\infty} c_n(f, \Psi)\psi_n \tag{1.9}$$

and gives a *greedy approximant*

$$G_m(f, \Psi) := S_{\Lambda_m}(f, \Psi) := \sum_{n \in \Lambda_m} c_n(f, \Psi)\psi_n.$$

Here, Λ_m is such that $|\Lambda_m| = m$ and

$$\min_{n \in \Lambda_m} |c_n(f, \Psi)| \geq \max_{n \notin \Lambda_m} |c_n(f, \Psi)|.$$

We note that we need some restrictions on the basis Ψ (see Sections 1.3 and 1.4 for a detailed discussion) in order to be able to run the greedy algorithm for each $f \in X$. It is sufficient to assume that Ψ is normalized. We make this assumption for our further discussion in the Introduction. In some later sections we continue to use the normalization assumption; in others, we do not.

An application of the greedy algorithm can also be seen as a rearrangement of the series from (1.9) in a special way: according to the size of coefficients. Let

$$|c_{n_1}(f, \Psi)| \geq |c_{n_2}(f, \Psi)| \geq \dots .$$

Then

$$G_m(f, \Psi) = \sum_{j=1}^{m} c_{n_j}(f, \Psi)\psi_{n_j}.$$

Thus, the greedy approximant $G_m(f, \Psi)$ is a partial sum of the rearranged series

$$\sum_{j=1}^{\infty} c_{n_j}(f, \Psi)\psi_{n_j}. \tag{1.10}$$

An immediate question arising from (1.10) is: When does this series converge? The theory of convergence of rearranged series is a classical topic in analysis. A series converges *unconditionally* if every rearrangement of this series converges. A basis Ψ of a Banach space X is said to be an *unconditional basis* if, for every $f \in X$, its expansion (1.9) converges unconditionally. For a set of indices Λ define

$$S_\Lambda(f, \Psi) := \sum_{n\in\Lambda} c_n(f, \Psi)\psi_n.$$

It is well known that if Ψ is unconditional then there exists a constant K such that, for any Λ,

$$\|S_\Lambda(f, \Psi)\| \leq K\|f\|. \tag{1.11}$$

This inequality is similar to $\|S_m(f, \Psi)\| \leq B\|f\|$ and implies an analog of inequality (1.7)

$$\|f - S_\Lambda(f, \Psi)\| \leq (K + 1)E_\Lambda(f, \Psi), \tag{1.12}$$

where

$$E_\Lambda(f, \Psi) := \inf_{\{c_n\}} \|f - \sum_{n\in\Lambda} c_n\psi_n\|.$$

Inequality (1.12) indicates that, in the case of an unconditional basis Ψ, it is sufficient for finding a near-best m-term approximant to optimize only over the sets of indices Λ. The greedy algorithm $G_m(\cdot, \Psi)$ gives a simple recipe for building Λ_m: pick the indices with largest coefficients. In Section 1.3 we discuss in detail when the above simple recipe provides a near-best m-term approximant. It turns out that the mere assumption that Ψ is unconditional does not guarantee that $G_m(\cdot, \Psi)$ provides a near-best m-term approximation. We also discuss a new class of bases (*greedy bases*) that has the property that $G_m(f, \Psi)$ provides a near-best m-term approximation for each $f \in X$. We show that the class of greedy bases is a proper subclass of the class of unconditional bases.

It follows from the definition of unconditional basis that any rearrangement of the series in (1.9) converges, and it is known that it converges to f. The rearrangement (1.10) is a specific rearrangement of (1.9). Clearly, for an unconditional basis Ψ, (1.10) converges to f. It turns out that unconditionality of Ψ is not a necessary condition for convergence of (1.10) for each $f \in X$. Bases that have the property of convergence of (1.10) for each $f \in X$ are exactly the *quasi-greedy bases* (see Section 1.4).

Let us summarize our discussion of bases in Banach spaces. Schauder bases are natural for convergence of $S_m(f, \Psi)$ and convenient for linear approximation theory. Other classical bases, namely unconditional bases, are natural for convergence of all rearrangements of expansions. The needs of nonlinear approximation, or, more specifically, the needs of greedy approximation lead us to new concepts of bases: greedy bases and quasi-greedy bases. The relations between these bases are as follows:

$$\{\text{greedy bases}\} \subset \{\text{unconditional bases}\}$$
$$\subset \{\text{quasi-greedy bases}\} \subset \{\text{Schauder bases}\}.$$

All the inclusions $\subset$ are proper inclusions.

In this chapter we provide a justification of the importance of the new classes of bases. With a belief in the importance of greedy bases and quasi-greedy bases, we discuss here the following natural questions: Could we weaken a rule of building $G_m(f, \Psi)$ and still have good approximation and convergence properties? We answer this question in Sections 1.5 and 1.6. What can be said about classical systems, say the Haar system and the trigonometric system, in this regard? We discuss this question in Sections 1.3 and 1.7. How can we build the approximation theory (mostly direct and inverse theorems) for m-term approximation with regard to greedy-type bases? Section 1.8 is devoted to this question.

1.2 Schauder bases in Banach spaces

Schauder bases in Banach spaces are used to associate a sequence of numbers with an element $f \in X$: these are coefficients of f with respect to a basis. This helps in studying properties of a Banach space X. We begin with some classical results on Schauder bases; see, for example, Lindenstrauss and Tzafriri (1977).

Definition 1.1 A sequence $\Psi := \{\psi_n\}_{n=1}^{\infty}$ in a Banach space X is called a Schauder basis of X (basis of X) if, for any $f \in X$, there exists a unique sequence $\{c_n(f)\}_{n=1}^{\infty} := \{c_n(f, \Psi)\}_{n=1}^{\infty}$ such that

$$f = \sum_{n=1}^{\infty} c_n(f)\psi_n.$$

Let

$$S_0(f) := 0, \quad S_m(f) := S_m(f, \Psi) := \sum_{n=1}^{m} c_n(f)\psi_n.$$

For a fixed basis Ψ, consider the quantity

$$|||f||| := \sup_m \|S_m(f, \Psi)\|.$$

It is clear that for any $f \in X$ we have

$$\|f\| \le |||f||| < \infty. \tag{1.13}$$

It is easy to see that $|||\cdot|||$ provides a norm on the linear space X. Denote this new normed linear space by X^s. The following known proposition is not difficult to prove.

Proposition 1.2 *The space X^s is a Banach space.*

Theorem 1.3 *Let X be a Banach space with a Schauder basis Ψ. Then the operators $S_m : X \to X$ are bounded linear operators and*

$$\sup_m \|S_m\| < \infty.$$

The proof of this theorem is based on the fundamental theorem of Banach.

Theorem 1.4 *Let U, V be Banach spaces and T be a bounded linear one-to-one operator from V to U. Then the inverse operator T^{-1} is a bounded linear operator from U to V.*

We specify $U = X$ and $V = X^s$, and let T be the identity map. It follows from (1.13) that T is a bounded operator from V to U. Thus, by Theorem 1.4, T^{-1} is also bounded. This means that there exists a constant C such that, for any $f \in X$, we have $|||f||| \le C\|f\|$. This completes the proof of Theorem 1.3.

The operators $\{S_m\}_{m=1}^{\infty}$ are called the natural projections associated with a basis Ψ. The number $\sup_m \|S_m\|$ is called the basis constant of the basis Ψ. A basis whose basis constant is unity is called a *monotone basis*. It is clear that an orthonormal basis in a Hilbert space is a monotone basis. Every Schauder basis Ψ is monotone with respect to the norm $|||f||| := \sup_m \|S_m(f, \Psi)\|$, which was used above. Indeed, we have

$$|||S_m(f)||| = \sup_n \|S_n(S_m(f))\| = \sup_{1\le n\le m} \|S_n(f)\| \le |||f|||.$$

The above remark means that, for any Schauder basis Ψ of X, we can renorm X (take X^s) to make the basis Ψ monotone for a new norm.

Theorem 1.5 *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of elements in a Banach space X. Then $\{x_n\}_{n=1}^{\infty}$ is a Schauder basis of X if and only if the following three conditions hold:*

(a) $x_n \neq 0$ *for all* n;

(b) *there is a constant* K *such that, for every choice of scalars* $\{a_i\}_{i=1}^{\infty}$ *and integers* $n < m$, *we have*

$$\|\sum_{i=1}^{n} a_i x_i\| \leq K \|\sum_{i=1}^{m} a_i x_i\|;$$

(c) *the closed linear span of* $\{x_n\}_{n=1}^{\infty}$ *coincides with* X.

We note that for a basis Ψ with the basis constant K, we have, for any $f \in X$,

$$\|f - S_m(f, \Psi)\| \leq (K+1) \inf_{\{c_k\}} \|f - \sum_{k=1}^{m} c_k \psi_k\|.$$

Thus, the partial sums $S_m(f, \Psi)$ provide near-best approximation from $\operatorname{span}\{\psi_1, \dots, \psi_m\}$.

Let a Banach space X, with a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$, be given. In order to understand the efficiency of an algorithm providing an m-term approximation, we compare its accuracy with the best-possible accuracy when an approximant is a linear combination of m terms from Ψ. We define the best m-term approximation with regard to Ψ as follows:

$$\sigma_m(f) := \sigma_m(f, \Psi)_X := \inf_{c_k, \Lambda} \|f - \sum_{k \in \Lambda} c_k \psi_k\|_X,$$

where the infimum is taken over coefficients c_k and sets of indices Λ with cardinality $|\Lambda| = m$. We note that in the above definition of $\sigma_m(f, \Psi)_X$ the system Ψ may be any system of elements from X, not necessarily a basis of X.

An immediate natural question is: When does the best m-term approximant exist? This question is more difficult than the corresponding question in linear approximation and it has not been studied thoroughly. In what follows, we present some results that may point us in the right direction.

Let us proceed directly to the setting of our approximation problem. Let a subset $A \subset X$ be given. For any $f \in X$, let

$$d(f, A) := d(f, A)_X := \inf_{a \in A} \|f - a\|$$

denote the distance from f to A, or, in other words, the best approximation error of f by elements from A in the norm of X. To illustrate some appropriate techniques, we prove existence theorems in two settings.

S1 Let $X = L_p(0, 2\pi)$, $1 \leq p < \infty$, or $X = L_\infty(0, 2\pi) := \mathcal{C}(0, 2\pi)$ be the set of 2π-periodic functions. Consider A to be the set Σ_m of all complex trigonometric polynomials or $\Sigma_m(R)$ of all real trigonometric polynomials which have at most m nonzero coefficients:

$$\Sigma_m := \left\{ t : t = \sum_{k\in\Lambda} c_k e^{ikx}, \quad |\Lambda| \le m \right\},$$

$$\Sigma_m(R) := \left\{ t : t = \sum_{k\in\Lambda_1} a_k \cos kx + \sum_{k\in\Lambda_2} b_k \sin kx, \quad |\Lambda_1| + |\Lambda_2| \le m \right\}.$$

We will also use the following notation in this case:

$$\sigma_m(f, \mathcal{T})_X := d(f, \Sigma_m)_X.$$

S2 Let $X = L_p(0,1)$, $1 \le p < \infty$, and let A be the set Σ_m^S of piecewise constant functions with at most $m-1$ break-points at $(0,1)$.

In the setting **S2** we prove here the following existence theorem (see DeVore and Lorenz (1993), p. 363).

Theorem 1.6 *For any $f \in L_p(0,1)$, $1 \le p < \infty$, there exists $g \in \Sigma_m^S$ such that*

$$d(f, \Sigma_m^S)_p = \|f - g\|_p.$$

Proof Fix the break-points $0 = y_0 \le y_1 \le \cdots \le y_{m-1} \le y_m = 1$, let $y := (y_0, \dots, y_m)$, and let $S_0(y)$ be the set of piecewise constant functions with break-points $y_1, \dots, y_{m-1}$. Further, let

$$e_m^y(f)_p := \inf_{a\in S_0(y)} \|f - a\|_p.$$

From the definition of $d(f, \Sigma_m^S)_p$, there exists a sequence y^i such that

$$e_m^{y^i}(f)_p \quad \to \quad d(f, \Sigma_m^S)_p$$

when $i \to \infty$. Considering a subsequence of $\{y^i\}$, if necessary we can assume that $y^i \to y^*$ for some $y^* \in \mathbb{R}^{m+1}$. Now we consider only those indices j for which $y_{j-1}^* \ne y_j^*$. Let Λ denote the corresponding set of indices. Take a positive number ϵ satisfying

$$\epsilon < \min_{j\in\Lambda}(y_j^* - y_{j-1}^*)/3,$$

and consider i such that

$$\|y^* - y^i\|_\infty < \epsilon, \quad \text{where} \quad \|y\|_\infty := \max_k |y_k|. \tag{1.14}$$

By the existence theorem in the case of approximation by elements of a subspace of finite dimension, for each y^i there exists

$$g(f, y^i, c^i) := \sum_{j=1}^{m} c^i_j \chi_{[y^i_{j-1}, y^i_j]},$$

where χ_E denotes the characteristic function of a set E, with the property

$$\|f - g(f, y^i, c^i)\|_p = e^{y^i}_m(f)_p.$$

For i satisfying (1.14) and $j \in \Lambda$ we have $|c^i_j| \le C(f, \epsilon)$, which allows us to assume (passing to a subsequence if necessary) the convergence

$$\lim_{i\to\infty} c^i_j = c_j, \quad j \in \Lambda.$$

Consider

$$g(f, c) := \sum_{j\in\Lambda} c_j \chi_{[y^*_{j-1}, y^*_j]}.$$

Let $U_\epsilon(y^*) := \cup_{j\in\Lambda}(y^*_j - \epsilon, y^*_j + \epsilon)$ and introduce $G := [0, 1] \setminus U_\epsilon(y^*)$. Then we have

$$\int_G |f - g(f, c)|^p = \lim_{i\to\infty} \int_G |f - g(f, y^i, c^i)|^p \le d(f, \Sigma^S_m)^p_p.$$

Making $\epsilon \to 0$, we complete the proof. $\square$

We proceed now to the trigonometric case **S1**. We will give the proof in the general d-variable case for $\mathcal{T}^d := \mathcal{T} \times \cdots \times \mathcal{T}$ (d times) because this generality does not introduce any complications. The following theorem was essentially proved in Baishanski (1983). The presented proof is taken from Temlyakov (1998c).

Theorem 1.7 *Let* $1 \le p \le \infty$. *For any* $f \in L_p(\mathbb{T}^d)$ *and any* $m \in \mathbb{N}$, *there exists a trigonometric polynomial* t_m *of the form*

$$t_m(x) = \sum_{n=1}^{m} c_n e^{i(k^n, x)} \tag{1.15}$$

such that

$$\sigma_m(f, \mathcal{T}^d)_p = \|f - t_m\|_p. \tag{1.16}$$

Proof We prove this theorem by induction. Let us use the abbreviated notation $\sigma_m(f)_p := \sigma_m(f, \mathcal{T}^d)_p$.

First step Let $m = 1$. We assume $\sigma_1(f)_p < \|f\|_p$, because in the case $\sigma_1(f)_p = \|f\|_p$ the proof is trivial: we take $t_1 = 0$. We now prove that