Where Does Smoothness Count the Most For Fredholm Equations of the Second Kind With Noisy Information?

Arthur G. Werschulz*
Department of Computer and Information Sciences
Fordham University, New York, NY 10023
Department of Computer Science
Columbia University, New York, NY 10027

Technical Report CUCS-017-01

November 5, 2002

Abstract

We study the complexity of Fredholm problems $(I - T_k)u = f$ of the second kind on the $I^d = [0, 1]^d$, where T_k is an integral operator with kernel k. Previous work on the complexity of this problem has assumed either that we had complete information about k or that k and f had the same smoothness. In addition, most of this work has assumed that the information about k and f was exact. In this paper, we assume that k and f have different smoothness; more precisely, we assume that $f \in W^{r,p}(I^d)$ with $r > d/p$ and that $k \in W^{s,\infty}(I^{2d})$ with $s > 0$. In addition, we assume that our information about k and f is contaminated by noise. We find that the nth minimal error is $\Theta(n^{-\mu} + \delta)$, where $\mu = \min\{r/d, s/(2d)\}$ and δ is a bound on the noise. We prove that a noisy modified finite element method has nearly minimal error. This algorithm can be efficiently implemented using multigrid techniques. We thus find tight bounds on the ε-complexity for this problem. These bounds depend on the cost $c(\delta)$ of calculating a δ-noisy information value. As an example, if the cost of a δ-noisy evaluation is proportional to δ^{-t}, then the ε-complexity is roughly $(1/\varepsilon)^{t+1/\mu}$.

*This research was supported in part by the National Science Foundation under Grant CCR-99-87858, as well as by a Fordham Univeristy Faculty Fellowship.

1 Introduction

We are interested in the worst case complexity of solving Fredholm problems of the second kind

$$(I - T_k)u = f \tag{1}$$

on the unit cube $I^d = [0, 1]^d$, where

$$T_k v = \int_{I^d} k(\cdot, y)v(y)\,dy \qquad \forall\, v \in L_p(I^d)$$

for a continuous kernel function $k\colon I^d \times I^d \to \mathbb{R}$. Here, $p \in [1, \infty]$, and error is measured in the $L_p(I^d)$-norm.

Previous work on this problem has either assumed that we have had complete information about k, or that k and f have had the same smoothness, see, e.g., [5], [6], [8], [10], [14], [15, Sec. 6.3], and the references contained therein.

What happens when we weaken these assumptions? There are two issues to deal with. First, we want to know where smoothness counts the most for Fredholm problems, as we did in [16] for two-point boundary value problems. That is, we would like to know which is more important—the smoothness of the kernel or of the right-hand side—in determining the complexity. In addition, note that (with the exception of [8]) the references listed above have all assumed that the available information is exact. But in practice, information evaluations are often contaminated by noise [11]. Hence we wish to know how noisy information affects the complexity, as well as which algorithms are optimal when the information is noisy.

In this paper, we study the worst case complexity of Fredholm problems under the following assumptions:

1. The right-hand side f belongs to the unit ball of $W^{r,p}(I^d)$, with $r > d/p$.
2. The kernel k belongs to a ball of $W^{s,\infty}(I^{2d})$, and $I - T_k$ is an invertible operator on $L_p(I^d)$.
3. Only noisy standard information is available. That is, for any $x, y \in I^d$, we can only calculate $f(x)$ or $k(x, y)$ with error at most δ, where $\delta \in [0, 1]$ is a known noise level.

We are able to determine $r_n(\delta)$, the nth minimal radius of δ-noisy information, i.e., the minimal error when we use n evaluations with a noise level of δ. We find that[1]

$$r_n(\delta) \asymp n^{-\mu} + \delta$$

[1] In this paper, we use $\preccurlyeq$, $\succcurlyeq$, and $\asymp$ to denote O-, Ω-, and Θ-relations.

with a proportionality factor independent of n and δ, where

$$\mu = \min\left\{\frac{r}{d}, \frac{s}{2d}\right\}. \tag{2}$$

Moreover, we describe an algorithm using n evaluations with noise level δ that is a nearly-minimal error algorithm. This algorithm is a modified finite element method (MFEM) using noisy information. The modification consists of replacing the kernel k and the right-hand side f that would appear in the "pure" finite element method by their piecewise-polynomial interpolants. Hence this algorithm uses noisy standard information, rather than continuous linear information. We shall refer to this algorithm as the "noisy MFEM." This is, of course, a bit of a misnomer, since the algorithm isn't noisy (only the information is noisy); but "noisy MFEM" is more succinct than "MFEM using noisy information."

We also analyze the cost of the noisy MFEM. Let $c(\delta)$ denote the cost of evaluating a function with a noise level δ. Then the information cost of this algorithm is $c(\delta)\,n$.

Let us now discuss the combinatory cost of the noisy MFEM. This algorithm requires the solution of an $n \times n$ linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$. Here, $\mathbf{A}$ is the Gram matrix of the finite element space, $\mathbf{B}$ depends on the kernel k and $\mathbf{f}$ depends on the right-hand side f. If we were considering only a single fixed kernel k, then we could precompute the LU-decomposition of the nonsingular matrix $\mathbf{A} - \mathbf{B}$, since this is independent of any particular f. We could then ignore the cost of this precomputation, considering it as a fixed overhead, since it need only be done once. Even so, the combinatory cost of our algorithm would be $\Theta(n^2)$, since the factors of the LU-decomposition of $\mathbf{A} - \mathbf{B}$ are dense $n \times n$ triangular matrices. Of course, things are much worse for our problem, since both the right-hand sides f and the kernels k are varying. Clearly, the factorization of $\mathbf{A} - \mathbf{B}$ is no longer independent of the problem element being considered, and so we would not be able to ignore the $O(n^3)$-cost of this factorization. Hence, we see that the combinatory cost of the noisy MFEM would overwhelm the information cost as n grows large.

We can overcome this difficulty by using a two-grid implementation of the noisy MFEM. This algorithm has the same order of error as the original noisy MFEM, and its combinatory cost is $O(n)$. Hence, we can calculate the two-grid approximation using $\Theta(n)$ arithmetic operations, which is optimal.

Using these results, we can determine tight bounds on the ε-complexity of the Fredholm problem. There exist positive constants C_1, C_2, and C_3, independent of ε, such that the problem complexity is bounded from below by

$$\operatorname{comp}(\varepsilon) \geq \inf_{0<\delta<C_1\varepsilon} \left\{ c(\delta) \left\lceil \left(\frac{1}{C_1\varepsilon - \delta} \right)^{1/\mu} \right\rceil \right\}$$

and from above by

$$\mathrm{comp}(\varepsilon) \le C_2 \inf_{0<\delta<C_3\varepsilon} \left\{ c(\delta) \left\lceil \left(\frac{1}{C_3\varepsilon - \delta} \right)^{1/\mu} \right\rceil \right\}.$$

These upper bounds are attained by two-grid implementations of the noisy modified FEM, with δ chosen to minimize the right-hand sides of the upper bound.

As a specific example, suppose that $c(\delta) = \delta^{-t}$ for some $t \ge 0$. We find that

$$\mathrm{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon} \right)^{t+1/\mu}.$$

Thus we have found sharp bounds on the ε-complexity.

How much do we lose when we go from exact information to noisy information? Suppose once again that $c(\delta) = \delta^{-t}$ for some $t \ge 0$. Since exact information is merely noisy information with $t = 0$, we see that the complexity for exact information is proportional to $c(1/\varepsilon)^{1/\mu}$, where c is the cost of one function evaluation. For the sake of comparison, let us write the complexity for noisy information as $(1/\varepsilon)^{1/\mu'}$, where

$$\mu' = \mu \cdot \frac{1}{1 + t\mu}.$$

Note that since the information is noisy, we have $t > 0$, and so $\mu' < \mu$. Hence we see that the complexity of our problem using noisy information of smoothness (r, s) is the same as the complexity using exact information of lesser smoothness (r', s'), where $r' = r/(1 + t\mu)$ and $s' = s/(1 + t\mu)$.

We now outline the rest of this paper. In Section 2, we precisely describe the problem to be solved. In Section 3, we prove a lower bound on the minimal error using noisy information. It is easy to find a matching upper bound using the general approach of interpolatory algorithms. However, this approach does not address the issue of combinatory cost. Since the problem is nonlinear, it is unclear whether there exists an interpolatory algorithm with (roughly) linear combinatory cost. The remainder of this paper deals with showing that such an algorithm exists, and is given as a two-grid implementation of a noisy modified finite element method (noisy MFEM). In Section 4, we define some useful finite element spaces, which are used in Section 5 to define the noisy MFEM. In Section 6, we establish an error bound for the noisy MFEM. In Section 7, we show that the noisy MFEM is a minimal error algorithm. In Section 8, we describe the two-grid implementation of the noisy MFEM, showing that its error is essentially the same as the noisy MFEM itself, and that its combinatory cost is essentially optimal. Finally, in Section 9, we determine the ε-complexity of the noisy Fredholm problem.

2 Problem description

In this section, we precisely describe the class of Fredholm problems whose solutions we wish to approximate.

For an ordered ring $\mathscr{R}$, we shall let $\mathscr{R}^+$ and $\mathscr{R}^{++}$ respectively denote the non-negative and positive elements of $\mathscr{R}$. Hence (for example), $\mathbb{Z}^+$ denotes the set of natural numbers (non-negative integers), whereas $\mathbb{Z}^{++}$ denotes the set of strictly positive integers. For a normed linear space $\mathscr{X}$, we let $\mathscr{B}\mathscr{X}$ denote the unit ball of $\mathscr{X}$. We assume that the reader is familiar with the standard concepts and notations involving Sobolev norms and spaces, as found in, e.g., [3].

We are given $d \in \mathbb{Z}^{++}$ and $p \in [1, \infty]$, as well as real numbers r and s satisfying $r > d/p$ and $s > 0$. Hence, the Sobolev space $W^{r,p}(I^d)$ is embedded in the space $C(I^d)$ of continuous functions, and $W^{s,\infty}(I^{2d})$ is embedded in $C(I^{2d})$, by the Sobolev embedding theorem.

For $k \in W^{s,\infty}(I^{2d})$, define $T_k \colon L_p(I^d) \to L_p(I^d)$ as

$$(T_k v)(x) = \int_{I^d} k(x, y) v(y)\, dy \qquad \forall\, x \in I^d.$$

The operator T_k is compact, see, e.g., [4, pg. 518], and hence $I - T_k$ is an invertible operator on $L_p(I^d)$ iff 1 is not an eigenvalue of T_k.

We are now ready to describe our class of problem elements. We first describe the class of kernels k. Let $c_1 > 0$ and $c_2 > 1$ be given. Then we let $\mathscr{K} = \mathscr{K}_{c_1,c_2}$ denote the class of all functions $k \in W^{s,\infty}(I^{2d})$ such that

$$\|k\|_{W^{s,\infty}(I^{2d})} \le c_1$$

and

$$\|(I - T_k)^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \le c_2.$$

Here, $\|\cdot\|_{\mathrm{Lin}[\mathscr{X}]}$ is the usual operator norm. The class of right-hand sides will be $\mathscr{B}W^{r,p}(I^d)$. Finally, we let

$$F = \mathscr{B}W^{r,p}(I^d) \times \mathscr{K}.$$

be our class of problem elements.

We are now ready to define our solution operator $S \colon F \to L_p(I^d)$ as

$$S([f, k]) = (I - T_k)^{-1} f \qquad \forall\, [f, k] \in F.$$

Hence $u = S([f, k])$ is the solution of (1) for $[f, k] \in F$.

We wish to calculate approximate solutions to this problem, using noisy standard information. To be specific, we will be using uniformly sup-norm-bounded noise. Our notation and terminology is essentially that of [11], although we sometimes use modifications found in [12].

Let $\delta \in [0, 1]$ be a *noise level*. For $[f, k] \in F$, we calculate δ-*noisy information*

$$z = [z_1, \dots, z_{n(z)}]$$

about $[f, k]$. Here, for each index $i \in \{1, \dots, n(z)\}$, either

$$|z_i - f(x_i)| \le \delta \text{ for } x_i \in I^d,$$

or

$$|z_i - k(x_i, y_i)| \le \delta \text{ for } (x_i, y_i) \in I^{2d}.$$

The choice of whether to evaluate k or f at the ith sample point, as well as the choice of the ith sample point itself, may be determined either nonadaptively or adaptively. Moreover, the information is allowed to be of varying cardinality.

For $[f, k] \in F$, we let $\mathbb{N}_\delta([f, k])$ denote the set of all such δ-noisy information z about $[f, k]$, and we let

$$Z = \bigcup_{[f,k]\in F} \mathbb{N}_\delta([f, k])$$

denote the set of all possible noisy information values. Then an *algorithm* using the noisy information $\mathbb{N}_\delta$ is a mapping $\phi\colon Z \to L_p(I^d)$.

Remark. Note that the permissible information consists of function values of f and k. One could allow the evaluation of derivatives as well. We restrict ourselves to function values alone, since this simplifies the exposition. There is no loss of generality in doing this, since the results of this paper also hold if derivative evaluations are allowed. $\square$

We want to solve the Fredholm problem in the worst case setting. This means that the *cardinality* of information $\mathbb{N}_\delta$ is given as

$$\operatorname{card} \mathbb{N}_\delta = \sup_{z\in Z} n(z)$$

and the *error* of an algorithm ϕ using $\mathbb{N}_\delta$ is given as

$$e(\phi, \mathbb{N}_\delta) = \sup_{[f,k]\in F} \; \sup_{z\in\mathbb{N}_\delta([f,k])} \|S([f, k]) - \phi(z)\|_{L_p(I^d)}.$$

As usual, we will need to know the minimal error achievable by algorithms using specific information, as well as by algorithms using information of specified cardinality. Let $n \in \mathbb{Z}^+$ and $\delta \in [0, 1]$. If $\mathbb{N}_\delta$ is δ-noisy information of cardinality at most n, then

$$r(\mathbb{N}_\delta) = \inf_{\phi \text{ using } \mathbb{N}_\delta} e(\phi, \mathbb{N}_\delta).$$

is the *radius of information*, i.e., the minimal error among all algorithms using given information $\mathbb{N}_\delta$. An algorithm ϕ^* using $\mathbb{N}_\delta$ is said to be an *optimal error algorithm*[2] if

$$e(\phi^*, \mathbb{N}_\delta) \asymp r(\mathbb{N}_\delta),$$

the proportionality constant being independent of n and δ. The nth *minimal radius*

$$r_n(\delta) = \inf\{\, r(\mathbb{N}_\delta) : \operatorname{card} \mathbb{N}_\delta \leq n \,\},$$

is the minimal error among all algorithms using δ-noisy information of cardinality at most n. Noisy information $\mathbb{N}_{n,\delta}$ of cardinality n such that

$$r(\mathbb{N}_{n,\delta}) \asymp r_n(\delta),$$

the proportionality factor being independent of both n and δ, is said to be nth *optimal information*. An optimal error algorithm using nth optimal information is said to be an nth *minimal error algorithm*.

Next, we describe our model of computation. We will use the model found in [11, Section 2.9]. (However, note that in the present paper, the accuracy δ is the same for all noisy observations, whereas δ may differ from one observation to another in [11].) Here are the most important features of this model:

1. For any $x \in I^d$ and any $f \in W^{r,p}(I^d)$, the cost of calculating a δ-noisy value of $f(x)$ is $c(\delta)$.

2. For any $(x, y) \in I^{2d}$ and any $k \in \mathscr{K}$, the cost of calculating a δ-noisy value of $k(x, y)$ is $c(\delta)$.

3. Real arithmetic operations and comparisons are done exactly, with unit cost.

Here, the cost function $c \colon \mathbb{R}^+ \to \mathbb{R}^{++}$ is nonincreasing.

For any noisy information $\mathbb{N}_\delta$ and any algorithm ϕ using $\mathbb{N}_\delta$, we shall let $\operatorname{cost}(\phi, \mathbb{N}_\delta)$ denote the worst case cost of computing $\phi(z)(x)$ for $z \in Z$ and $x \in I^d$. We can decompose this as follows. Let

$$\operatorname{cost}^{\text{info}}(\mathbb{N}_\delta) = \sup_{z \in Z} \{\text{cost of computing } z\}$$

[2] In this paper, we ignore constant multiplicative factors in our definitions of optimality. The more fastidious may use the term “quasi-optimal” if they desire.

denote the worst case *information cost.* Note that if $\mathbb{N}_\delta$ is information of cardinality n, then

$$\text{cost}^{\text{info}}(\mathbb{N}_\delta) \geq c(\delta)\, n.$$

Here, equality holds for nonadaptive information, but strict inequality can hold for adaptive information, since we must be concerned with the cost of choosing each new adaptive sample point. We also let

$$\text{cost}^{\text{comb}}(\phi, \mathbb{N}_\delta) = \sup_{z\in Z} \sup_{x\in I^d} \{\text{cost of computing } \phi(z)(x), \text{ given } z \in Z\}$$

denote the worst case *combinatory cost.* Then

$$\text{cost}(\phi, \mathbb{N}_\delta) \leq \text{cost}^{\text{info}}(\mathbb{N}_\delta) + \text{cost}^{\text{comb}}(\phi, \mathbb{N}_\delta).$$

Now that we have defined the error and cost of an algorithm, we can finally define the complexity of our problem. We shall say that

$$\text{comp}(\varepsilon) = \inf\{\, \text{cost}(\phi, \mathbb{N}_\delta) : \mathbb{N}_\delta \text{ and } \phi \text{ such that } e(\phi, \mathbb{N}_\delta) \leq \varepsilon \,\}$$

is the *ε-complexity* of our problem. An algorithm ϕ using noisy information $\mathbb{N}_\delta$ for which

$$e(\phi, \mathbb{N}_\delta) \leq \varepsilon \qquad \text{and} \qquad \text{cost}(\phi, \mathbb{N}_\delta) \asymp \text{comp}(\varepsilon),$$

the proportionality factor being independent of both δ and ε, is said to be an *optimal algorithm.*

3 Lower bounds

In this section, we prove a lower bound on the nth minimal error using δ-noisy information.

Theorem 3.1. *Recall from* (2) *that*

$$\mu = \min\left\{\frac{r}{d}, \frac{s}{2d}\right\}.$$

There is a constant M_0, independent of n and δ, such that

$$r_n(\delta) \geq M_0(n^{-\mu} + \delta)$$

for all $n \in \mathbb{Z}^+$ and $\delta \in [0, 1]$.

Proof. We first claim that

$$r_n(\delta) \succcurlyeq n^{-r/d} + \delta. \tag{3}$$

Indeed, since $T_0 = 0$, we find that $S([f, 0]) = f$ for all $f \in W^{r,p}(I^d)$. Thus APP, the problem of approximating functions from $\mathscr{B}W^{r,p}(I^d)$ in the $L_p(I^d)$-norm, is a special instance of our problem, and so

$$r_n(\delta) \geq r_n(\delta; \text{APP}),$$

the latter denoting the nth minimal radius of δ-noisy information for APP. Clearly

$$r_n(\delta; \text{APP}) \geq r_n(0; \text{APP}). \tag{4}$$

Moreover,

$$r_n(0; \text{APP}) \succcurlyeq n^{-r/d},$$

see, e.g., [9, pg. 34]. Hence

$$r_n(\delta; \text{APP}) \succcurlyeq n^{-r/d}. \tag{5}$$

Thus, to establish (3), we only need to prove that

$$r_n(\delta; \text{APP}) \succcurlyeq \delta. \tag{6}$$

Let $\mathbb{N}_\delta$ be noisy information of cardinality at most n. By the results in [11, Chapter 2.7], there exists nonadpative information $\mathbb{N}_\delta^{\text{non}}$ of cardinality l' such that

$$r(\mathbb{N}_\delta; \text{APP}) \geq \tfrac{1}{2} r(\mathbb{N}_\delta^{\text{non}}; \text{APP}).$$

By [11, Lemma 2.8.2],

$$r(\mathbb{N}_\delta^{\text{non}}; \text{APP}) \succcurlyeq \delta.$$

Hence

$$r(\mathbb{N}_\delta; \text{APP}) \succcurlyeq \delta.$$

Since $\mathbb{N}_\delta$ is arbitrary information of cardinality at most n, we find that (6) holds. Using (4)–(6), we find that (3) holds, as claimed.

We now claim that

$$r_n(0) \succcurlyeq n^{-s/2d} \tag{7}$$

holds. Our approach follows that outlined in [5, pp. 260–261].

Let

$$\theta_1 \in (c_2^{-1}, 1) \qquad \text{and} \qquad k_0 = \min\left\{\theta_1 c_1, 1 - \frac{1}{\theta_1 c_2}\right\},$$

and define

$$f^* \equiv 1 \qquad \text{and} \qquad k^* \equiv k_0.$$

Now

$$\|k^*\|_{W^{s,\infty}(I^{2d})} = k_0 < c_1. \tag{8}$$

It is easy to see that

$$\|T_k\|_{\mathrm{Lin}[L_p(I^d)]} \le \|k\|_{C(I^{2d})} \qquad \forall\, k \in \mathscr{K}. \tag{9}$$

In particular, we have

$$\|T_{k^*}\|_{\mathrm{Lin}[L_p(I^d)]} \le k_0 < 1,$$

so that

$$\|(I - T_{k^*})^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \le \frac{1}{1-k_0} \le \theta_1 c_2 < c_2. \tag{10}$$

From (8) and (10), we see that $k^* \in \mathscr{K}$. Since it is clear that $f^* \in \mathscr{B}W^{r,p}(I^d)$, we find that $[f^*, k^*] \in F$.

Let N be noiseless information of cardinality at most n. Then we may write

$$N([f^*, k^*]) = [z_1, \ldots, z_l]$$

for some $l \le n$, where each z_i is an evaluation of either f^* or k^*. Suppose that there are l' evaluations of k^*. Without loss of generality, we may assume that that these evaluations have the form

$$z_i = k^*(x_i, y_i) \qquad (1 \le i \le l').$$

From [2] (see also [9, pg. 34]), we can find a function $w \in \mathscr{B}W^{s,\infty}(I^{2d})$ such that

$$\begin{gathered} 0 \le w(x,y) \le k_0 \qquad \forall\, x, y \in I^d, \\ w(x_i, y_i) = 0 \qquad (1 \le i \le l'), \\ \|w\|_{W^{s,\infty}(I^{2d})} = 1, \\ \int_{I^{2d}} w(x,y)\,dx\,dy \ge \frac{\theta_2}{(l')^{s/2d}}, \end{gathered}$$

where θ_2 is a positive constant that is independent of the points (x_i, y_i) and of l'.

Let

$$\theta_3 = \min\{(1-\theta_1)c_1, 1 - c_2^{-1} - k_0\}.$$

Note that since $\theta_1 < 1$ and $k_0 \le 1 - (\theta_1 c_2)^{-1}$, we have $k_0 < 1 - c_2^{-1}$, and so $\theta_3 > 0$. We define

$$k^{**} = k_0 + \theta_3 w.$$

We claim that $k^{**} \in \mathscr{K}$. Indeed, we have

$$\begin{aligned}\|k^{**}\|_{W^{s,\infty}(I^{2d})} &\le \|k_0\|_{W^{s,\infty}(I^{2d})} + \theta_3 \|w\|_{W^{s,\infty}(I^{2d})} = k_0 + \theta_3 \\ &\le \theta_1 c_1 + \theta_3 \le c_1.\end{aligned}$$

Moreover,

$$\|T_{k^{**}}\|_{\mathrm{Lin}[L_p(I^d)]} \le \|T_{k_0}\|_{\mathrm{Lin}[L_p(I^d)]} + \theta_3 \|T_w\|_{\mathrm{Lin}[L_p(I^d)]} \le k_0 + \theta_3 < 1, \qquad (11)$$

and thus

$$\|(I - T_{k^{**}})^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \le \frac{1}{1 - (k_0 + \theta_3)} \le c_2.$$

Hence, $k^{**} \in \mathscr{K}$.

Letting $f^* \equiv 1$, we let

$$u^* = S([f^*, k^*]) \qquad \text{and} \qquad u^{**} = S([f^*, k^{**}]).$$

Since

$$[f^*, k^*], [f^*, k^{**}] \in F \qquad \text{with} \qquad N([f^*, k^*]) = N([f^*, k^{**}]),$$

we have

$$r(N) \ge \tfrac{1}{2} \|u^* - u^{**}\|_{L_p(I^d)}, \qquad (12)$$

see, e.g., [13, pp. 45, 49].

We claim that $u^{**} > 1$ on I^d. Indeed, since (11) holds, the Neumann series

$$(I - T_{k^{**}})^{-1} = \sum_{j=0}^{\infty} T_{k^{**}}^j$$

converges in $\mathrm{Lin}[L_p(I^d)]$. Now

$$T_{k^{**}}^j = T_{k_j^{**}} \qquad \text{for } j \ge 1,$$

where $\{k_j^{**}\}_{j=1}^{\infty}$ is defined inductively as

$$k_j^{**}(x, y) = \begin{cases} k^{**}(x, y) & \text{if } j = 1, \\ \displaystyle\int_{I^d} k^{**}(x, t)\, k_{j-1}^{**}(t, y)\, dt & \text{if } j \ge 2 \end{cases} \qquad \forall\, x, y \in I^d.$$

Hence

$$u^{**} = \sum_{j=0}^{\infty} T_{k^{**}}^j f^*.$$

By induction, we find that

$$k_j^{**}(x, y) \geq k_0^j \qquad \forall\, x, y \in I^d, \forall j \geq 1,$$

and thus for $x \in I^d$, we have

$$u^{**}(x) = 1 + \sum_{j=1}^{\infty} \int_{I^d} k_j^{**}(x, y)\, dy \geq 1 + \sum_{j=1}^{\infty} k_0^j = \frac{1}{1 - k_0} > 1,$$

as claimed.

Hence

$$u^{**}(x) - u^*(x) = k_0 \int_{I^d} [u^{**}(y) - u^*(y)]\, dy + \theta_3 \int_{I^d} w(x, y) u^{**}(y)\, dy.$$

Since $u^{**} > 1$ on I^d and $w > 0$ on I^{2d}, we find that

$$\begin{aligned}(1 - k_0) \int_{I^d} [u^{**}(x) - u^*(x)]\, dx &= \theta_3 \int_{I^{2d}} w(x, y) u^{**}(y)\, dy\, dx \\ &> \theta_3 \int_{I^{2d}} w(x, y)\, dy\, dx \\ &\geq \frac{\theta_2 \theta_3}{(l')^{s/2d}} \geq \frac{\theta_2 \theta_3}{n^{s/2d}},\end{aligned}$$

the latter since $l' \leq n$. By Minkowski's inequality, we have

$$(1 - k_0) \int_{I^d} [u^{**}(x) - u^*(x)]\, dx \leq (1 - k_0) \| u^{**} - u^* \|_{L_p(I^d)}.$$

Using the last two inequalities and (12), we get

$$r(N) \geq \frac{\theta_2 \theta_3}{2(1 - k_0) n^{s/2d}}.$$

Since N is arbitrary information of cardinality at most n, the inequality (7) holds, as claimed.

From (3), we see that

$$r_n(\delta) \succcurlyeq \delta,$$

which, together with (7), implies that

$$r_n(\delta) \succcurlyeq n^{-s/2d} + \delta.$$

The theorem now follows immediately from this inequality and (3). $\square$

4 Some finite element spaces

Now that we have a lower bound on the nth minimal radius for our problem, the next task will be to find a matching upper bond and an nth minimal error algorithm. This algorithm will be a modified finite element method using noisy information. Before describing the algorithm, we need to define some finite element spaces.

In what follows, our notation is based on the standard one found in, e.g., [3] and [15, Chapter 5].

Let $m \in \mathbb{Z}^+$. For $K \subseteq \mathbb{R}^d$, let

$$Q_m(K) = \left\{ \sum_{0 \le \alpha_1, \ldots, \alpha_d \le m} a_\alpha x^\alpha : x \in K \right\}$$

denote the polynomials of degree at most m in each variable, with the domain restricted to K. Here, we recall that $x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$ for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$. Clearly $Q_m(K)$ is a function space over K, with

$$\dim Q_m(K) = (m+1)^d.$$

In particular, we note that the space $Q_m(I^d)$ has a basis $\{\hat{s}_1, \ldots, \hat{s}_a\}$ consisting of tensor products. More precisely, let

$$\hat{p}_i(\hat{\xi}) = \prod_{\substack{0 \le j \le m \\ j \ne i}} \frac{\hat{\xi} - \hat{\xi}_j}{\hat{\xi}_i - \hat{\xi}_j}, \qquad (0 \le i \le m)$$

be the usual one-dimensional Lagrange basis polynomials, where $0 < \hat{\xi}_1 < \cdots < \hat{\xi}_m < 1$. Let $\{\alpha^{(i)}\}_{i=1}^{(m+1)^d}$ be an enumeration of the multi-indices $\alpha \in (\mathbb{Z}^+)^d$ satisfying $\max_{1 \le j \le d} \alpha_j \le m$; we write $\alpha^{(i)} = (\alpha_1^{(i)}, \ldots, \alpha_d^{(i)})$. We can set

$$\hat{s}_i(\hat{\xi}_1, \ldots, \hat{\xi}_d) = \prod_{j=1}^{d} \hat{p}_{\alpha_j^{(i)}}(\hat{\xi}_j)$$

and

$$\hat{x}_i = \left(\hat{\xi}_1^{\alpha^{(i)}}, \ldots, \hat{\xi}_d^{\alpha^{(i)}} \right).$$

Then $\{\hat{s}_1, \ldots, \hat{s}_{(m+1)^d}\}$ is a basis for $Q_m(I^d)$ such that

$$\hat{s}_j(\hat{x}_i) = \delta_{i,j} \qquad \text{for } 1 \le i, j \le (m+1)^d.$$

Associated with the space $Q_m(I^d)$, we have an interpolation operator $\hat{\Pi}\colon C(I^d) \to Q_m(I^d)$ defined as

$$\hat{\Pi}\hat{v} = \sum_{i=1}^{(m+1)^d} \hat{v}(\hat{x}_i)\hat{s}_i \qquad \forall\, \hat{v} \in C(I^d).$$

Now let K be a cube in $\mathbb{R}^d$ whose sides are parallel to the coordinate axes. Then K can be written as the image of I^d under an affine bijection $F_K\colon I^d \to K$ having the form

$$F_K(\hat{x}) = h_K\hat{x} + b_K \qquad \forall\, \hat{x} \in I^d,$$

where h_K is the length of any side of K and b_K is the element in K closest to the origin, i.e., the smallest corner of K. We get a basis $\{s_{1,K}, \ldots, s_{(m+1)^d,K}\}$ for $Q_m(K)$ by taking

$$s_{j,K} = \hat{s}_j \circ F_K^{-1},$$

that is,

$$s_{j,K}(x) = \hat{s}_j(\hat{x}) \qquad \text{where } \hat{x} = F_K^{-1}(x) = \frac{x - b_K}{h_K}\,,$$

for $1 \le j \le (m+1)^d$. Defining

$$x_{j,K} = F_K(\hat{x}_j) \qquad \text{for } 1 \le j \le (m+1)^d,$$

we find that

$$s_{j,K}(x_{i,K}) = \delta_{i,j} \qquad \text{for } 1 \le i,\, j \le (m+1)^d.$$

Associated with the polynomial space $Q_m(K)$, we have an interpolation operator $\Pi_K\colon C(K) \to Q_m(K)$ defined as

$$\Pi_K v = \sum_{j=1}^{(m+1)^d} v(x_{j,K})s_{j,K} \qquad \forall\, v \in C(K),$$

so that

$$(\Pi_K v)(x) = (\hat{\Pi}\hat{v})(\hat{x}) \qquad \text{for } \hat{v} = v \circ F_K \text{ and } \hat{x} = F_K^{-1}(x).$$

We are finally ready to define finite element spaces. Choose $h > 0$ such that $1/h$ is an integer. Let $\mathcal{Q}_h$ be a decomposition of I^d into congruent cubes whose sides parallel the coordinate axes and have length h. Then

$$\mathscr{S}_h = \left\{ I^d \xrightarrow{v} \mathbb{R} : v\Big|_K \in Q_m(K) \text{ for } K \in \mathcal{Q}_h \right\}$$

is our finite element space. Note that since $|\mathscr{Q}_h| = h^{-d}$, we have

$$n_h := \dim \mathscr{S}_h = \left(\frac{m+1}{h}\right)^d. \tag{13}$$

We now construct a basis $\{s_1, \dots, s_{n_h}\}$ for $\mathscr{Q}_h$. Let $b_{K_1}, \dots, b_{K_{h^{-d}}}$ be an enumeration of the points $\{b_K\}_{K\in\mathscr{Q}_h}$ by lexicographic ordering. This induces an enumeration $K_1, \dots, K_{h^{-d}}$ of the cubes $K \in \mathscr{Q}_h$. We then let

$$s_{h^{-d}(i-1)+j} = s_{i,K_j} \qquad \text{for } 1 \le j \le h^{-d}, 1 \le i \le (m+1)^d,$$

with each $s_{i,K}$ being extended from K to I^d as being zero outside K. Analogously, we let

$$x_{h^{-d}(i-1)+j} = x_{i,K_j} \qquad \text{for } 1 \le j \le h^{-d}, 1 \le i \le (m+1)^d.$$

We then find that

$$s_j(x_i) = \delta_{i,j} \qquad \text{for } 1 \le i, j \le n_h.$$

Associated with the finite element space $\mathscr{S}_h$, we have an interpolation operator $\Pi_h \colon C(I^d) \to \mathscr{S}_h$, defined as

$$\Pi_h v = \sum_{K\in\mathscr{Q}_h} \Pi_K v \qquad \forall\, v \in C(I^d),$$

where each $\Pi_K v$ is extended from K to I^d as being zero outside K. Alternatively, we may write

$$\Pi_h v = \sum_{j=1}^{n_h} v(x_j) s_j \qquad \forall\, v \in C(I^d).$$

We have a second interpolation operator $\Pi_{h\otimes h} \colon C(I^{2d}) \to \mathscr{S}_h \otimes \mathscr{S}_h$, defined as

$$\begin{aligned}(\Pi_{h\otimes h} v)(x, y) &= \Pi_h\big[x \mapsto \Pi_h\big(y \mapsto v(x, y)\big)\big] \\ &= \sum_{i,j=1}^{n_h} v(x_i, x_j) s_j(y) s_i(x)\end{aligned}$$

for $x, y \in I^d$ and $v \in C(I^{2d})$.

Remark. In the sequel, we shall often write $s_{i,h}$ and $x_{j,h}$ rather than s_i and x_j, to indicate their dependence on h.

We now present some standard error estimates, which will be useful in the sequel.

Lemma 4.1. *Let $t \geq 0$ and $q \in [1, \infty]$. There exists $M_1 > 0$ such that the following hold:*

1. *Let $v \in W^{t,q}(I^d)$. Then*

$$\|v - \Pi_h v\|_{L_q(I^d)} \leq M_1 h^{\min\{m+1,t\}} \|v\|_{W^{t,q}(I^d)}.$$

2. *Let $w \in W^{t,q}(I^{2d})$. Then*

$$\|w - \Pi_{h\otimes h} w\|_{L_q(I^{2d})} \leq M_1 h^{\min\{m+1,t\}} \|w\|_{W^{t,q}(I^{2d})}.$$

Proof. For $K \subseteq \mathbb{R}^d$, let

$$P_m(K) = \left\{ \sum_{|\alpha| \leq m} a_\alpha x^\alpha : x \in K \right\}$$

denote the polynomials of total degree at most m. Since $P_m(I^d) \subseteq Q_m(I^d)$, we see that $\hat{\Pi}\hat{v} = \hat{v}$ for all $v \in P_m(I^d)$. Hence the local estimates of [3, pp. 118–122] hold. Since there are no inter-element continuity relations to deal with, the global estimates of [3] hold as well. This suffices to establish the lemma. $\square$

Let $h > 0$. Recall that the mapping $P_h \colon L_2(I^d) \to L_2(I^d)$, defined as

$$\langle P_h v, w \rangle = \langle v, w \rangle \qquad \forall\, v \in L_2(I^d), w \in \mathscr{S}_h, \tag{14}$$

is the *orthogonal projector* of $L_2(I^d)$ onto $\mathscr{S}_h$. Here, $\langle \cdot, \cdot \rangle$ is the standard duality pairing

$$\langle v, w \rangle = \int_{I^d} v(x) w(x)\, dx \qquad \forall\, v \in L_p(I^d), w \in L_{p'}(I^d),$$

with

$$p' = \frac{p}{p-1}$$

denoting the exponent conjugate to p. It is well-known that P_h is a self-adjoint operator with range $\mathscr{S}_h$ and unit norm. The next lemma shows that $\{P_h\}_{h>0}$ is uniformly bounded in the other $L_q(I^d)$-norms.

Lemma 4.2. *Let $q \in [1, \infty]$. There exists $\pi_q > 0$ such that for any $h > 0$,*

$$\|P_h v\|_{L_q(I^d)} \leq \pi_q \|v\|_{L_q(I^d)} \qquad \forall\, v \in L_q(I^d).$$

Proof. See, e.g., [15, pp. 177-178], and the references cited therein. $\square$

5 The noisy modified FEM

We now define the noisy modified finite element method (noisy MFEM). This is an algorithm using information consisting of noisy function evaluations. As mentioned in the Introduction, it would be somewhat more accurate to describe this method as the "MFEM using noisy information," but the conciseness of "noisy MFEM" outweighs its mild inaccuracy.

The easiest way to describe the noisy MFEM is by following three steps. First, we describe the pure finite element method, which uses inner product information. Next, we describe the noise-free MFEM, which uses noise-free standard information. Finally, we describe the noisy MFEM, which uses noisy standard information.

We first recall how the pure finite element method is defined. Let $[f, k] \in F$ and $h > 0$. Then the pure *finite element method* (pure FEM) consists of finding $u_h \in \mathscr{S}_h$ such that

$$B(u_h, w; k) = \langle f, w \rangle \qquad \forall\, w \in \mathscr{S}_h,$$

where

$$B(v, w; k) = \langle (I - T_k)v, w \rangle \qquad \forall\, v \in L_p(I^d), w \in L_{p'}(I^d).$$

Alternatively, we have

$$(I - P_h T_k)u_h = P_h f.$$

If we write

$$u_h(x) = \sum_{j=1}^{n_h} \upsilon_j s_{j,h}(x) \qquad \forall\, x \in I^d,$$

then we see that the vector $\mathbf{u} = [\upsilon_1, \ldots, \upsilon_{n_h}]^T$ is the solution of the linear system

$$(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f},$$

where

$$a_{i,j} = \langle s_{j,h}, s_{i,h} \rangle \quad \text{and} \quad b_{i,j} = \langle T_k s_{j.h}, s_{i,h} \rangle \qquad \text{for } 1 \le i, j \le n_h$$

and

$$\mathbf{f} = [\langle f, s_{1,h} \rangle \ldots \langle f, s_{n_h,h} \rangle]^T.$$

Of course, the pure FEM requires the calculation of $\langle f, s_i \rangle$ and $\langle T_k s_j, s_i \rangle$. These are weighted integrals of f and k. Since we are only using (noisy) standard information, such information about f and k is not available to us. Instead, we replace

f and k by their interpolants. This gives us an approximation, the modified MFEM, that uses only standard information.

More precisely, let $h, \bar{h} > 0$. For $[f, k] \in F$, we define

$$B_{\bar{h}}(v, w; k) = B(v, w; \Pi_{\bar{h}\otimes\bar{h}} k) \qquad \forall\, v \in L_p(I^d),\, w \in L_{p'}(I^d)$$

and let

$$f_h(w) = \langle \Pi_h f, w\rangle \qquad \forall\, w \in L_{p'}(I^d).$$

Note that for $v \in L_{p'}(I^d)$ and $w \in L_{p'}(I^d)$, we have

$$\langle T_{\Pi_{\bar{h}\otimes\bar{h}} k} v, w\rangle = \sum_{i,j=1}^{n_h} k(x_{i,\bar{h}}, x_{j,\bar{h}})\langle s_{j,\bar{h}}, v\rangle\langle s_{i,\bar{h}}, w\rangle,$$

so that

$$B_{\bar{h}}(v, w; k) = \langle v, w\rangle - \sum_{i,j=1}^{n_h} k(x_{i,\bar{h}}, x_{j,\bar{h}})\langle s_{j,\bar{h}}, v\rangle\langle s_{i,\bar{h}}, w\rangle.$$

Moreover

$$f_h(w) = \sum_{j=1}^{n_h} f(x_{j,h})\langle s_{j,h}, w\rangle \qquad \forall\, w \in L_{p'}(I^d).$$

The *modified finite element method* (MFEM) consists of finding $u_{h,\bar{h}} \in \mathscr{S}_h$ such that

$$B_{\bar{h}}(u_{h,\bar{h}}, w; k) = f_h(w) \qquad \forall\, w \in \mathscr{S}_h.$$

If we write

$$u_{h,\bar{h}}(x) = \sum_{j=1}^{n_h} \upsilon_j s_{j,h}(x) \qquad \forall\, x \in I^d,$$

then we see that the vector $\mathbf{u} = [\upsilon_1, \dots, \upsilon_{n_h}]^T$ is the solution of the linear system

$$(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}.$$

Here

$$a_{i,j} = \langle s_{j,h}, s_{i,h}\rangle \quad \text{and} \quad b_{i,j} = \langle T_{\Pi_{\bar{h}\otimes\bar{h}} k} s_{j,h}, s_{i,h}\rangle \qquad \text{for } 1 \le i, j \le n_h,$$

and

$$\mathbf{f} = [f_h(s_{1,h}) \dots f_h(s_{n_h,h})]^T.$$

Of course, the MFEM uses noise-free information. If we allow noisy evaluations in the MFEM, we get the noisy MFEM. More precisely, let $h, \bar{h}, \delta > 0$. For $[f, k] \in F$, we calculate

$$\tilde{f}_{i,\delta} \in \mathbb{R} \text{ such that } |f(x_{i,h}) - \tilde{f}_{i,\delta}| \le \delta \qquad \text{for } 1 \le i \le n_h,$$

and

$$\tilde{k}_{i,j,\delta} \in \mathbb{R} \text{ such that } |k(x_{i,h}, x_{j,h}) - \tilde{k}_{i,j,\delta}| \le \delta \qquad \text{for } 1 \le i, j \le n_{\bar{h}}.$$

Let

$$T_{k;\bar{h},\delta} v = \sum_{i,j=1}^{n_{\bar{h}}} \tilde{k}_{i,j,\delta} \langle s_{j,\bar{h}}, v \rangle s_{i,\bar{h}} \qquad \forall\, v \in L_p(I^d).$$

and

$$\Pi_{h,\delta} f = \sum_{j=1}^{n_h} \tilde{f}_{j,\delta} s_{j,h}.$$

For $k \in \mathscr{K}$, define a bilinear form $B_{\bar{h},\delta}(\cdot,\cdot;k)$ approximating $B_{\bar{h}}(\cdot,\cdot;k)$ as

$$B_{\bar{h},\delta}(v,w;k) = \langle v - T_{k;\bar{h},\delta} v, w \rangle \qquad \forall\, v \in L_p(I^d), w \in L_{p'}(I^d)$$

and a linear form $f_{h,\delta}$ approximating f_h as

$$f_{h,\delta}(w) = \langle \Pi_{h,\delta} f, w \rangle \qquad \forall\, w \in L_{p'}(I^d).$$

The *noisy modified finite element method* (noisy MFEM) consists of finding $u_{h,\bar{h},\delta} \in \mathscr{S}_h$ such that

$$B_{\bar{h},\delta}(u_{h,\bar{h},\delta}, w; k) = f_{h,\delta}(w) \qquad \forall\, w \in \mathscr{S}_h.$$

Writing

$$u_{h,\bar{h},\delta}(x) = \sum_{j=1}^{n_h} \upsilon_j s_{j,h}(x) \qquad \forall\, x \in I^d,$$

we see that the vector $\mathbf{u} = [\upsilon_1, \dots, \upsilon_{n_h}]^T$ is the solution of the linear system

$$(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}. \tag{15}$$

Here

$$a_{i,j} = \langle s_{j,h}, s_{i,h} \rangle \quad \text{and} \quad b_{i,j} = \langle T_{k;\bar{h},\delta} s_{j,h}, s_{i,h} \rangle \qquad \text{for } 1 \le i, j \le n_h,$$

and

$$\mathbf{f} = [f_{h,\delta}(s_{1,h}) \dots f_{h,\delta}(s_{n_h,h})]^T.$$

Let

$$\mathbb{N}_{h,\bar{h},\delta}([f,k]) = [\mathbb{N}_{h,\delta}(f), \overline{\mathbb{N}}_{\bar{h},\delta}(k)],$$

where

$$\mathbb{N}_{h,\delta}(f) = [\tilde{f}_{1,\delta}, \dots, \tilde{f}_{n_h,\delta}]$$

and

$$\overline{\mathbb{N}}_{\bar{h},\delta}(k) = [\overline{\mathbb{N}}^{(1)}_{\bar{h},\delta}(k), \dots, \overline{\mathbb{N}}^{(n_{\bar{h}})}_{\bar{h},\delta}(k)],$$

with

$$\overline{\mathbb{N}}^{(i)}_{\bar{h},\delta}(k) = [\tilde{k}_{i,1,\delta}, \dots, \tilde{k}_{i,n_{\bar{h}},\delta}] \qquad \text{for } 1 \le i \le n_{\bar{h}}.$$

If $u_{h,\bar{h},\delta}$ is well-defined, then we can write

$$u_{h,\bar{h},\delta} = \phi_{h,\bar{h},\delta}\big(\mathbb{N}_{h,\bar{h},\delta}([f,k])\big),$$

where

$$\operatorname{card} \mathbb{N}_{h,\bar{h},\delta} = n_{\bar{h}}^2 + n_h = \left(\frac{m+1}{\bar{h}}\right)^{2d} + \left(\frac{m+1}{h}\right)^{d} \asymp \bar{h}^{\,-2d} + h^{-d}.$$

6 Error analysis of the noisy modified FEM

In this section, we establish an error bound for the noisy modified FEM. We do this as follows. First, we establish the uniform weak coercivity of the bilinear forms $B(\cdot,\cdot;k)$ for $k \in \mathscr{K}$. Once we know that the bilinear forms are uniformly weakly coercive, we can obtain an error estimate by using Strang's lemma (see below). The remaining task is then to estimate the various terms appearing in Strang's lemma.

So, the first task is to establish uniform weak coercivity. Before doing so, we establish two auxiliary lemmas.

The first lemma shows that the inverses of certain operators are uniformly bounded. Let

$$h_0 = \left(\frac{1}{2\,c_1 c_2 M_1}\right)^{1/\min\{m+1,s\}}.$$

Recall that the *adjoint* of a linear transformation $A\colon L_p(I^d) \to L_p(I^d)$ of normed linear spaces is the linear operator $A^*\colon L_{p'}(I^d) \to L_{p'}(I^d)$ satisfying

$$\langle A^* v, w\rangle = \langle v, Aw\rangle \qquad \forall\, v \in L_p(I^d),\, w \in L_{p'}(I^d).$$

In particular, for any $k \in \mathscr{K}$, we have

$$(T_k^* w)(y) = \int_{I^d} k(x,y) w(x)\, dx \qquad \forall\, w \in L_{p'}(I^d).$$

Lemma 6.1. *Let $h \in (0, h_0]$ and $k \in \mathscr{K}$. Then $I - T^*_{\Pi_{h\otimes h}k}$ is invertible on $L_{p'}(I^d)$, with*

$$\|(I - T^*_{\Pi_{h\otimes h}k})^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]} \le 2\,c_2.$$

Proof. Let $h \in (0, h_0]$ and $k \in \mathscr{K}$. Note that since $(A^*)^{-1} = (A^{-1})^*$ for any invertible linear transformation A, we find that $I - T_k^*$ is invertible and

$$\|(I - T_k^*)^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]} \le c_2.$$

Let us write

$$I - T^*_{\Pi_{h\otimes h}k} = (I - T_k^*) + T^*_{k-\Pi_{h\otimes h}k}.$$

From (9) and Lemma 4.1, along with the definition of the class $\mathscr{K}$, we find

$$\begin{aligned}\|T^*_{k-\Pi_{h\otimes h}k}\|_{\mathrm{Lin}[L_{p'}(I^d)]} &\le \|k - \Pi_{h\otimes h}k\|_{L_\infty(I^{2d})} \le M_1 h^{\min\{m+1,s\}}\|k\|_{W^{s,\infty}(I^{2d})} \\ &\le M_1 h_0^{\min\{m+1,s\}} \cdot c_1 = \frac{1}{2c_2},\end{aligned}$$

and so

$$\|T^*_{k-\Pi_{h\otimes h}k}\|_{\mathrm{Lin}[L_{p'}(I^d)]}\|(I - T_k^*)^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]} \le \frac{1}{2\,c_2} \cdot c_2 = \tfrac{1}{2}.$$

From this inequality and [7, Lemma 1.3.14] we see that $I - T^*_{\Pi_{h\otimes h}k}$ is invertible, with

$$\begin{aligned}\|(I - T^*_{\Pi_{h\otimes h}k})^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]} &\le \frac{\|(I - T_k^*)^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]}}{1 - \|T^*_{k-\Pi_{h\otimes h}k}\|_{\mathrm{Lin}[L_{p'}(I^d)]}\|(I - T_k^*)^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]}} \\ &\le 2\,c_2,\end{aligned}$$

as required. □

Remark. Note that $T^*_{\Pi_{h\otimes h}k} \colon \mathscr{S}_h \to \mathscr{S}_h$. Hence if $h \in (0, h_0]$, the mapping $I - T^*_{\Pi_{h\otimes h}k}$ is an invertible linear operator on $\mathscr{S}_h$.

Our second auxiliary lemma shows that certain inner products can be bounded from below by products of norms.

Lemma 6.2. *Let $v \in L_p(I^d)$ be nonzero. For any $\tau \in (0, \|v\|_{L_p(I^d)})$, there is a nonzero function $g \in L_{p'}(I^d)$ such that*

$$\langle v, g\rangle \ge (\|v\|_{L_p(I^d)} - \tau)\|g\|_{L_{p'}(I^d)}.$$

Proof. Suppose first that $p < \infty$. Let $g = (\operatorname{sgn} v)|v|^{p-1}$. Then g is nonzero, with

$$\langle v, g\rangle = \|v\|_{L_p(I^d)}\|g\|_{L_{p'}(I^d)},$$

which is a stronger result than that which we want to prove. Hence it only remains to show that the lemma holds when $p = \infty$. We use an idea found on [1, pg. 26]. For $\tau \in (0, \|v\|_{L_\infty(I^d)})$, let

$$E = \{ x \in I^d : |v(x)| > \|v\|_{L_\infty(I^d)} - \tau \}.$$

From the definition of the essential supremum, meas $E > 0$. Let $g = (\operatorname{sgn} v)\chi_E$ be the characteristic function of E. Then g is a nonzero function, with

$$\|g\|_{L_1(I^d)} = \int_{I^d} \chi_E(x)\, dx = \text{meas}\, E.$$

Hence we have

$$\langle v, g \rangle = \int_E |v(x)|\, dx \geq (\|v\|_{L_\infty(I^d)} - \tau)\, \text{meas}\, E = (\|v\|_{L_\infty(I^d)} - \tau)\|g\|_{L_1(I^d)}.$$

Hence the lemma holds when $p = \infty$. ☐

Let

$$p' = \frac{p}{p-1}$$

denote the exponent conjugate to p. We are now ready to prove uniform weak coercivity of the bilinear forms $B(\cdot, \cdot; k)$ over all $k \in \mathscr{K}$.

Lemma 6.3. *There exist $h_1 > 0$ and $\gamma > 0$ such that the following holds: for any $k \in \mathscr{K}$, any $h \in (0, h_1]$, and any $v \in \mathscr{S}_h$, there exists nonzero $w \in \mathscr{S}_h$ such that*

$$B(v, w; k) \geq \gamma \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{16}$$

Proof. Let $k \in \mathscr{K}$ and $h \in (0, h_0]$. Let $v \in \mathscr{S}_h$. If $v = 0$, then this inequality holds for any nonzero $w \in \mathscr{S}_h$. So we may restrict our attention to the case $v \neq 0$.

By Lemma 6.2, there exists nonzero $g \in L_{p'}(I^d)$ such that

$$\langle v, g \rangle \geq \tfrac{1}{2} \|v\|_{L_p(I^d)} \|g\|_{L_{p'}(I^d)}.$$

Recalling the definition of the orthogonal projector P_h from (14) and using the remark following Lemma 6.1, we see that

$$w = (I - T^*_{\Pi_{h\otimes h}k})^{-1} P_h g$$

is a well-defined element of $\mathscr{S}_h$. Since $v \in \mathscr{S}_h$, we clearly have

$$\langle v, (I - T^*_{\Pi_{h\otimes h}k}) w \rangle = \langle v, g \rangle \geq \tfrac{1}{2} \|v\|_{L_p(I^d)} \|g\|_{L_{p'}(I^d)}.$$

Moreover, from Lemmas 4.2 and 6.1, we have

$$\begin{aligned}\|w\|_{L_{p'}(I^d)} &\le \|(I - T^*_{\Pi_{h\otimes h}k})^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]}\|P_h g\|_{L_{p'}(I^d)}\\ &\le 2\,c_2\|P_h g\|_{L_{p'}(I^d)} \le 2\,\pi_{p'}c_2\|g\|_{L_{p'}(I^d)}.\end{aligned}$$

Hence

$$\langle (I - T_{\Pi_{h\otimes h}k})v, w\rangle = \|v\|_{L_p(I^d)}\|g\|_{L_{p'}(I^d)} \ge \frac{1}{4\,\pi_{p'}c_2}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.$$

Since g and v are nonzero, this inequality implies that $\langle (I - T_{\Pi_{h\otimes h}k})v, w\rangle$ is nonzero. Since the latter is linear in w, we see that $w \neq 0$.

Using (9) and Lemma 4.1, we find

$$\begin{aligned}|\langle T_{k-\Pi_{h\otimes h}k}v, w\rangle| &\le \|T_{k-\Pi_{h\otimes h}k}v\|_{\mathrm{Lin}[L_p(I^d)]}\|w\|_{\mathrm{Lin}[L_{p'}(I^d)]}\\ &\le \|k - \Pi_{h\otimes h}k\|_{L_\infty(I^{2d})}\|v\|_{\mathrm{Lin}[L_p(I^d)]}\|w\|_{\mathrm{Lin}[L_{p'}(I^d)]}\\ &\le M_1 h^{\min\{m+1,s\}}\|k\|_{W^{s,\infty}(I^{2d})}\|v\|_{\mathrm{Lin}[L_p(I^d)]}\|w\|_{\mathrm{Lin}[L_{p'}(I^d)]}\\ &\le c_2 M_1 h^{\min\{m+1,s\}}\|v\|_{\mathrm{Lin}[L_p(I^d)]}\|w\|_{\mathrm{Lin}[L_{p'}(I^d)]}.\end{aligned}$$

Hence

$$\begin{aligned}B(v, w; k) &= \langle (I - T_{\Pi_{h\otimes h}k})v, w\rangle - \langle T_{k-\Pi_{h\otimes h}k}v, w\rangle\\ &\ge \left[\frac{1}{4\,\pi_{p'}c_2} - c_2 M_1 h^{\min\{m+1,s\}}\right]\|v\|_{\mathrm{Lin}[L_p(I^d)]}\|w\|_{\mathrm{Lin}[L_{p'}(I^d)]}.\end{aligned}$$

Letting

$$h_1 = \min\left\{\left(\frac{1}{8\,\pi_{p'}c_2^2 M_1}\right)^{1/\min\{m+1,s\}}, h_0\right\}$$

and

$$\gamma = \frac{1}{8\,\pi_{p'}c_2},$$

we see that the desired estimate (16) holds for $h \in (0, h_1]$. □

Since the bilinear forms $B(\cdot, \cdot; k)$ are uniformly weakly coercive for $k \in \mathscr{K}$, we have *Strang's lemma*:

Lemma 6.4. *Suppose there exist $\delta_0 \in (0, 1]$ and $h_2 \in (0, h_1]$ such that the following holds: for any $\delta \in [0, \delta_0]$, any $h, \bar{h} \in (0, h_2]$, and any $k \in \mathscr{K}$, we have*

$$|B(v, w; k) - B_{\bar{h},\delta}(v, w; k)| \le \tfrac{1}{2}\gamma\,\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \qquad \forall\, v, w \in \mathscr{S}_h,$$

where γ is as in Lemma 6.3. Then there exists $M_2 > 0$ such that the following hold for any $\delta \in [0, \delta_0]$ and any $h, \bar{h} \in (0, h_2]$:

1. *The noisy modified* FEM *is well-defined. That is, there exists a unique* $u_{h,\bar{h},\delta} \in \mathscr{S}_h$ *such that*

$$B_{\bar{h},\delta}(u_{h,\bar{h},\delta}, w) = f_{h,\delta}(w) \qquad \forall\, w \in \mathscr{S}_h.$$

2. *Let* $u = S([f,k])$*. Then*

$$\begin{aligned}\|u - u_{h,\bar{h},\delta}\|_{L_{p'}(I^d)} \le M_2 \inf_{v \in \mathscr{S}_h} \Bigg[&\|u - v_h\|_{L_p(I^d)} \\ &+ \sup_{w \in \mathscr{S}_h} \left(\frac{|B(v,w;k) - B_{\bar{h},\delta}(v,w;k)|}{\|w\|_{L_{p'}(I^d)}} + \frac{|\langle f, w\rangle - f_{h,\delta}(w)|}{\|w\|_{L_{p'}(I^d)}} \right) \Bigg]. \end{aligned} \tag{17}$$

We now estimate the quantities appearing on the right-hand side of (17).

Lemma 6.5. *There exists* $M_3 > 0$ *such that*

$$|B(v,w;k) - B_{\bar{h},\delta}(v,w;k)| \le M_3(\bar{h}^{\,\min\{m+1,s\}} + \delta)\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}$$

for any positive h*,* $\bar{h}$*, and* δ*, for any* $k \in \mathscr{K}$*, and for any* $v, w \in \mathscr{S}_h$*.*

Proof. Choose positive h, $\bar{h}$, and δ, along with $k \in \mathscr{K}$ and $v, w \in \mathscr{S}_h$. Then

$$|B(v,w;k) - B_{\bar{h},\delta}(v,w;k)| \le |A_1| + |A_2|, \tag{18}$$

where

$$A_1 = B(v,w;k) - B(v,w;\Pi_{\bar{h}\otimes\bar{h}}k) = \langle T_{k - \Pi_{\bar{h}\otimes\bar{h}}k} v, w\rangle$$

and

$$A_2 = B(v,w;\Pi_{\bar{h}\otimes\bar{h}}k) - B_{\bar{h},\delta}(v,w;k)| = \langle (T_{\Pi_{\bar{h}\otimes\bar{h}}k} - T_{k;\bar{h},\delta})v, w\rangle.$$

We first estimate $|A_1|$. Using (9) and Lemma 4.1, we find

$$\begin{aligned}|A_1| &\le \|T_{k-\Pi_{\bar{h}\otimes\bar{h}}k}\|_{\mathrm{Lin}[L_p(I^d)]}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \\ &\le \|k - \Pi_{\bar{h}\otimes\bar{h}}k\|_{L_\infty(I^{2d})}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \\ &\le c_1 M_1\, \bar{h}^{\,\min\{m+1,s\}}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.\end{aligned} \tag{19}$$

To estimate $|A_2|$, let

$$\zeta(x,y) = \sum_{i,j=1}^{n_{\bar{h}}} \Big(k(x_{i,\bar{h}}, x_{j,\bar{h}}) - \tilde{k}_{i,j,\delta}\Big)\, s_{j,\bar{h}}(y) s_{i,\bar{h}}(x).$$

Then

$$
\begin{aligned}
|A_2| &\le \left| \int_{I^d} \int_{I^d} \zeta(x, y) v(y) w(x) \, dy \, dx \right| \\
&\le \sup_{x, y \in I^{2d}} |\zeta(x, y)| \int_{I^d} |v(y)| \, dy \int_{I^d} |w(x)| \, dx \qquad (20) \\
&\le \|\zeta\|_{L_\infty(I^{2d})} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}.
\end{aligned}
$$

Now for $x \in I^d$, define $\operatorname{supp}_{\bar h} x$ as

$$
i \in \operatorname{supp}_{\bar h} x \qquad \text{iff} \qquad i \in \{1, \dots, n_{\bar h}\} \text{ and } x \text{ is in the support of } s_{i,\bar h}. \qquad (21)
$$

By construction of the basis functions for $\mathscr{S}_{\bar h}$, there exists positive constants σ_1 and σ_2, independent of x, j, and $\bar h$, such that

$$
|\operatorname{supp}_{\bar h} x| \le \sigma_1. \qquad (22)
$$

and

$$
\|s_{j,\bar h}\|_{L_\infty(I^d)} \le \sigma_2. \qquad (23)
$$

Hence for any $x, y \in I^d$, we have

$$
\begin{aligned}
|\zeta(x, y)| &\le \sum_{\substack{i \in \operatorname{supp}_{\bar h} x \\ j \in \operatorname{supp}_{\bar h} y}} \left| k(x_{i,\bar h}, x_{j,\bar h}) - \tilde k_{i,j,\delta} \right| |s_{j,\bar h}(y)| \, |s_{i,\bar h}(x)| \\
&\le \sigma \delta \sup_{1 \le j \le n_{\bar h}} \|s_{j,\bar h}\|^2_{L_\infty(I^d)} \le \sigma_1 \sigma_2^2 \, \delta.
\end{aligned}
$$

Since $x, y \in I^d$ are arbitrary, we thus have

$$
\|\zeta\|_{L_\infty(I^{2d})} \le \sigma_1 \sigma_2^2 \, \delta. \qquad (24)
$$

Using this inequality in (20), we obtain

$$
|A_2| \le \sigma_1 \sigma_2^2 \, \delta \, \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}.
$$

Combining this result with (20), recalling the decomposition (18), and letting

$$
M_3 = \max\{c_1 M_1, \sigma_1 \sigma_2^2\},
$$

we obtain the desired result. ☐

Lemma 6.6. *There exists $M_4 > 0$ such that*

$$
|\langle f, w \rangle - f_{h,\delta}(w)| \le M_4 (h^{\min\{m+1, r\}} + \delta) \|w\|_{L_{p'}(I^d)}
$$

for any positive h and δ, for any $f \in \mathscr{B}W^{r,p}(I^d)$, and for any $w \in \mathscr{S}_h$.

Proof. Choose positive h and δ, along with $f \in \mathscr{B}W^{r,p}(I^d)$ and $w \in \mathscr{S}_h$. Then

$$|\langle f, w\rangle - f_{h,\delta}(w)| \le |A_3| + |A_4|, \tag{25}$$

where

$$A_3 = \langle f - \Pi_h f, w\rangle$$

and

$$A_4 = \left\langle \Pi_h f - \sum_{j=1}^{n_h} \tilde{f}_{j,\delta} s_{j,h}, w \right\rangle.$$

We first estimate $|A_3|$. Using Lemma 4.1, we have

$$|A_3| \le \|f - \Pi_h f\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)} \le M_1 h^{\min\{m+1,r\}} \|w\|_{L_{p'}(I^d)}. \tag{26}$$

We now estimate $|A_4|$. We find

$$\begin{aligned} |A_4| &\le \left\| \sum_{j=1}^{n_h} [f(x_{j,h}) - \tilde{f}_{j,\delta}] s_{j,h} \right\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)} \\ &\le \delta \left\| \sum_{j=1}^{n_h} |s_{j,h}| \right\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \end{aligned}$$

Now

$$\left\| \sum_{j=1}^{n_h} |s_{j,h}| \right\|_{L_p(I^d)} \le \left\| \sum_{j=1}^{n_h} |s_{j,h}| \right\|_{L_\infty(I^d)}.$$

But for any $x \in I^d$, we may use (21)–(23) to see that

$$\sum_{j=1}^{n_h} |s_{j,h}(x)| = \sum_{j \in \operatorname{supp}_{\tilde{h}} x} |s_{j,h}(x)| \le \sigma_1 \sigma_2,$$

and thus

$$|A_4| \le \sigma_1 \sigma_2 \delta.$$

Using this inequality, along with (26), in (25), and setting

$$M_4 = \max\{M_1, \sigma_1\sigma_2\},$$

the desired result follows immediately. $\square$

The final preparatory step is to prove a "shift theorem," which relates the smoothness of $(I - T_k)^{-1}$ to the smoothnesses of f and of k.

Lemma 6.7. *Let* $0 \le t \le \min\{r, s\}$. *For* $k \in \mathscr{K}$ *and* $f \in W^{t,p}(I^d)$, *we have*

$$\|(I - T_k)^{-1}\|_{\mathrm{Lin}[W^{t,p}(I^d)]} \le 1 + c_2 c_3,$$

where

$$c_3 = \begin{cases} \binom{d+s}{d}^{1/p} c_1 & \text{if } p < \infty, \\ c_1 & \text{if } p = \infty. \end{cases} \tag{27}$$

Proof. Let $k \in \mathscr{K}$. First, we show that

$$\|T_k\|_{\mathrm{Lin}[L_p(I^d), W^{s,p}(I^d)]} \le c_3, \tag{28}$$

with $\|\cdot\|_{\mathrm{Lin}[L_p(I^d), W^{s,p}(I^d)]}$ denoting the usual operator norm. We shall prove only the case $p < \infty$, the case $p = \infty$ being analogous. Let α be a multi-index of order at most s. Then for any $v \in L_p(I^d)$, we have

$$\begin{aligned} |(\partial^\alpha T_k v)(x)| &= \left| \int_{I^d} \partial_x^\alpha k(x, y) v(y)\, dy \right| \\ &\le \sup_{y \in I^d} |\partial_x^\alpha k(x, y)| \|v\|_{L_p(I^d)} \\ &\le \|k\|_{W^{s,\infty}(I^{2d})} \|v\|_{L_p(I^d)}, \end{aligned}$$

so that

$$\|\partial^\alpha T_k v\|_{L_p(I^d)} \le \|k\|_{W^{s,\infty}(I^{2d})} \|v\|_{L_p(I^d)}.$$

Since α is an arbitrary multi-index of order at most s in d variables, we obtain

$$\|T_k v\|_{W^{s,p}(I^d)} = \left[\sum_{|\alpha| \le s} \|\partial^\alpha T_k v\|_{L_p(I^d)} \right]^{1/p} \le \binom{d+s}{s}^{1/p} \|k\|_{W^{s,\infty}(I^{2d})} \|v\|_{L_p(I^d)},$$

from which the desired result (28) follows.

Now let $f \in W^{t,p}(I^d)$, and set $u = (I - T_k)^{-1} f$. Since $u = f + T_k u$, we get

$$\|u\|_{W^{t,p}(I^d)} \le \|f\|_{W^{t,p}(I^d)} + \|T_k u\|_{W^{t,p}(I^d)}.$$

Now

$$\begin{aligned} \|T_k u\|_{W^{t,p}(I^d)} &\le \|T_k\|_{\mathrm{Lin}[L_p(I^d), W^{t,p}(I^d)]} \|u\|_{L_p(I^d)} \\ &\le \|T_k\|_{\mathrm{Lin}[L_p(I^d), W^{s,p}(I^d)]} \|u\|_{L_p(I^d)} \\ &\le c_3 \|u\|_{L_p(I^d)} \\ &\le c_3 \|(I - T_k)^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \|f\|_{L_p(I^d)} \\ &\le c_2 c_3 \|f\|_{L_p(I^d)}. \end{aligned}$$

Hence

$$\|(I - T_k)^{-1} f\|_{W^{t,p}(I^d)} \le \|f\|_{L_p(I^d)} + c_2 c_3 \|f\|_{L_p(I^d)} = (1 + c_2 c_3)\|f\|_{L_p(I^d)},$$

which establishes the desired result. □

We are now ready to show that the noisy modified FEM is well-defined, as well as to establish an upper bound on its error.

Theorem 6.1. *Let the degree m of the finite element spaces $\mathscr{S}_h$ and $\mathscr{S}_{\bar{h}}$ be chosen as*

$$m = \min\{r, s\} - 1.$$

Choose positive h_2 and δ_0 such that

$$M_3(h_2^s + \delta_0) \le \tfrac{1}{2}\gamma. \tag{29}$$

Then there exists $M_5 > 0$ such that the following hold for $h \in (0, h_1]$, $\bar{h} \in (0, h_2]$, and $\delta \in [0, \delta_0]$:

1. *The noisy modified FEM is well-defined.*
2. *We have the error bound*

$$e(\phi_{h,\bar{h},\delta}, \mathbb{N}_{h,\bar{h},\delta}) \le M_5(h^{\min\{r,s\}} + \bar{h}^{\,s} + \delta).$$

Proof. Let $h \in (0, h_1]$, $\bar{h} \in (0, h_2]$, and $\delta \in [0, \delta_0]$. Using Lemmas 6.4 and 6.5, we see that the noisy modified FEM is well defined. It only remains to establish the error bound.

For $[f, k] \in F$, let $u = S([f, k])$ and $u_{h,\bar{h},\delta} = \phi_{h,\bar{h},\delta}\big(\mathbb{N}_{h,\bar{h},\delta}([f, k])\big)$. Using Lemmas 4.1 and 6.7, and setting

$$C_4 = M_2(1 + c_2 c_3),$$

we find

$$\begin{aligned} \|u - \Pi_h u\|_{L_p(I^d)} &\le M_1 h^{\min\{r,s\}} \|u\|_{W^{\min\{r,s\},p}(I^d)} \\ &\le M_1(1 + c_2 c_3) h^{\min\{r,s\}} \|f\|_{W^{\min\{r,s\},p}(I^d)} \\ &\le C_4 h^{\min\{r,s\}}. \end{aligned} \tag{30}$$

Now let $w \in \mathscr{S}_h$. By the definition of c_2, we find

$$\begin{aligned} \|\Pi_h u\|_{L_p(I^d)} &\le \|u - \Pi_h u\|_{L_p(I^d)} + \|u\|_{L_p(I^d)} \\ &\le C_4 h_p^{\min\{r,s\}} + c_2 \|f\|_{L_p(I^d)} \le C_4 + c_2, \end{aligned}$$

and thus using Lemma 6.5, we find that

$$\begin{aligned}|B(v,w;k)-B_{\bar h,\delta}(v,w;k)| &\le M_3(\bar h^{\,s}+\delta)\|\Pi_h u\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}\\ &\le (C_4+c_2)M_3(\bar h^{\,s}+\delta)\|w\|_{L_{p'}(I^d)}.\end{aligned} \tag{31}$$

Moreover using Lemma 6.6, we have

$$|\langle f,w\rangle - f_{h,\delta}| \le M_4(h^r+\delta)\|w\|_{L_{p'}(I^d)}. \tag{32}$$

Hence using (30)–(32) in Lemma 6.4, we get

$$\|u-u_{h,\bar h,\delta}\|_{L_p(I^d)} \le M_2\left(C_4 h^{\min\{r,s\}} + (C_4+c_2)M_3(\bar h^{\,s}+\delta) + M_4(h^r+\delta)\right).$$

Taking

$$M_5 = M_2\big(C_4 + (C_4+c_2)M_3 + M_4\big),$$

we get the desired error bound. □

Remark. We have a wide amount of latitude in choosing parameters h_2 and δ_0 such that (29) holds. One simple choice is to pick

$$h_2 = \left(\frac{\gamma}{4M_3}\right)^{1/s} \qquad \text{and} \qquad \delta_0 = \frac{\gamma}{4M_3}.$$

7 The noisy modified FEM is a minimal error algorithm

Let $n \in \mathbb{Z}^+$. In this section, we show how to choose the meshsizes h and $\bar h$ such that the noisy modified FEM is an nth minimal error algorithm.

We define integer parameters l and $\bar l$, as follows:

1. Suppose that $s < 2r$. In this case, we have $s < 2\min\{r,s\}$. Take

$$l = \lceil n^{s/(2\min\{r,s\})}\rceil \qquad \text{and} \qquad \bar l = \left\lfloor \sqrt{n-l}\right\rfloor.$$

2. Suppose that $s = 2r$. Take

$$l = \lceil \tfrac12 n\rceil \qquad \text{and} \qquad \bar l = \left\lfloor \sqrt{\tfrac12 n}\right\rfloor.$$

3. Suppose that $s > 2r$. Take

$$\bar l = \lceil n^{r/s}\rceil \qquad \text{and} \qquad l = n - \bar l^{\,2}.$$

With these definitions for l and $\bar{l}$, define

$$h = \frac{\min\{r, s\}}{l^{1/d}} \qquad \text{and} \qquad \bar{h} = \frac{\min\{r, s\}}{\bar{l}^{\,1/d}}.$$

Recalling that the degree m of our finite element spaces is given by

$$m = \min\{r, s\} - 1,$$

we see that

$$n_h = l \qquad \text{and} \qquad n_{\bar{h}} = \bar{l}$$

by (13). With these choices of h and $\bar{h}$, let

$$\mathbb{N}_{n,\delta} = \mathbb{N}_{h,\bar{h},\delta} \qquad \text{and} \qquad \phi_{n,\delta} = \phi_{h,\bar{h},\delta}.$$

That is, for any $[f, k] \in F$, we have

$$\mathbb{N}_{n,\delta}([f, k]) = [\mathbb{N}_{l,\delta}(f), \overline{\mathbb{N}}_{\bar{l}^2,\delta}(k)],$$

where

$$\mathbb{N}_{l,\delta}(f) = \mathbb{N}_{h,\delta}(f) \qquad \text{and} \qquad \overline{\mathbb{N}}_{\bar{l}^2,\delta}(k) = \overline{\mathbb{N}}_{\bar{h},\delta}(k).$$

Since $\mathbb{N}_{n,\delta}([f, k])$ uses $\bar{l}^2$ noisy evaluations of k and l of f, we have

$$\operatorname{card} \mathbb{N}_{n,\delta} = \bar{l}^2 + l \le n.$$

We now have

Theorem 7.1. *Recall from* (2) *that*

$$\mu = \min\left\{\frac{r}{d}, \frac{s}{2d}\right\}.$$

1. *There exists $n_0^* \in \mathbb{Z}^+$ such that the $\phi_{n,\delta}$ is well-defined for all $n \ge n_0^*$ and all $\delta \in [0, \delta_0]$.*
2. *There exists $M_6 > 0$ such that*

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \le M_6(n^{-\mu} + \delta) \qquad \text{for } n \ge n_0^* \text{ and } \delta \in [0, \delta_0]. \tag{33}$$

3. *The nth minimal radius satisfies*

$$r_n(\delta) \asymp n^{-\mu} + \delta.$$

4. *The information $\mathbb{N}_{n,\delta}$ is nth optimal information, and $\phi_{h,\delta}$ is an nth minimal error algorithm.*

Proof. The first item follows from Theorem 6.1. Once we establish (33), the remaining items will then follow immediately from (33) and Theorem 3.1. Hence, it remains to prove (33).

We prove (33) on a case-by-case basis. Suppose first that $r < 2s$. We then have

$$h \asymp l^{-1/d} \asymp n^{-s/(2\min\{r,s\}d)} \qquad \text{and} \qquad \bar{h} \asymp \bar{l}^{\,-1/d} \asymp n^{-1/(2d)}.$$

Since $s < 2r$, we have $\mu = s/(2d)$. Hence

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \preccurlyeq h^{\min\{r,s\}} + \bar{h}^{\,s} + \delta \preccurlyeq n^{-s/(2d)} + \delta \asymp n^{-\mu} + \delta.$$

Next, suppose that $r = 2s$. We have

$$h \asymp l^{-1/d} \asymp n^{-1/d} \qquad \text{and} \qquad \bar{h} \asymp l^{-1/d} \asymp n^{-1/(2d)}.$$

Since $r < 2r = s$, we have $\min\{r, s\} = r$. Thus

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \preccurlyeq h^{\min\{r,s\}} + \bar{h}^{\,s} + \delta \preccurlyeq n^{-r/d} + n^{-s/(2d)} + \delta \asymp n^{-\mu} + \delta.$$

Finally, suppose that $r > 2s$. We have

$$h \asymp l^{-1/d} \asymp n^{-1/d} \qquad \text{and} \qquad \bar{h} \asymp \bar{l}^{\,-1/d} \asymp n^{-r/(ds)}.$$

Since $r < 2r < s$, we have $\min\{r, s\} = r$. Thus

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \preccurlyeq h^{\min\{r,s\}} + \bar{h}^{\,s} + \delta \preccurlyeq n^{-r/d} + \delta.$$

But since $s > 2r$, we have $\mu = r/d$. Thus

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \preccurlyeq n^{-\mu} + \delta.$$

Hence (33) holds in all three cases. $\square$

8 Two-grid implementation of the noisy modified FEM

We have just seen that $\phi_{n,\delta}$ is an nth minimal error algorithm. Its information cost is $c(\delta)\, n$. Hence if we were only interested in informational complexity, then we would have a source of optimal algorithms, see, e.g., [13, Section 4.4].

Unfortunately, the combinatory cost of this algorithm is generally much worse than $\Theta(n)$. Indeed, for any $[f, k] \in F$ and any $n \geq n_0^*$, this algorithm presents us with a linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$. The matrix $\mathbf{B}$ is a full $l \times l$ matrix, where

$$l \asymp \begin{cases} n^{s/(2\min\{r,s\})} & \text{if } s < 2r, \\ n & \text{if } s \geq 2r. \end{cases}$$

Hence, if we were to use Gaussian elimination to solve this linear system, the combinatory cost would be proportional to n^κ, where

$$\kappa = \begin{cases} \dfrac{3s}{2\min\{r,s\}} & \text{if } s < 2r, \\ 3 & \text{if } s \ge 2r. \end{cases}$$

Since $\kappa \in [\frac{3}{2}, 3]$, the combinatory cost is not $O(n)$.

Rather than using Gaussian elimination to directly solve the linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$, we shall use a two-grid algorithm to obtain a sufficiently accurate approximation of the solution $\mathbf{u}$. This will give us a nearly optimal approximation at nearly optimal cost.

Our approach will closely follow that of [7]. For given n, we shall define l, $\bar{l}$, h and $\bar{h}$ as at the beginning of Section 7. This will give us a linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$ whose solution we wish to approximate. We let n^* be a second integer, satisfying $n^* = \Theta(n^{1/3})$. If we were to set up the linear system corresponding to the noisy MFEM using information of cardinality n^*, we would get an $l^* \times l^*$ linear system $(\tilde{\mathbf{A}} - \tilde{\mathbf{B}})\tilde{\mathbf{u}} = \tilde{\mathbf{f}}$. Here, l^*, $\overline{l^*}$, h^*, and $\overline{h^*}$ are the parameters for the noisy MFEM using information of cardinality n^*, as defined at the beginning of Section 7.

Before describing the two-grid method, we need to introduce some prolongation and restriction operators, as described in Sections 5.2 and 5.3 of [7]. Let $X = L_p(I^d)$, $X_l = (\mathbb{R}^l, \|\cdot\|_{\ell_p})$, and $X_{l^*} = (\mathbb{R}^{l^*}, \|\cdot\|_{\ell_p})$. We define the *canonical prolongation* $P_h \colon X_l \to X$ as

$$P_h\mathbf{v} = \sum_{j=1}^{l} v_j s_{j,h} \qquad \forall\, \mathbf{v} = [v_1 \ldots v_l] \in \mathbb{R}^l.$$

The *canonical restriction* $R_h \colon X \to X_l$ is defined as

$$R_h w = \mathbf{A}^{-1}[\langle w, s_{1,h}\rangle \ldots \langle w, s_{l,h}\rangle]^T \qquad \forall\, w \in X.$$

Note that P_h and R_h are uniformly bounded mappings, i.e., there exist positive constants C_{P} and C_{R} such that

$$\|P_h\|_{\mathrm{Lin}[X_l,X]} \le C_{\mathrm{P}} \quad \text{and} \quad \|R_h\|_{\mathrm{Lin}[X,X_l]} \le C_{\mathrm{R}} \qquad \forall\, h > 0. \tag{34}$$

Moreover

$$R_h P_h = I \qquad \text{and} \qquad P_h R_h = \Pi_h. \tag{35}$$

(See [7, pg. 161].)

We then define the *intergrid prolongation operator* $\mathfrak{p} \colon X_{l^*} \to X_l$ and the *intergrid restriction operator* $\mathfrak{r} \colon X_l \to X_{l^*}$ as

$$\mathfrak{p} = R_h P_{h^*} \qquad \text{and} \qquad \mathfrak{r} = R_{h^*} P_h.$$

We will also need to use the adjoint operator $\mathfrak{p}^*\colon X_l \to X_{l^*}$, defined as

$$\mathfrak{p}^*\mathbf{v}\cdot\mathbf{w} = \mathbf{v}\cdot\mathfrak{p}\mathbf{w} \qquad \forall\,\mathbf{v}\in X_l,\,\mathbf{w}\in X_{l^*}.$$

We are now ready to define the two-grid iteration scheme. This is the variant TGM$'$ found on [7, pg. 179].

function TG($n : \mathbb{Z}^+;\ \mathbf{A},\mathbf{B} : \mathbb{R}^{l\times l}; \mathbf{f} : \mathbb{R}^l$) : $\mathbb{R}^l$;
begin
  if n is sufficiently small **then**
    compute $\mathbf{u}\in\mathbb{R}^l$ such that $(\mathbf{A}-\mathbf{B})\mathbf{u}=\mathbf{f}$
  else
    begin
      $\mathbf{u} := \mathbf{0}$;
      for $i := 1$ **to** 3 **do**
        begin
          Solve the linear system $\mathbf{A}\tilde{\mathbf{u}} = \mathbf{f}+\mathbf{B}\mathbf{u}$; {Picard iteration}
          $\mathbf{d} := \mathfrak{p}^*(\mathbf{A}\tilde{\mathbf{u}}-\mathbf{f}-\mathbf{B}\tilde{\mathbf{u}})$; {compute defect}
          solve the system $(\tilde{\mathbf{A}}-\tilde{\mathbf{B}})\delta = \mathbf{d}$; {coarse-grid solution}
          $\mathbf{u} := \mathbf{u}-\mathfrak{p}\delta$ {coarse-grid correction}
        end
    end;
  TG := $\mathbf{u}$
end;

Finally, let

$$\check{\mathbb{N}}_{n,\delta} = [\overline{\mathbb{N}}_{\bar{l}^2,\delta}, \overline{\mathbb{N}}_{l^2,\delta}, \mathbb{N}_{l,\delta}]$$

be *two-grid information* of cardinality at most n. Let

$$\check{u}_{n,\delta} = P_h[\mathrm{TG}(n,\mathbf{A},\mathbf{B},\mathbf{f})] = \sum_{j=1}^{l} \upsilon_j s_{j,h}, \tag{36}$$

Then $\check{u}_{n,\delta}$ depends on $[f,k]\in F$ only through the information $\check{\mathbb{N}}_{n,\delta}([f,k])$, and so we may write $\check{u}_{n,\delta} = \check{\phi}_{n,\delta}(\check{\mathbb{N}}_{n,\delta}([f,k]))$, where $\check{\phi}_{n,\delta}$ is an algorithm using the information $\check{\mathbb{N}}_{n,\delta}$. We call $\check{\phi}_{n,\delta}$ the *two-grid algorithm*.

Our first task is to analyze the cost of the two-grid algorithm. Before doing this, we prove the following

Lemma 8.1. *Let $n\in\mathbb{Z}^+$. For $\mathbf{v}\in\mathbb{R}^l$, we can calculate* $\mathbf{B}\mathbf{v}$ *using at most $O(n)$ operations.*

Proof. Let $\mathbf{S} \in \mathbb{R}^{\bar{l}\times l}$ have $(\bar{j}, j)$ entry

$$\sigma_{\bar{j},j} = \langle s_{\bar{j},\bar{h}}, s_{j,h}\rangle \qquad \text{for } 1 \le \bar{j} \le \bar{l},\, 1 \le j \le l,$$

and let $\mathbf{C} = [\tilde{k}_{\bar{\imath},\bar{j},\delta}]_{1\le\bar{\imath},\bar{j}\le\bar{l}}$. We then have $\mathbf{B} = \mathbf{S^T C S}$. For $\mathbf{v} \in \mathbb{R}^l$, we calculate $\mathbf{Bv}$ as follows:

1. Let $\mathbf{a} = \mathbf{Sv} \in \mathbb{R}^{\bar{l}}$. Since each row of $\mathbf{S}$ has only $O(1)$ nonzero elements, this matrix/vector multiplication can be done in at most $O(l)$ operations.

2. Let $\mathbf{b} = \mathbf{Ca} \in \mathbb{R}^{\bar{l}}$. This is the usual multiplication of an $\bar{l} \times \bar{l}$ matrix by an $\bar{l}$-vector, which can be done in at most $O(\bar{l}^2)$ operations.

3. Let $\mathbf{c} = \mathbf{S^T b} \in \mathbb{R}^l$. Since each row of $\mathbf{S}^T$ has only $O(1)$ nonzero elements, this matrix/vector multiplication can be done in at most $O(l)$ operations.

Then $\mathbf{Bv} = \mathbf{c}$. Moreover, the cost of calculating $\mathbf{z}$ is clearly $O(\bar{l}^2 + l) = O(n)$ operations, as required. $\square$

We then have

Lemma 8.2. *The cost of the two-grid algorithm satisfies*

$$\operatorname{cost}(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \preccurlyeq c(\delta)\, n.$$

Proof. By construction, the information $\check{\mathbb{N}}_{n,\delta}$ has cardinality proportional to n. Hence the information cost of the two-grid algorithm is at most $c(\delta)\, n$. Hence, it remains to determine the combinatory cost.

Let $[f, k] \in F$. We need to find the cost of computing $\mathrm{TG}(n, \mathbf{A}, \mathbf{B}, \mathbf{f})$.

1. We first do the Picard iteration. From Lemma 8.1, evaluating $\mathbf{Bu}$ costs $O(n)$, and hence the cost of evaluating $\mathbf{z} = \mathbf{f} + \mathbf{Bu}$ is also $O(n)$. Furthermore, the bandwidth of $\mathbf{A}$ is bounded, independent of n, since there are no interelement continuity requirements. Thus the cost of the Picard iteration step is $O(n)$ operations.

2. Next, we compute the defect. Since the number of elements in any row of $\mathbf{A}$ is bounded, the cost of evaluating $\mathbf{A\tilde{u}}$ is $O(l)$ operations. By Lemma 8.1, we can calculate $\mathbf{B\tilde{u}}$ in $O(n)$ operations. Thus we can calculate $\mathbf{w} = \mathbf{A\tilde{u}} - \mathbf{f} - \mathbf{B\tilde{u}}$ in $O(n)$ operations. It only remains to calculate $\mathfrak{p}^*\mathbf{w}$, which can clearly be done in $O(l)$ operations.

3. To calculate the coarse-grid solution, we need to solve an $n^* \times n^*$ linear system. Since $n^* = \Theta(n^{1/3})$, we can do this in $O(n)$ operations.

4. The coarse-grid correction can clearly be done in $O(l)$ operations.

Thus we can compute $\mathrm{TG}(n, \mathbf{A}, \mathbf{B}, \mathbf{f})$ with a cost of at most $O(\bar{l}^2 + l) = O(n)$ operations, as required. $\square$

Our next task is to analyze the error of the two-grid approximation. Before doing this, we need to do a little groundwork. Write $Y = W^{\min\{r,s\},\infty}(I^d)$. Let $Y_l = (\mathbb{R}^l, \|\cdot\|_{Y_l})$, where

$$\|\mathbf{v}\|_{Y_l} = \inf_{v \in R_l^{-1}(\mathbf{v})} \|v\|_Y.$$

The norm $\|\cdot\|_{Y_{l^*}}$ and space Y_{l^*} are defined analogously.

For future reference, we note that the linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$ may be rewritten in the form $(\mathbf{I} - \mathbf{K})\mathbf{u} = \mathbf{g}$, where

$$\mathbf{K} = \mathbf{A}^{-1}\mathbf{B} \qquad \text{and} \qquad \mathbf{g} = \mathbf{A}^{-1}\mathbf{f}.$$

We will also have cause to refer to the matrix $\tilde{\mathbf{K}} = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}$. We have the following

Lemma 8.3. *There exist positive constants C_{S}, C_{K}, C_{B}, C_{I}, and C_{C}, which are independent of n, such that the following hold:*

1. *Stability:* $\|(\mathbf{I} - \mathbf{K})^{-1}\|_{\mathrm{Lin}[X_l]} \le C_{\mathrm{S}}$.
2. *Discrete regularity:* $\|\mathbf{K}\|_{\mathrm{Lin}[Y_l]} \le C_{\mathrm{K}}$.
3. *Uniform boundedness of prolongations:* $\|\mathfrak{p}\|_{\mathrm{Lin}[X_{l^*}, X_l]} \le C_{\mathrm{B}}$.
4. *Interpolation error:* $\|\mathbf{I} - \mathfrak{p}\mathfrak{r}\|_{\mathrm{Lin}[Y_l, X_l]} \le C_{\mathrm{I}}(l^*)^{-\min\{r,s\}/d}$.
5. *Relative consistency:* $\|\mathfrak{r}\mathbf{K} - \tilde{\mathbf{K}}\mathfrak{r}\|_{\mathrm{Lin}[Y_l, X_{l^*}]} \le C_{\mathrm{C}}\left(\overline{l^*}\right)^{-s/d} + \delta$.

Proof. We first prove stability. Let $\mathbf{f} = [\alpha_1 \ldots \alpha_l]^T \in X_l$ and $\mathbf{u} = (\mathbf{I} - \mathbf{K})^{-1}\mathbf{f}$. Let $\breve{u} = P_h\mathbf{u}$ and $\breve{f} = P_h\mathbf{f}$. Then

$$B_{h,\bar{h},\delta}(\breve{u}, w) = \breve{f}_{h,\delta}(w) \qquad \forall\, w \in \mathscr{S}_h.$$

Using Lemmas 6.3 and 6.5, there exists nonzero $w \in \mathscr{S}_h$ such that

$$B_{h,\bar{h},\delta}(\breve{u}, w) \ge \tfrac{1}{2}\gamma \|\breve{u}\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}.$$

Using Lemma 6.6, we easily find that

$$|\breve{f}_{h_\ell,\delta}(v)| \le 3M_4 \|\breve{f}\|_{L_p(I^d)} \|v\|_{L_{p'}(I^d)},$$

and thus

$$\|\breve{u}\|_{L_p(I^d)} \le \frac{6M_4}{\gamma}\|\breve{f}\|_{L_p(I^d)}. \tag{37}$$

From (35), we have $\mathbf{u} = R_h\breve{u}$. Using (34), we see that

$$\|\mathbf{u}\|_{X_l} \le C_{\mathrm{R}}\|\breve{u}\|_X. \tag{38}$$

Since $\breve{f} = \sum_{j=1}^{l}\alpha_j s_{j,h}$, we use the discrete Minkowski inequality to find

$$\begin{aligned}\|\breve{f}\|_X &= \left(\int_{I^d}\left|\sum_{j=1}^{l}\alpha_j s_{j,h}(x)\right|^p\right)^{1/p} dx \\ &\le \left\{\int_{I^d}\left[\left(\sum_{j=1}^{l}|\alpha_j|^p\right)^{1/p}\left(\sum_{j=1}^{l}|s_{j,h}(x)|^{p'}\right)^{1/p'}\right]^p dx\right\}^{1/p} \\ &= \theta\|\mathbf{f}\|_{X_l},\end{aligned} \tag{39}$$

where

$$\theta = \left[\int_{I^d}\left(\sum_{j=1}^{l}|s_{j,h}(x)|^{p'}\right)^{p/p'} dx\right]^{1/p}.$$

Now

$$\sum_{j=1}^{l}|s_{j,h}(x)|^{p'} \le \max_{1\le j\le l}\|s_{j,h}\|_{L_\infty(I^d)}^{p'}\,|\operatorname{supp}_l x|,$$

where

$$j \in \operatorname{supp}_l x \qquad \text{iff} \qquad j \in \{1,\ldots,l\} \text{ and } x \text{ is in the support of } s_{j,h}.$$

As in the proof of Lemma 6.5, there exist positive constants σ_1 and σ_2, independent of x, j, and ℓ, such that

$$\max_{1\le j\le l}\|s_{j,h}\|_{L_\infty(I^d)} \le \sigma_1 \qquad \text{and} \qquad |\operatorname{supp}_l x| \le \sigma_2.$$

Hence

$$\theta \le \sigma_1\sigma_2^{1/p'}.$$

Using (39), we find that

$$\|\breve{f}\|_X \le \sigma_1\sigma_2^{1/p'}\|\mathbf{f}\|_{X_l}. \tag{40}$$

Let $C_{\mathrm{S}} = 6M_4C_{\mathrm{R}}\sigma_1\sigma_2^{1/p'}/\gamma$. Using (37), (38), and (40), we obtain

$$\|(\mathbf{I}-\mathbf{K})^{-1}\mathbf{f}\|_{X_l} = \|\mathbf{u}\|_{X_l} \le C_{\mathrm{S}}\|\mathbf{f}_l\|_{X_\ell}.$$

Since $\mathbf{f} \in X_l$ is arbitrary, we find that part 1 holds, as required.

We next check that discrete regularity holds. From [7, Remark 5.2.3], we find that

$$\mathbf{K} = R_h T_{k;h,\bar{h},\delta} P_h. \tag{41}$$

Using the definition of the norm $\|\cdot\|_{Y_l}$, we find that $\|R_h\|_{\mathrm{Lin}[Y,Y_l]} = 1$, and so

$$\begin{aligned}\|\mathbf{K}\|_{\mathrm{Lin}[X_l,Y_l]} &\le \|R_h\|_{\mathrm{Lin}[Y,Y_l]} \|T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]} \|P_h\|_{\mathrm{Lin}[X_l,X]} \\ &\le C_{\mathrm{P}}\, \|T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]},\end{aligned}$$

where C_{P} is defined by (34). Now

$$\|T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]} \le \|T_k\|_{\mathrm{Lin}[Y,X]} + \|T_k - T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]}.$$

From (28), we have

$$\|T_k\|_{\mathrm{Lin}[Y,X]} \le c_3,$$

whereas from the proofs of Lemma 6.5 and Theorem 7.1, we find that

$$\|T_k - T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]} \preccurlyeq \bar{h}^s + \delta \preccurlyeq n^{-\mu} + \delta.$$

Combining the previous inequalities, we see that part 2 holds.

To prove uniform boundedness of prolongations, we use Exercise 5.3.6(a) on [7, pg. 171], finding that part 3 holds with $C_{\mathrm{B}} = C_{\mathrm{R}} C_{\mathrm{P}}$.

Next, we establish the interpolation error. Note that since (35) holds, we have

$$I - \mathfrak{pr} = R_h P_h - R_h P_{h^*} R_{h^*} P_h = R_h(I - \Pi_{h^*}) P_h.$$

Hence using Lemma 4.1, we find

$$\begin{aligned}\|I - \mathfrak{pr}\|_{\mathrm{Lin}[Y_l,X_l]} &\le \|R_h\|_{\mathrm{Lin}[X,X_l]} \|I - \Pi_{h_{l^*}}\|_{\mathrm{Lin}[Y,X]} \|P_h\|_{\mathrm{Lin}[Y_l,Y]} \\ &\le C_{\mathrm{R}} C_{\mathrm{P}} M_1 (h^*)^{\min\{r,s\}} \preccurlyeq (l^*)^{-\min\{r,s\}/d},\end{aligned}$$

so that part 4 holds with $C_{\mathrm{I}} = C_{\mathrm{R}} C_{\mathrm{P}} M_1$.

We now establish relative consistency, using a perturbation of the proof of [7, Lemma 5.3.11]. Using (41), we have

$$\mathfrak{r}\mathbf{K}\mathfrak{p} = (\mathfrak{r}R_h) T_{k;h,\bar{h},\delta}(P_h \mathfrak{p}) = R_{h^*} T_{k;h,\bar{h},\delta} P_{h^*} = \tilde{\mathbf{K}} + R_{h^*} E P_{h^*},$$

where

$$E = T_{k;h^*,\overline{h^*},\delta} - T_{k;h,\bar{h},\delta}.$$

Hence

$$\mathfrak{r}\mathbf{K} - \tilde{\mathbf{K}}\mathfrak{r} = \mathfrak{r}\mathbf{K}(I - \mathfrak{pr}) + R_{h^*} E P_{h^*} \mathfrak{r}.$$

Now

$$\|\mathfrak{r}\mathbf{K}(I-\mathfrak{p}\mathfrak{r})\|_{\mathrm{Lin}[Y_l,X_l]} \le C_{\mathrm{B}}C_{\mathrm{K}}C_{\mathrm{I}}(l^*)^{-\mu}.$$

Moreover,

$$\|R_{h^*}EP_{h^*}\mathfrak{r}\|_{\mathrm{Lin}[Y_l,X_l]} \le C_{\mathrm{R}}C_{\mathrm{P}}C_{\mathrm{B}}\|E\|_{\mathrm{Lin}[Y,X]}.$$

Now

$$\begin{aligned}\|E\|_{\mathrm{Lin}[Y,X]} &\le \|I-T_{k;h^*,\overline{h^*},\delta}\|_{\mathrm{Lin}[Y,X]} + \|I-T_{k;h,\bar{h},\delta}\|_{\mathrm{Lin}[Y,X]}\\ &\preccurlyeq \left(\overline{h^*}\right)^s + \delta \preccurlyeq \left(\overline{l^*}\right)^{s/d} + \delta\end{aligned}$$

the latter following from the proofs of Lemma 6.5 and Theorem 7.1. Combining these results, we see that part 5 holds, as claimed. ☐

Using some of the ideas found in the proofs of [7, Theorem 5.5.7 and Theorem 5.6.4], we are now ready to estimate the distance between the exact solution $\mathbf{u}$ of the linear system $(\mathbf{I}-\mathbf{K})\mathbf{u}=\mathbf{f}$ and the solution $\tilde{\mathbf{u}} = \mathrm{TG}(n,\mathbf{A},\mathbf{B},\mathbf{f})$ produced by the two-grid method.

Lemma 8.4. *We have*

$$\|\tilde{\mathbf{u}}-\mathbf{u}\|_{X_l} \preccurlyeq \left(n^{-\mu}+\delta\right)\|\mathbf{f}\|_{X_l}.$$

Proof. It is no loss of gereality to assume that n is sufficiently large that we do not solve the linear system $(\mathbf{A}-\mathbf{B})\mathbf{u}=\mathbf{f}$ directly. Let

$$\mathbf{M}^{\mathrm{TG}} = \mathbf{I} - (\mathbf{I}-\tilde{\mathbf{K}})^{-1}\mathfrak{r}(\mathbf{I}-\tilde{\mathbf{K}})\mathbf{K}$$

and

$$\mathbf{c} = (\mathbf{I}-\tilde{\mathbf{K}})^{-1}\mathfrak{r}(\mathbf{I}-\tilde{\mathbf{K}})\mathbf{K}(\mathbf{I}-\mathbf{K})^{-\mathbf{1}}\mathbf{f}.$$

Using Lemma 8.3 and [7, Theorem 5.4.3], we have

$$\|\mathbf{M}^{\mathrm{TG}}\|_{\mathrm{Lin}[X_l]} \le C_{\mathrm{TG}}\left((l^*)^{-\min\{r,s\}/d} + \left(\overline{l^*}\right)^{-s/d} + \delta\right),$$

where $C_{\mathrm{TG}} = (C_{\mathrm{I}} + C_{\mathrm{B}}C_{\mathrm{S}}C_{\mathrm{C}})C_{\mathrm{K}}$. Arguing as in the proof of Theorem 7.1, we find that

$$\|\mathbf{M}^{\mathrm{TG}}\|_{\mathrm{Lin}[X_l]} \le\preccurlyeq (n^*)^{-\mu} + \delta.$$

Since $n^* = \Theta(n^{1/3})$, it follows that

$$\|\mathbf{M}^{\mathrm{TG}}\|_{\mathrm{Lin}[X_l]} \preccurlyeq n^{-\mu/3} + \delta,$$

It is fairly easy to check (see also [7, Theorem 5.4.3]) that $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}^{(3)}$, where

$$\begin{aligned}&\tilde{\mathbf{u}}^{(0)} = \mathbf{0},\\ &\tilde{\mathbf{u}}^{(i)} = \mathbf{M}^{\mathrm{TG}}\tilde{\mathbf{u}}^{(i-1)} + \mathbf{c} \qquad (1 \le i \le 3).\end{aligned}$$

Moreover, it is also easy to see that $\mathbf{u} = \mathbf{M}^{\mathrm{TG}}\mathbf{u} + \mathbf{c}$, so that

$$\|\mathbf{u} - \tilde{\mathbf{u}}^{(i)}\|_{X_l} \le \|\mathbf{M}^{\mathrm{TG}}\|_{\mathrm{Lin}[X_l]}\|\mathbf{u} - \tilde{\mathbf{u}}^{(i-1)}\|_{X_l}.$$

Combining these results, we find

$$\begin{aligned}\|\tilde{\mathbf{u}} - \mathbf{u}\|_{X_l} &\le \|\mathbf{M}^{\mathrm{TG}}\|^3_{\mathrm{Lin}[X_l]}\|\mathbf{u}\|_{X_l} \preccurlyeq \left(n^{-\mu} + \delta\right)\|\mathbf{u}\|_{X_l} \\ &\preccurlyeq \left(n^{-\mu} + \delta\right)\|\mathbf{f}\|_{X_l},\end{aligned}$$

the latter following from part 1 of Lemma 8.3. □

We are now ready to state and prove the main result of this section.

Theorem 8.1. *There exist positive constants M_7 and M_8 such that for any $n \in \mathbb{Z}^+$ and $\delta \in [0, \delta_0]$, the full multigrid algorithm $\check{\phi}_{n,\delta}$ satisfies*

$$e(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \le M_8(n^{-1/\mu} + \delta),$$

with

$$\mathrm{cost}(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \le M_7 c(\delta)\, n.$$

Proof. For $[f, h] \in F$, let $u_{n,\delta} = \phi_{n,\delta}\big(\mathbb{N}_{n,\delta}([f, k])\big)$ and $\check{u}_{n,\delta} = \check{\phi}_{n,\delta}(\check{\mathbb{N}}_{n,\delta}([f, k]))$. By Theorem 7.1 and Lemma 8.2, it suffices to show that

$$\|u_{n,\delta} - \check{u}_{n,\delta}\|_{L_p(I^d)} \preccurlyeq (n^{-1/\mu} + \delta)\|f\|_{L_p(I^d)}. \tag{42}$$

Now $u_{n,\delta} = P_h\mathbf{u}$, where $\mathbf{u}$ is the exact solution of the linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$ given by (15), and $\check{u}_{n,\delta} = P_h\tilde{\mathbf{u}}$, where $\tilde{\mathbf{u}} = \mathrm{TG}(n, \mathbf{A}, \mathbf{B}, \mathbf{f})$. Using (34) along with Lemma 8.4, we obtain

$$\|u_{n,\delta} - \check{u}_{n,\delta}\|_{L_p(I^d)} \le C_{\mathrm{P}}\|\mathbf{u} - \tilde{\mathbf{u}}\|_{X_l} \preccurlyeq (n^{-\mu+\delta})\|\mathbf{f}\|_{X_l}.$$

Hence (42) holds if

$$\|\mathbf{f}\|_{X_l} \preccurlyeq \|f\|_{L_p(I^d)}. \tag{43}$$

For $i \in \{1, \dots, l\}$, define

$$\varepsilon_i = \langle f, s_{i,h}\rangle - f_{h,\delta}(s_{i,h}).$$

Let

$$\mathbf{e} = [\varepsilon_1, \dots, \varepsilon_l]^T$$

and

$$\mathbf{f}^* = [\langle f, s_{1,h}\rangle, \dots, \langle f, s_{l,h}\rangle]^T.$$

Then

$$\|\mathbf{f}\|_{X_l} \le \|\mathbf{e}\|_{X_l} + \|\mathbf{f}^*\|_{X_l}.$$

Since

$$\|\mathbf{e}\|_{X_l} = \|\mathbf{e}\|_{\ell_p(\mathbb{R}^l)} \le l^{1/p'}\|\mathbf{e}\|_{\ell_\infty(\mathbb{R}^l)}$$

and

$$|\varepsilon_i| \preccurlyeq (n^{-\mu}+\delta)\|f\|_{L_p(I^d)}\|s_{i,h}\|_{L_{p'}(I^d)} \preccurlyeq (n^{-\mu}+\delta)l^{-1/p'}\|f\|_{L_p(I^d)},$$

we see that

$$\|\mathbf{e}\|_{X_l} \preccurlyeq (n^{-\mu}+\delta)\|f\|_{L_p(I^d)}. \tag{44}$$

On the other hand, we have $P_h\mathbf{f}^* = \Pi_h f$, so that $\mathbf{f}^* = R_h P_h \mathbf{f}^* = R_h \Pi_h f$ by (35). From (35) and Lemma 4.1, we obtain

$$\|\mathbf{f}^*\|_{X_l} \le C_{\mathrm{R}}\|\Pi_h f\|_{L_p(I^d)} \le C_{\mathrm{R}}(1+M_1)\|f\|_{L_p(I^d)}.$$

Using this inequality and (44), we obtain our desired result (43), which completes the proof of the theorem. □

9 Complexity

In this section, we determine the ε-complexity of the noisy Fredholm problem. We recall from (2) that

$$\mu = \min\left\{\frac{r}{d}, \frac{s}{2d}\right\}.$$

Our main result is

Theorem 9.1. *Let $\varepsilon > 0$. There exist positive numbers C_1, C_2, and C_3, depending only on the global parameters of the problem but independent of ε, such that the following hold:*

1. *The problem complexity is bounded from below by*

$$\mathrm{comp}(\varepsilon) \ge \inf_{0<\delta<C_1\varepsilon} c(\delta)\left\lceil\left(\frac{1}{C_1\varepsilon-\delta}\right)^{1/\mu}\right\rceil.$$

2. *The problem complexity is bounded from above by*

$$\mathrm{comp}(\varepsilon) \le C_2 \inf_{0<\delta<C_3\varepsilon} c(\delta)\left\lceil\left(\frac{1}{C_3\varepsilon-\delta}\right)^{1/\mu}\right\rceil. \tag{45}$$

The upper bound is attained by using the noisy MFEM $\check{\phi}_{n,\delta}$ *using information* $\check{\mathbb{N}}_{n,\delta}$, *where*

$$n = \left\lceil \left(\frac{1}{C_3 \varepsilon - \delta} \right)^{1/\mu} \right\rceil, \tag{46}$$

with $C_3 = M_8^{-1}$ *from Theorem* 8.1 *and where* δ *is chosen to minimize the appropriate right hand side appearing in* (45).

Proof. To prove the lower bound, suppose that ϕ is an algorithm using noisy information $\mathbb{N}_\delta$ such that $e(\phi, \mathbb{N}_\delta) \le \varepsilon$. Then $\operatorname{card} \mathbb{N}_\delta \ge n$, where n must be large enough to make $r_n(\delta) \le \varepsilon$. Theorem 3.1 immediately tells us that we must choose $\delta < M_0^{-1}\varepsilon$ and that we must have

$$n \ge \left\lceil \left(\frac{1}{M_0^{-1}\varepsilon - \delta} \right)^{1/\mu} \right\rceil.$$

The cost of any algorithm using n information evaluations must be at least $n\, c(\delta)$, and so

$$\operatorname{cost}(\phi, \mathbb{N}_\delta) \ge c(\delta) \left\lceil \left(\frac{1}{M_0^{-1}\varepsilon - \delta} \right)^{1/\mu} \right\rceil.$$

Since ϕ and $\mathbb{N}_\delta$ are an arbitrary algorithm and information such that $e(\phi, \mathbb{N}_\delta) \le \varepsilon$, we find that

$$\operatorname{comp}(\varepsilon) \ge c(\delta) \left\lceil \left(\frac{1}{M_0^{-1}\varepsilon - \delta} \right)^{1/\mu} \right\rceil.$$

Finally, since $\delta > 0$ is arbitrary, we get the desired lower bound with $C_1 = M_0^{-1}$.

To prove the upper bound, let $\delta > 0$. If (46) holds, then we may use Theorem 8.1 to see that $e(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \le \varepsilon$. Moreover, we have

$$\operatorname{cost}(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta})(\varepsilon) \le M_7 c(\delta) \left\lceil \left(\frac{1}{M_8^{-1}\varepsilon - \delta} \right)^{1/\mu} \right\rceil,$$

Set $C_2 = M_7$ and $C_3 = M_8^{-1}$. Choosing δ minimizing the right-hand side in these inequalities, the desired result follows. ☐

The lower and upper bounds in Theorem 9.1 are very tight. For an error level ε and a constant C, define the function $g_{\varepsilon,C} \colon \mathbb{R}^{++} \to \mathbb{R}^{++}$ as

$$g_{\varepsilon,C}(\delta) = c(\delta) \left(\frac{1}{C\varepsilon - \delta} \right)^{1/\mu} \qquad \forall\, \delta > 0,$$

and set

$$g^*_{\varepsilon,C} = \inf_{0<\delta<C\varepsilon} g_{\varepsilon,C}(\delta).$$

By Theorem 9.1, we see that

$$g^*_{\varepsilon,C_1} \le \mathrm{comp}(\varepsilon) \le C_2\, g^*_{\varepsilon,C_2}.$$

This inequality allows us to determine the complexity for various cost functions $c(\cdot)$. In particular, if the cost function $c(\cdot)$ is differentiable, then the optimal δ must satisfy $g'_{\varepsilon,C}(\delta) = 0$, i.e., we must have

$$-\frac{c(\delta)}{c'(\delta)} = \mu(C\varepsilon - \delta). \tag{47}$$

As a specific example, consider the cost function $c(\delta) = \delta^{-t}$, where $t > 0$. We find that for $\varepsilon > 0$, the optimal δ is

$$\delta^* = \frac{C\mu t\varepsilon}{\mu t + 1}, \tag{48}$$

so that

$$g^*_{e,C} \asymp \left(\frac{1+\mu t}{C}\right)^{t+1/\mu} \left(\frac{1}{\mu t}\right)^{t} \left(\frac{1}{\varepsilon}\right)^{t+1/\mu}.$$

Thus we see that the optimal δ^* is proportional to ε, and that

$$\mathrm{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{t+1/\mu}.$$

Acknowledgments

I would like to thank W. Hackbusch, S. Heinrich, E. Novak, R. Israel, and S. Pereverzev for their various suggestions. I would also like to thank H. Woźniakowski for his careful reading of an early draft of this paper.

References

[1] R. A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[2] N. S. Bakhvalov. On approximate calculation of integrals. *Vestnik MGU, Ser. Mat. Mekh. Astron. Fiz. Khim.*, 4:3–18, 1959. (In Russian.).

[3] P. G. Ciarlet. *The Finite Element Method For Elliptic Problems*. North-Holland, Amsterdam, 1978.

[4] N. Dunford and J. T. Schwartz. *Linear Operators. Part I: General Theory*, volume 7 of *Pure and Applied Mathematics*. Interscience, New York, 1958.

[5] K. V. Emelyanov and A. M. Ilin. Number of arithmetic operations necessary for the approximate solution of Fredholm integral equations. *USSR Comp. Math. Math. Phys.*, 7(4):259–267, 1967.

[6] K. Frank, S. Heinrich, and S. Pereverzev. Information complexity of multivariate Fredholm integral equations in Sobolev classes. *J. Complexity*, 12(1):17–34, 1996.

[7] W. Hackbusch. *Integral Equations: Theory and Numerical Treatment*, volume 120 of *International Series of Numerical Mathematics*. Birkhäuser, Basel, 1995.

[8] T. Jiang. The worst case complexity of the Fredholm equation of the second kind with non-periodic free term and noise information. *Numer. Funct. Anal. Optim.*, 19(3–4):329–343, 1998.

[9] E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 1988.

[10] S. V. Pereverzev. Complexity of the Fredholm problem of second kind. In *Optimal Recovery*, pages 255–272. Nova Science, New York, 1992.

[11] L. Plaskota. *Noisy Information and Computational Complexity*. Cambridge University Press, Cambridge, 1996.

[12] L. Plaskota. Worst case complexity of problems with random information noise. *J. Complexity*, 12:416–439, 1996.

[13] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Academic Press, New York, 1988.

[14] A. G. Werschulz. What is the complexity of the Fredholm problem of the second kind? *Journal of Integral Equations*, 9:213–241, 1985.

[15] A. G. Werschulz. *The Computational Complexity of Differential and Integral Equations: An Information-Based Approach*. Oxford University Press, New York, 1991.

[16] A. G. Werschulz. Where does smoothness count the most for two-point boundary-value problems? *J. Complexity*, 15:360–384, 1999.