The Complexity of Fredholm Equations of the Second Kind: Noisy Information About Everything

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-003-04

January 21, 2004

Abstract

We study the complexity of Fredholm problems of the second kind $u - \int_\Omega k(\cdot, y)u(y)\,dy = f$. Previous work on the complexity of this problem has assumed that Ω was the unit cube I^d. In this paper, we allow Ω to be part of the data specifying an instance of the problem, along with k and f. More precisely, we assume that Ω is the diffeomorphic image of the unit d-cube under a C^{r_1} mapping $\rho\colon I^d \to I^l$. In addition, we assume that $k \in C^{r_2}(I^{2l})$ and $f \in W^{r_3,p}(I^l)$ with $r_3 > l/p$. Our information about the problem data is contaminated by δ-bounded noise. Error is measured in the L_p-sense. We find that the nth minimal error is bounded from below by $\Theta(n^{-\mu_1} + \delta)$ and from above by $\Theta(n^{-\mu_2} + \delta)$, where

$$\mu_1 = \min\left\{\frac{r_1}{d}, \frac{r_2}{2d}, \frac{r_3 - (d-l)/p}{d}\right\} \qquad \text{and} \qquad \mu_2 = \min\left\{\frac{r_1 - \nu}{d}, \frac{r_2}{2d}, \frac{r_3 - (l-d)/p}{d}\right\},$$

with

$$\nu = \begin{cases} \dfrac{d}{p} & \text{if } r_1 \geq 2,\ r_2 \geq 2, \text{ and } d \leq p, \\ 1 & \text{otherwise.} \end{cases}$$

In particular, the nth minimal error is proportional to $\Theta(n^{-\mu_1} + \delta)$ when $p = \infty$. The upper bound is attained by a noisy modified Galerkin method, which can be efficiently implemented using multigrid techniques. We thus find bounds on the ε-complexity of the problem, these bounds depending on the cost $\mathbf{c}(\delta)$ of calculating a δ-noisy function value. As an example, if $\mathbf{c}(\delta) = \delta^{-b}$, we find that the ε-complexity is between $(1/\varepsilon)^{b+1/\mu_1}$ and $(1/\varepsilon)^{b+1/\mu_2}$.

1 Introduction

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

$$u(s) - \int_\Omega k(s,t)v(t)\,dt = f(s) \qquad \forall\, s \in \Omega. \tag{1.1}$$

*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.

Previous work on the complexity of this problem has dealt with the case where the domain Ω of the integral equation has been the unit cube I^d. Moreover, most of this work has either assumed that we have had complete information about k, or that k and f have had the same smoothness (see, e.g., [8], [9], [11], [13], [18], [19, Sec. 6.3], and the references contained therein). Furthermore, most of the work (with the exception of [11] and a few papers referenced therein) has assumed that the information was exact.

In [20], we studied the complexity of this problem under the assumption that we had noisy standard information about the kernel k and the right-hand side f, with k and f having different smoothness. This lifted many of the restrictions in the previous studies of this problem. However, [20] still assumed that the problem was being solved over the unit cube.

Clearly, the assumption $\Omega = I^d$ is exceptionally restrictive. We need to be able to solve Fredholm problems over whatever domains they naturally arise. Examples include the following:

- The solution of Poisson's equation can be written in terms of integral equations involving single layer potentials, see (e.g.) [6, pg. 390] and [10, Chap. 8].
- The solution of the exterior Helmholtz problem (which arises in scattering theory) can expressed in terms of the solution of a Fredholm problem, see [2].

Note that the integral equations arising in these examples need to be solved over whatever domain the particular problem is defined, and not merely (say) a cube. For problems defined over boundaries of regions (such as the examples given above), the domain in question is a d-dimensional subset of $\mathbb{R}^{d+1}$. This motivates our interest in solving Fredholm problems over general d-dimensional subsets of $\mathbb{R}^l$, where $d \le l$.

In this paper, we study the worst case complexity of Fredholm problems, assuming that we have noisy standard information about all the elements that prescribe our problem. Roughly speaking, this means the following:

1. Error is measured in the L_p-sense, for some $p \in [1, \infty]$.
2. The domain Ω is the image $\rho(I^d)$ of the unit cube under an injection $\rho \in C^{r_1}(I^d; I^l)$. Hence Ω is a subregion of I^l when $d = l$, whereas Ω is a d-dimensional surface in I^l if $d < l$.
3. The kernel k belongs to a ball of $C^{r_2}(I^{2l})$. Moreover, the operator appearing on the left-hand side of (1.1) is invertible, with all such operators satisfying a "uniform invertibility" condition.
4. The right-hand side f belongs to the unit ball of $W^{r_3,p}(I^l)$, with the Sobolev embedding condition $r_3 > l/p$ holding.
5. Only δ-noisy standard information (i.e., noisy function values) is available about the functions determining a particular problem instance.

See Section 2 for the full details.

We are able to determine bounds on the nth minimal radius $r(n, \delta)$ of δ-noisy information, i.e., the minimal error when we use n evaluations with a noise level of δ. In Section 3, we establish the following lower bounds:[1]

1. Let $d < l$ and $r_1 = 1$. Then
$$r(n, \delta) \asymp 1.$$

[1] In this paper, we use $\preccurlyeq$, $\succcurlyeq$, and $\asymp$ to denote O-, Ω-, and Θ-relations. Here, all proportionality factors are independent of n and δ.

2. Let $d = l$ or $r_1 \geq 2$. Then

$$r(n, \delta) \succcurlyeq \left(\frac{1}{n}\right)^{\mu_1} + \delta,$$

where

$$\mu_1 = \min\left\{\frac{r_1}{d}, \frac{r_2}{2d}, \frac{r_3 - (l - d)/p}{d}\right\}. \tag{1.2}$$

Note that the problem is *unsolvable* if $d < l$ and $r_1 = 1$, i.e., we cannot make the error arbitrarily small using finitely many noisy evaluations, no matter how small the noise level nor how large the number of evaluations. Hence, we the problem is solvable only if $d = l$ or if $r_1 = 1$.

Next, we seek upper bounds on the nth minimal noisy error. These bounds are given by a noisy Galerkin method, described in Section 4. This method uses two meshsizes $\bar{h}$ and h, for approximating the Fredholm kernel k and the right-hand side f (respectively). In Section 5, we analyze the error of this method in terms of h, $\bar{h}$, and δ. Then in Section 6, we show how to choose h and $\bar{h}$ minimizing, for a given number n of δ-noisy function evaluations, the upper bound on the error of the noisy Galerkin method. We find that if $d = l$ or $r_1 \geq 1$, then

$$r(n, \delta) \preccurlyeq \left(\frac{1}{n}\right)^{\mu_2} + \delta,$$

where

$$\mu_2 = \min\left\{\frac{r_1 - \nu}{d}, \frac{r_2}{2d}, \frac{r_3 - (l - d)/p}{d}\right\}, \tag{1.3}$$

with

$$\nu = \begin{cases} \dfrac{d}{p} & \text{if } r_1 \geq 2,\ r_2 \geq 2, \text{ and } d \leq p, \\ \\ 1 & \text{otherwise}. \end{cases} \tag{1.4}$$

When do we have tight bounds on the minimal error? Since the problem is unsolvable if $d < l$ and $r_1 = 1$, we can restrict our attention to the case where $d = l$ or $r_1 \geq 2$. Our lower and upper bounds match, yielding

$$r(n, \delta) \asymp \left(\frac{1}{n}\right)^{\mu} + \delta, \tag{1.5}$$

in the following cases:

- If

  $$\min\left\{\frac{r_2}{2d}, \frac{r_3 - (l - d)/p}{d}\right\} \leq \frac{r_1 - \nu}{d}$$

  then (1.5) holds with

  $$\mu = \min\left\{\frac{r_2}{2d}, \frac{r_3 - (l - d)/p}{d}\right\}.$$

- If $r_1 \geq 2$, $r_2 \geq 2$, and $p = \infty$, then $\nu = 0$, and so (1.5) holds with

  $$\mu = \mu_1 = \min\left\{\frac{r_1}{d}, \frac{r_2}{2d}, \frac{r_3}{d}\right\}.$$

However, our upper and lower bounds are not always tight. For an especially appalling case, suppose that $d = l$ and $r_1 = 1$. Then the upper bound on the minimal error does *not* converge to zero as $n \to \infty$, whereas the lower bound *does* converge to zero as $n \to \infty$, and so we don't even know whether the problem is convergent when $d = l$ and $r_1 = 1$. The task of determining tight bounds on the minimal error in the remaining cases is currently an open problem.

Let us discuss the cost of the noisy Galerkin method. Let $\mathbf{c}(\delta)$ denote the cost of evaluating a function with a noise level δ. Then the information cost of this algorithm is $\mathbf{c}(\delta)\, n$. However, since this algorithm involves the solution of a full linear system of equations, the combinatory cost is much worse than $\Theta(n)$. As in [20], we overcome this difficulty by using a two-grid implementation of the noisy Galerkin method. This algorithm has the same order of error as the original noisy Galerkin , and its combinatory cost is $\Theta(n)$. Hence, we can calculate the two-grid approximation using $\Theta(n)$ arithmetic operations, which is optimal. The details are given in Section 7.

We use these results in Section 8 to determine bounds on the ε-complexity of the Fredholm problem. First, suppose that $d < l$ and $r_1 = 1$. Since the nth minimal radius is bounded away from zero, there exists $\varepsilon_0 > 0$ such that $\text{comp}(\varepsilon) = \infty$ for $0 \le \varepsilon \le \varepsilon_0$. So, we consider the case where $d = l$ or $r_1 \ge 2$. We find that there exist positive constants C_1, C_2, and C_3, independent of ε, such that the problem complexity is bounded from below by

$$\text{comp}(\varepsilon) \ge \inf_{0<\delta<C_1\varepsilon} \left\{ c(\delta) \left\lceil \left(\frac{1}{C_1\varepsilon - \delta} \right)^{1/\mu_1} \right\rceil \right\}$$

and from above by

$$\text{comp}(\varepsilon) \le C_2 \inf_{0<\delta<C_3\varepsilon} \left\{ c(\delta) \left\lceil \left(\frac{1}{C_3\varepsilon - \delta} \right)^{1/\mu_2} \right\rceil \right\}.$$

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

In particular, suppose that $\mathbf{c}(\delta) = \delta^{-b}$ for some $b > 0$. We find that

$$\left(\frac{1}{\varepsilon}\right)^{b+1/\mu_1} \preccurlyeq \text{comp}(\varepsilon) \preccurlyeq \left(\frac{1}{\varepsilon}\right)^{b+1/\mu_2},$$

Note that when $\mu_1 = \mu_2 = \mu$, we have tight bounds

$$\text{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{b+1/\mu}$$

on the ε-complexity. This holds (for example) when $r_1 \ge 2$, $r_2 \ge 2$, and $p = \infty$.

We close this Introduction by mentioning future extensions of this work.

1. One glaring problem is that the lower and upper bounds are not always tight; we hope to remedy this problem in the near future.

2. The other major issue to raise is that not all domains of interest are images of cubes. For example, a smooth region (such as a ball or sphere) is not the diffeomorphic image of a cube. One way of getting around this difficulty is to consider domains that are images of balls. This approach was studied (for the surface approximation and integration problems) in [21, Sect. 5]; it appears that the results of this paper also apply to the case where the domain is the image of a ball, the main difference being a

slight extra complication appearing in the definitions of certain integrals that will appear in the noisy Galerkin method.

Another idea is to use oriented cellulated regions (OCRs) [7, pp. 369-370], which essentially means that the domains are finite unions of images of cubes. Clearly, OCRs can represent a wide variety of domains. We hope to analyze the complexity of Fredholm problems over OCRs in a future paper.

2 Problem description

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

For an ordered ring $\mathscr{X}$, we shall let $\mathscr{X}^+$ and $\mathscr{X}^{++}$ respectively denote the non-negative and positive elements of $\mathscr{X}$. 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 $\mathscr{Y}$, we let $\mathscr{B}\mathscr{Y}$ denote the unit ball of $\mathscr{Y}$. We assume that the reader is familiar with the standard concepts and notations involving Sobolev norms and spaces, as found in, e.g., [4].

As in [21], we shall deal only with nondegenerate domains that are bijective images of I^d (see Figure 1), the nondegeneracy meaning that the Jacobian associated with the domain never vanishes.

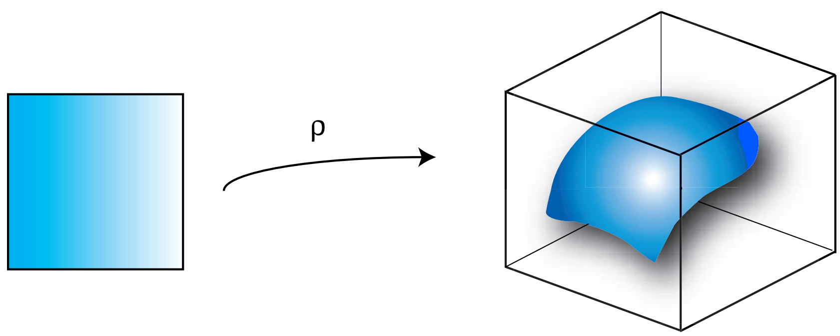

Figure 1: An admissible domain

More precisely, let $\rho\colon I^d \to I^l$ be a continuously differentiable injection, so that $d \le l$ must necessarily hold. The *gradient* of ρ at $x \in I^d$ is

$$(\nabla\rho)(x) = \left[\frac{\partial\rho_i}{\partial x_j}(x)\right]_{1\le i\le l, 1\le j\le d} \in \mathbb{R}^{l\times d},$$

where $\rho_1, \dots, \rho_l$ are the components of ρ. Then the *Jacobian* of $\rho(I^d)$ at $x \in I^d$ is defined to be

$$J(x;\rho) = \sqrt{\det A(x;\rho)},$$

where

$$A(x;\rho) = [(\nabla\rho)(x)]^{\mathsf{T}}[(\nabla\rho)(x)] \in \mathbb{R}^{d\times d}.$$

The region $\rho(I^d)$ is *nondegenerate* if $J(x;\rho) \ne 0$ for all $x \in I^d$. For a nondegenerate region, we have the change of variables formula

$$\int_{\Omega_\rho} v(t)\,dt = \int_{I^d} v\big(\rho(x)\big) J(x;\rho)\,dt, \tag{2.1}$$

and so $J_\rho(x;\rho)\,dx$ is the volume element (if $d = l$) or surface area element (if $d < l$) of Ω_ρ at $x \in I^d$. See [7, p. 334 ff.] for further discussion.

Given such ρ, let $\Omega_\rho = \rho(I^d)$, and suppose that u is the solution of the Fredholm problem (1.1) over the domain Ω_ρ, i.e.,

$$u(s) - \int_{\Omega_\rho} k(s,t)u(t)\,dt = f(s) \qquad \forall\, s \in \Omega_\rho.$$

Writing $s = \rho(x)$ and using the change of variables formula (2.1), this may be rewritten

$$u\bigl(\rho(x)\bigr) - \int_{I^d} k\bigl(\rho(x), \rho(y)\bigr)u\bigl(\rho(y)\bigr)J(y;\rho)\,dy = f\bigl(\rho(x)\bigr) \qquad \forall\, x \in I^d \tag{2.2}$$

as a problem over I^d.

It will be convenient to write (2.2) as an operator equation. Define

$$k_\rho(x,y) = k\bigl(\rho(x), \rho(y)\bigr) \qquad \forall\, x, y \in I^d.$$

Moreover, for $v\colon I^d \to \mathbb{R}$ and $g\colon I^{2d} \to \mathbb{R}$, let us write

$$T_{\rho,g}v = \int_{I^d} g(\cdot, y)v(y)J(y;\rho)\,dy$$
$$v_\rho = v \circ \rho$$

Then we may rewrite (2.2) in the form

$$(I - T_{\rho,k_\rho})u_\rho = f_\rho. \tag{2.3}$$

We are now ready to describe the admissible problem elements $[\rho, k, f]$. We begin with the class $\mathscr{R}$ of functions $\rho\colon I^d \to I^l$ determining the domains $\Omega_\rho = \rho(I^d)$. Let positive numbers c_1 and c_2 be given, along with $r_1 \geq 1$. Then $\mathscr{R}$ consists of the functions $\rho \in C^{r_1}(I^d; I^d)$ that satisfy the condition

$$\|\rho\|_{C^{r_1}(I^d;I^l)} \leq c_1,$$

where

$$\|\rho\|_{C^{r_1}(I^d;I^l)} = \max_{1\leq i\leq d} \|\rho_i\|_{C^{r_1}(I^d)},$$

as well as the "uniform nondegeneracy condition"

$$\min_{x\in I^d} J_\rho(x) \geq c_2.$$

For simplicity, we shall assume that $c_1 < 1 \leq c_2$ in this paper.

Remark. The mapping $\mathrm{id}\colon I^d \to I^l$, defined as

$$\mathrm{id}(x) = (x, \underbrace{0, \ldots, 0}_{l-d\text{ zeros}}) \qquad \forall\, x \in I^d,$$

belongs to $\mathscr{R}$. □

Remark. Why do we require $\rho(I^d) \subseteq I^l$? Under this condition, any $k\colon I^{2l} \to \mathbb{R}$ will be defined on $\Omega_\rho \times \Omega_\rho$, and any $f\colon I^l \to \mathbb{R}$ will be defined on Ω_ρ. Thus any such k and f will be allowable in our integral equation (1.1). Had we not imposed this condition on ρ, we would have needed to impose more complicated conditions on our k and f than those stated below. □

Remark. The conditions defining $\mathscr{R}$ imply that we have an a priori bound on the volume or surface area element of Ω_ρ, which is independent of ρ, namely

$$J(\cdot;\rho) \le \kappa_{d,l} = \begin{cases} 1 & \text{if } d = l, \\ \sqrt{d!\, l^d} c_1^d & \text{if } d < l. \end{cases} \qquad \forall\, \rho \in \mathscr{R}. \tag{2.4}$$

Indeed, the bound for the case $d = l$ follows from the fact that the codomain of ρ is the unit cube, whereas a *very* rough calculation shows the bound for the case $d < l$. Hence for any $\rho \in \mathscr{R}$, the volume (or surface area) of Ω_ρ, which is merely $\|J(\cdot;\rho)\|_{L_1(I^d)}$, is at most $\kappa_{d,l}$. □

Next, we describe our class $\mathscr{K}$ of kernels k. Let $c_3 > 0$ and $c_4 > 1$ be given, along with $r_2 \ge 0$ and $p \in [1,\infty]$. Then $\mathscr{K}$ consists of the functions $k \in C^{r_2}(I^{2l})$ satisfying

$$\|k\|_{C^{r_2}(I^{2l})} \le c_3,$$

for which the "uniform invertibility condition"

$$\|(I - T_{\rho,k_\rho})^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \le c_4 \qquad \forall\, \rho \in \mathscr{R}$$

holds. Here, $\|\cdot\|_{\mathrm{Lin}[\mathscr{Y}]}$ is the usual operator norm.

Remark. Suppose that

$$c_3 < 1 \qquad \text{and} \qquad c_4 > \frac{1}{1 - c_3}.$$

If $k \in C^{r_2}(I^{2l})$ with

$$\|k\|_{C^{r_2}(I^{2l})} \le \frac{c_3}{\kappa_{d,l}},$$

where $\kappa_{d,l}$ is given by (2.4), then $k \in \mathscr{K}$. This follows from the easily-proven fact that $T_{\rho,k_\rho} \in L_p(I^d)$, with

$$\|T_{\rho,k_\rho}\|_{\mathrm{Lin}\, L_p(I^d)} \le \|k_\rho\|_{C(I^{2d})} \|J(\cdot,\rho)\|_{L_{p'}(I^d)}, \tag{2.5}$$

where

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

is the exponent conjugate to p, along with the Neumann series for $(I - T_{\rho,k_\rho})^{-1}$. □

Our class of right-hand sides will be $\mathscr{B}W^{r_3,p}(I^l)$, where we will require $r_3 > l/p$, so that the Sobolev embedding theorem will hold. Hence our class of problem elements will be

$$\mathscr{F} = \mathscr{R} \times \mathscr{K} \times \mathscr{B}W^{r_3,p}(I^l).$$

Now we can define our solution operator $S\colon \mathscr{F} \to L_p(I^d)$ as

$$S([\rho,k,f]) = (I - T_{\rho,k_\rho})^{-1} f_\rho \qquad \forall\, [\rho,k,f] \in \mathscr{F}.$$

Hence $u_\rho = S([\rho,k,f])$ is the solution of the operator equation (2.3).

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 [14], although we sometimes use modifications found in [15].

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

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

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

$$|z_i - \rho(x^{(i)})| \le \delta \text{ for some } x^{(i)} \in I^d,$$

or

$$\left| z_i - k\big(\rho(x^{(i)}), \rho(y^{(i)})\big) \right| \le \delta \text{ for some } (x^{(i)}, y^{(i)}) \in I^{2d}, \tag{2.6}$$

or

$$\left| z_i - f\big(\rho(x^{(i)})\big) \right| \le \delta \text{ for some } x^{(i)} \in I^d. \tag{2.7}$$

The choice of whether to evaluate ρ, k_ρ or f_ρ 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.

Remark. The reader may have expected that instead of using the "noisy composite information" (2.6)–(2.7), that we would use the simpler noisy information

$$|z_i - k(s_i, t_i)| \le \delta \text{ for some } s_i, t_i \in I^l$$

about the kernel and

$$|z_i - f(s_i)| \le \delta \text{ for some } s_i \in I^l$$

about the right-hand side. The main reason for using noisy composite information is that the algorithm that gives us our upper bounds uses this information. However, there is essentially no loss of generality in using noisy composite information instead of the the simpler noisy information, since

1. any lower bound on the error of algorithms using noisy composite information is a lower on algorithms using the simpler noisy information, and

2. if k and f satisfy a Lipschitz condition, then the simpler noisy information is also noisy composite information (albeit with a different value of δ involving the Lipschitz constant).

Since the definition of $\mathscr{F}$ is already fairly complicated, we prefer using noisy composite information to imposing additional conditions on $\mathscr{F}$. □

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

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

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

Remark. Note that the permissible information consists of noisy function values of k_ρ and f_ρ. 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. □

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

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

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

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

Remark. Rather than measuring the error in approximating the composite function $u_\rho = S([\rho, k, f])$ in $L_p(I^d)$, one might maintain that it is more natural to directly approximate u itself in $L_p(\Omega_\rho)$. It turns out that these tasks are essentially equivalent. More precisely, let ϕ be an algorithm using $\mathbb{N}_\delta$. For $z \in \mathbb{N}_\delta([\rho, k, f])$, let $\psi(z) = \phi(z) \circ \rho^{-1}$. Then

$$c_2^{1/p} \| u_\rho - \phi(z) \|_{L_p(I^d)} \le \| u - \psi(z) \|_{L_p(\Omega_\rho)} \le c_1^{1/p} \| u_\rho - \phi(z) \|_{L_p(I^d)}.$$

(This follows easily from the conditions defining $\mathscr{R}$.) Although the direct approximation of u may be more natural than approximating the composite function u_ρ, the technical details for handling the latter are simpler. $\square$

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*

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

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 [14, 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 [14].) Here are the most important features of this model:

1. The cost of calculating a δ-noisy function evaluation is $\mathbf{c}(\delta)$.

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

[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.

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

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

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

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

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

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

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

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 radius of δ-noisy information. One tool for doing this is to show that some other problem is a special case of our Fredholm problem, whence the minimal radius of this other problem is a lower bound on that of our problem. Hence, we will sometimes need to discuss the nth minimal radius of δ-noisy information for a problem given by another solution operator. This means that it will sometimes be necessary to explicitly show how the minimal radius depends on the solution operator.

Theorem 3.1.

1. *If $d < l$ and $r_1 = 1$, then*
$$r(n, \delta) \asymp 1.$$

2. *If $d = l$ or $r_1 \geq 2$, let μ_1 be defined as in* (1.2). *There is a constant M_0, independent of n and δ, such that*
$$r(n, \delta) \geq M_0(n^{-\mu_1} + \delta)$$
for all $n \in \mathbb{Z}^+$ and $\delta \in [0, 1]$.

Proof. We first consider the case $d < l$ and $r_1 = 1$. Let

$$\rho^*(x) = (0, x_2, \dots, x_d, x_1, \underbrace{0, \dots, 0}_{l-d-1 \text{ zeros}}) \qquad \forall\, x = (x_1, \dots, x_d) \in I^d,$$

and define $k^* \equiv \frac{1}{2}$ and $f^* \equiv 1$. Since $J(\cdot, \rho^*) \equiv 1$, it follows that $[\rho^*, k^*, f^*] \in \mathscr{F}$. Moreover, $u^*_{\rho^*} = S([\rho^*, k^*, f^*])$ satisfies

$$u^*\big(\rho^*(x)\big) = \tfrac{1}{2} \int_{I^d} u^*\big(\rho^*(y)\big) J(y, \rho^*)\, dy + 1$$

and so

$$u^*\big(\rho^*(x)\big) \equiv \frac{1}{1 - \frac{1}{2}\operatorname{area}(\Omega_{\rho^*})} = 2.$$

Let N be noise-free information of cardinality at most n. Without loss of generality, assume that the ρ-evaluation points in $N([\rho^*, k^*, f^*])$ are $x^{(1)}, \dots, x^{(n')}$. As on [22, pg. 461], we can construct a function $z \colon I^d \to \mathbb{R}$ such that

$$z(x^{(1)}) = \dots = z(x^{(n')}) = 0$$

and

$$\|z\|_{C^{r_1}(I^d; I^l)} = 1.$$

Let

$$\rho^{**}(x) = \left[\frac{z(x)}{2\sqrt{d}}, x_2, \dots, x_d, x_1, \underbrace{0, \dots, 0}_{l-d-1 \text{ zeros}}\right] \qquad \forall\, x = (x_1, \dots, x_d) \in I^d.$$

We find

$$J(x, \rho^{**}) = \sqrt{1 + \frac{1}{4d}\sum_{j=1}^{d}\left(\frac{\partial z}{\partial x_j}\right)^2(x)} \geq 1,$$

from which it follows that $\rho^{**} \in \mathscr{R}$. Hence $[\rho^{**}, k^*, f^*] \in \mathscr{F}$. Moreover, $u^{**}_{\rho^{**}}$ satisfies

$$u^{**}\big(\rho^{**}(x)\big) = \tfrac{1}{2} \int_{I^d} u^{**}\big(\rho^{**}(y)\big) J(y, \rho^{**})\big)\, dy + 1,$$

and so

$$u^{**}\big(\rho^{**}(x)\big) \equiv \frac{1}{1 - \frac{1}{2}\operatorname{area}(\Omega_{\rho^{**}})}.$$

Using [17, pp. 45, 49], we see that

$$\begin{aligned} r(N) &\geq \tfrac{1}{2}\|u^{**}_{\rho^{**}} - u^*_{\rho^*}\|_{L_p(I^d)} = \tfrac{1}{2}\left|\frac{1}{1 - \frac{1}{2}\operatorname{area}(\Omega_{\rho^{**}})} - \frac{1}{1 - \frac{1}{2}\operatorname{area}(\Omega_{\rho^*})}\right| \\ &= \tfrac{1}{4}\,\frac{|\operatorname{area}(\Omega_{\rho^{**}}) - \operatorname{area}(\Omega_{\rho^*})|}{|1 - \frac{1}{2}\operatorname{area}(\Omega_{\rho^{**}})|}\,. \end{aligned}$$

Now

$$\operatorname{area}(\Omega_{\rho^{**}}) = \int_{I^d} J(x; \rho^{**})\, dx \leq \tfrac{5}{4},$$

and so

$$1 - \tfrac{1}{2}\,\text{area}(\Omega_{\rho^{**}}) \geq \tfrac{3}{8}.$$

Hence

$$r(N) \geq \tfrac{2}{3}|\,\text{area}(\Omega_{\rho^{**}}) - \text{area}(\Omega_{\rho^{*}})| = \tfrac{2}{3}\int_{I^d} \frac{\displaystyle\sum_{j=1}^{d}\left(\frac{\partial z}{\partial x_j}\right)^2(x)}{\sqrt{1 + \displaystyle\frac{1}{4d}\sum_{j=1}^{d}\left(\frac{\partial z}{\partial x_j}\right)^2(x)} + 1}\,dx$$

$$\geq \tfrac{1}{3}\int_{I^d} \sum_{j=1}^{d}\left(\frac{\partial z}{\partial x_j}\right)^2(x)\,dx\ .$$

Following the proof of [22, Thm. 4.3], we now see that

$$r(N) \succcurlyeq 1,$$

from which we see that

$$r(n,\delta) \geq r(n,0) \succcurlyeq 1.$$

To see the matching upper bound, let $\mathbb{N}_\delta$ be noisy information of cardinality at most n, and let ϕ^0 be the zero algorithm

$$\phi^0(z) \equiv 0 \qquad \forall\, z \in \mathscr{Z}(\mathbb{N}_\delta).$$

It is easy to see that the error of ϕ^0 is bounded, independent of n and δ, and so

$$r(n,\delta) \leq e(\phi^0, \mathbb{N}_\delta) \preccurlyeq 1.$$

Thus

$$r(n,\delta) \asymp 1,$$

as claimed.

We now treat the case where $d = l$ or $r_1 \geq 2$. First, we claim that

$$r(n,0) \succcurlyeq n^{-r_1/d}. \tag{3.1}$$

Indeed, let $\rho \in \mathscr{R}$ and define $f^* \in \mathscr{B}W^{r_3,p}(I^l)$ as

$$f^*(s) \equiv s \qquad \forall\, s \in I^l.$$

Since $T_{\rho,0} = 0$, we see that $S([\rho, 0, f^*]) = \rho_1$, the first component of ρ. Define a solution operator $\tilde{S}\colon \mathscr{R} \to L_p(I^d)$ as

$$\tilde{S}(\rho) = S([\rho, 0, f^*]) = \rho_1 \qquad \forall\, \rho \in \mathscr{R}.$$

Since the problem given by this solution operator is a special case of our Fredholm problem, we see that the nth minimal noise-free radius of S is bounded from below by that for $\tilde{S}$, i.e.,

$$r(n, 0; S) \geq r(n, 0; \tilde{S}).$$

Following the proof of [21, Lemma 3.4], we have

$$r(n, 0; \tilde{S}) \succcurlyeq n^{-r_1/d},$$

and so (3.1) holds, as claimed.

We next claim that

$$r(n,\delta) \succcurlyeq \delta. \tag{3.2}$$

Define a solution operator $\tilde{S}\colon \mathscr{B}W^{r_3,p}(I^l) \to \mathbb{R}$ as

$$\tilde{S}(f) = S([\mathrm{id}, 0, f]) = f_{\mathrm{id}} \qquad \forall f \in \mathscr{B}W^{r_3,p}(I^l).$$

Since $\tilde{S}$ is a special case of S, we have

$$r(n,\delta;S) \geq r(n,\delta;\tilde{S}).$$

Replicating the proof of [20, inequality (6)], we find that

$$r(n,\delta;\tilde{S}) \succcurlyeq \delta,$$

and hence (3.2) holds,

We next claim that

$$r(n,0) \succcurlyeq n^{-r_2/2d} \tag{3.3}$$

holds. Our argument based on that found in the proof of the analogous bound in the second part of [20, Thm. 3.1]. Let

$$\theta_1 \in (c_4^{-1}, 1) \qquad \text{and} \qquad k_0 = \min\left\{\theta_1 c_3, 1 - \frac{1}{\theta_1 c_4}\right\}.$$

Define $f^*\colon I^l \to \mathbb{R}$ and $k^*\colon I^{2l} \to \mathbb{R}$ as

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

Clearly $\mathrm{id} \in \mathscr{R}$ and $f^* \in \mathscr{B}W^{r_3,p}(I^l)$. We have

$$\|k^*\|_{C^{r_2}(I^{2l})} = k_0 < c_1.$$

From (2.5), we have

$$\|T_{\rho,k^*_{\mathrm{id}}}\|_{\mathrm{Lin}[L_p(I^d)]} \leq k_0,$$

so that

$$\|(I - T_{\rho,k^*_{\mathrm{id}}})^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \leq \frac{1}{1-k_0} \leq \theta_1 c_2 < c_2.$$

Thus $k^* \in \mathscr{K}$.

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

$$N([\mathrm{id}, k^*, f^*]) = [z_1, \dots, z_{n'}]$$

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

$$z_i = k^*\big(\mathrm{id}(x^{(i)}), \mathrm{id}(y^{(i)})\big) \qquad (1 \leq i \leq n'').$$

From [3] (see also [12, pg. 34]), we can find a function $w \in \mathscr{B}C^{r_2}(I^{2l})$ such that

$$0 \le w(x, y) \le k_0 \qquad \forall\, x, y \in I^d,$$
$$w(x^{(i)}, y^{(i)}) = 0 \qquad (1 \le i \le n''),$$
$$\|w\|_{C^{r_2}(I^{2l})} = 1,$$
$$\int_{I^{2d}} w(x, y)\, dx\, dy \ge \frac{\theta_2}{(n'')^{r_2/2d}},$$

where θ_2 is a positive constant that is independent of the points $(x^{(i)}, y^{(i)})$ and of n''. Let

$$\theta_3 = \min\{(1-\theta_1)c_3, 1 - c_4^{-1} - k_0\},$$

and define

$$k^{**} = k_0 + \theta_3 w \circ \mathrm{id}^{-1},$$

where $\mathrm{id}^{-1} \colon I^l \to I^d$ is the right inverse of id given by

$$\mathrm{id}^{-1}(s) = (s_1, \dots, s_d) \qquad \forall\, s = (s_1, \dots, s_l) \in I^l.$$

As in the proof of [20, Thm. 3.1], we find

$$\|k^{**}\|_{C^{r_2}(I^{2l})} \le \theta_1 c_3 + \theta_3 \le c_3$$

and thus

$$\|(I - T_{k^{**}, k^{**}_{\mathrm{id}}})^{-1}\|_{\mathrm{Lin}[L_p(I^d)]} \le \frac{1}{1-(k_0+\theta_3)} \le c_4.$$

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

Thus we have found $[\mathrm{id}, k^{**}, f^*], [\mathrm{id}, k^*, f^*] \in \mathscr{F}$ such that

$$N([\mathrm{id}, k^{**}, f^*]) = N([\mathrm{id}, k^*, f^*]).$$

From the proof of [20, Thm. 3.1], we have

$$r(N) \ge \tfrac{1}{2}\|S([\mathrm{id}, k^{**}, f^*]) - S([\mathrm{id}, k^*, f^*])\|_{L_p(I^d)} \succcurlyeq (n'')^{-r_2/2d}.$$

Since $n'' \le n$ and N is arbitrary information of cardinality at most n, the desired bound (3.3) holds.

We next claim that

$$r(n, 0) \succcurlyeq n^{-[r_3-(l-d)/p]/d}. \tag{3.4}$$

Define the solution operator $\tilde{S} : \mathscr{B}W^{r_3,p}(I^l) \to L_p(I^d)$ as

$$\tilde{S}(f) = S([\mathrm{id}, 0, f]) \equiv f_{\mathrm{id}} \qquad \forall\, f \in \mathscr{B}W^{r_3,p}(I^l).$$

That is, $\tilde{S}$ is the solution operator corresponding to the $L_p(I^d)$ approximation problem for the unit ball of $W^{r_3,p}(I^l)$. Since $\tilde{S}$ is a special case of S, we have

$$r(n, 0; S) \ge r(n, 0; \tilde{S}),$$

and thus it suffices to show that

$$r(n, 0; \tilde{S}) \succcurlyeq n^{-[r_3-(l-d)/p]/d}. \tag{3.5}$$

Let N be noiseless information of cardinality at most n for $\tilde{S}$. By the results in [17, Sect. 4.5.2], we can assume that N is nonadaptive without loss of generality. Hence, there exists $n' \leq n$ and points $x^{(1)}, \dots, x^{(n')} \in I^d$ such that

$$N(f) = [f(x^{(1)}) \dots f(x^{(n')})] \qquad \forall f \in \mathscr{B}W^{r_3,p}(I^l).$$

Suppose that n'' of these points lie in $\mathrm{id}(I^d)$; without loss of generality, assume that these points are $x^{(1)}, \dots, x^{(n'')}$. It is well-known (see, e.g., [12]) that we can construct $w \in W^{r_3,p}\big(\mathrm{id}(I^d)\big)$ such that

$$\begin{aligned} w(x^{(1)}) = \cdots = w(x^{(n'')}) = 0, \\ \|w\|_{W^{r_3-(l-d)/p,p}(\mathrm{id}(I^d))} = 1, \\ \|w\|_{L_p(I^d)(\mathrm{id}(I^d))} \geq \frac{\theta_3}{(n'')^{[r_3-(l-d)/p]/d}} \end{aligned},$$

where $\theta_3 > 0$ is independent of w and n''. Using the Sobolev trace theorem $l - d$ times, the function w can be extended to all of I^l, with

$$\|w\|_{W^{r_3,p}(I^l)} \leq \theta_4 \|w\|_{W^{r_3-(l-d)/p,p}(\mathrm{id}(I^d))},$$

where $\theta_4 > 0$ is independent of w, see (e.g.) [1, § 7.56]. Now let

$$\tilde{w} = \frac{1}{\theta_4} w.$$

Then

$$\begin{aligned} N(\tilde{w}) = 0, \\ \|\tilde{w}\|_{W^{r_3,p}(I^l)} \leq 1, \\ \|w_{\mathrm{id}}\|_{L_p(I^d)} = \|w\|_{L_p(\mathrm{id}(I^d))} \geq \frac{\theta_3}{(n'')^{[r_3-(l-d)/p]/d}}. \end{aligned}$$

Using [17, Lemma 5.2.1], we see that

$$r(N) \geq \frac{\theta_3}{\theta_4\,(n'')^{[r_3-(l-d)/p]/d}}$$

Since $n'' \leq n$ and N is arbitrary information of cardinality at most n, this establishes (3.5), and hence (3.4).

Combining (3.1)–(3.4), we get

$$r(n,\delta) \succcurlyeq \left(\frac{1}{n}\right)^{\mu_1} + \delta,$$

as required. □

4 The noisy modified Galerkin method

Having established a lower bound on the nth minimal radius for our problem, we now seek an upper bound. Of course, since our problem is unsolvable when $d < l$ and $r_1 = 1$, we shall assume that $d = l$ or $r_1 \geq 2$ in the sequel. Our upper bound will be provided by a modified Galerkin method using noisy standard information. In this section, we describe the method; we analyze its error in the next section.

We first present a weak formulation of our problem. For $[\rho, k] \in \mathscr{R} \times \mathscr{K}$, define a bilinear form $B(\cdot,\cdot;\rho,k_\rho)$ on $L_p(I^d) \times L_{p'}(I^d)$ as

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

For $f \in \mathscr{B}W^{r_3,p}(I^l)$, we see that $u_\rho = S([\rho, k, f]) \in L_p(I^d)$ satisfies

$$B(u_\rho, w; \rho, k_\rho) = \langle f_\rho, w\rangle \qquad \forall\, w \in L_{p'}(I^d).$$

Here, $\langle\cdot,\cdot\rangle$ denotes the standard duality pairing

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

Next, we describe a class of useful spline spaces; for further details, see [22]. Let $m \in \mathbb{Z}^{++}$ (to be determined later) and $h > 0$. Then $\mathscr{S}_h$ denotes a d-fold tensor product of one-dimensional C^1-splines of degree m, over a uniform grid of mesh-size h.

Let $n_h = \dim \mathscr{S}_h$, noting that $n_h \asymp h^{-d}$. Associated with $\mathscr{S}_h$ is a quasi-interpolation operator

$$(Q_h w)(x) = \sum_{j=1}^{n_h} \lambda_{j,h}(w) s_{j,h}(x) \qquad \forall\, x \in I^d,\, w \in C(I^d), \tag{4.1}$$

where each $s_{j,h}$ is a d-fold tensor product of one-dimensional splines and we can write

$$\lambda_{j,h}(w) = \lambda_j\bigl(\{w(x_{i,h})\}_i\bigr) \qquad \forall\, w \in C(I^d)$$

where each $\lambda_j(w)$ can be computed with cost independent of h, once the values $w(x_{1,h}), \ldots w(x_{n_h,h})$ have been computed. Associated with $\mathscr{S}_h$ (for any $q \in [1,\infty]$) is a projection operator $\mathscr{P}_h \colon L_q(I^d) \to \mathscr{S}_h$ defined by

$$\langle \mathscr{P}_h v, w\rangle = \langle v, w\rangle \qquad \forall\, v \in L_q(I^d),\, w \in L_{p'}(I^d).$$

Not only is this projection operator well-defined, but we have the stronger result that

$$\pi_q = \sup_{0<h\le 1} \|\mathscr{P}_h\|_{\mathrm{Lin}[L_q(I^d)]}, \tag{4.2}$$

is finite, see [19, pp. 177–178] and the references cited therein.

We will also have need of a $2d$-variate spline space $\mathscr{S}_{\bar h} \otimes \mathscr{S}_{\bar h}$ involving a (possibly) different mesh-size $\bar h$. The quasi-interpolation operator $Q_{\bar h\otimes\bar h}$ of $\mathscr{S}_{\bar h} \otimes \mathscr{S}_{\bar h}$ takes the form

$$(Q_{\bar h\otimes\bar h} w)(x,y) = \sum_{i,j=1}^{n_{\bar h}} \lambda_{i,j,\bar h}\bigl(\{w(x_{i',\bar h}, x_{j',\bar h})\}_{i',j'}\bigr) s_{j,\bar h}(y) s_{i,\bar h}(x) \qquad \forall x, y \in I^d,\, w \in C(I^{2d}). \tag{4.3}$$

Remark. Note since the maximum continuous differentiability of a degree-m spline is $m-2$, we must have $m \ge 3$ to guarantee that $\mathscr{S}_h$ and $\mathscr{S}_{\bar h} \otimes \mathscr{S}_{\bar h}$ are globally C^1. We also note that $\mathscr{S}_h$ and $\mathscr{S}_{\bar h} \otimes \mathscr{S}_{\bar h}$ are subspaces of $W^{2,\infty}(I^d)$, since $\mathscr{S}_h$ is piecewise polynomial and globally C^1; this follows from the L_∞ version of [4, Thm. 2.1.1]. □

Now that we have a bilinear form and a family of spline spaces, we can define a "pure" Galerkin method. Let $[\rho, k, f] \in \mathscr{F}$ and let $h > 0$. Then the *pure Galerkin method* consists of finding $u_h \in \mathscr{S}_h$ such that

$$B(u_h, w; \rho, k_\rho) = \langle f_\rho, w\rangle \qquad \forall\, w \in \mathscr{S}_h.$$

Alternatively, we seek $u_h \in \mathscr{S}_h$ satisfying

$$(I - \mathscr{P}_h T_{\rho,k_\rho}) u_h = \mathscr{P}_h f_\rho,$$

where $\mathscr{P}_h$ is the projection operator mentioned above. Note that u_h is an approximation of u_ρ, and not of u.

If we write

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

then $\mathbf{u} = [\upsilon_1, \ldots, \upsilon_{n_h}]$ satisfies the linear system

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

where

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

and, for $1 \le i, j \le n_h$, we have

$$a_{i.j} = \langle s_{j,h}, s_{i,h}\rangle \qquad \text{and} \qquad b_{i,j} = \langle T_{\rho,k_\rho} s_{j,h}, s_{i,h}\rangle = \int_{I^{2d}} k\big(\rho(x), \rho(y)\big) s_{j,h}(y) J(y;\rho) s_{i,h}(x)\, dy\, dx\,.$$

Since the pure Galerkin method requires the calculation of weighted integrals involving ρ, k_ρ and f_ρ, and we are only using (noisy) standard information, the pure Galerkin method is not admissible for us. Instead, we shall replace ρ, k_ρ, and f_ρ by their noisy quasi-interpolants (defined below); this will give us an algorithm using permissible information.

Let $h, \bar{h}, \delta > 0$, and let $[\rho, k, f] \in \mathscr{F}$. For $j \in \{1, \ldots, n_h\}$, calculate $\tilde{\rho}_{j;h,\delta}$ satisfying

$$|\tilde{\rho}_{j,h,\delta} - \rho(x_{j,h})| \le \delta$$

and $\tilde{f}_{j,h,\delta}$ satisfying

$$\big|\tilde{f}_{j,h,\delta} - f\big(\rho(x_{j,h})\big)\big| \le \delta.$$

For $i, j \in \{1, \ldots, n_{\bar{h}}\}$, calculate $\tilde{k}_{i,j,\delta}$ satisfying

$$\big|\tilde{k}_{i,j,\delta} - k\big(\rho(x_{i,\bar{h}}), \rho(x_{j,\bar{h}})\big)\big| \le \delta.$$

Define noisy quasi-interpolants of ρ, f_ρ, and k_ρ by using the quasi-interpolants (4.1) and (4.3), but using noisy function values instead of exact function values. Thus

$$(Q_{h,\delta}\rho)(x) = \sum_{j=1}^{n_h} \lambda_{j,h}\big(\{\tilde{\rho}_{i,h,\delta}\}_i\big) s_{j,h}(x),$$

$$(Q_{h,\delta} f_\rho)(x) = \sum_{j=1}^{n_h} \lambda_{j,h}\big(\{\tilde{f}_{i,h,\delta}\}_i\big) s_{j,h}(x),$$

$$(Q_{h,\bar{h},\delta} k_\rho)(x, y) = \sum_{i,j=1}^{n_{\bar{h}}} \lambda_{i,j,\bar{h}}\big(\{\tilde{k}_{i',j',h,\bar{h}\delta}\}_{i',j'}\big) s_{j,\bar{h}}(y) s_{i,\bar{h}}(x).$$

For $[\rho, k] \in \mathscr{R} \times \mathscr{K}$, we define a new bilinear form $B_{h,\bar{h},\delta}(\cdot,\cdot;\rho,k_\rho)$ on $L_p(I^d) \times L_{p'}(I^d)$ as

$$B_{h,\bar{h},\delta}(v, w; \rho, k_\rho) = B(v, w; Q_{h,\delta}\rho, Q_{h,\bar{h},\delta} k_\rho) \qquad \forall\, v \in L_p(I^d),\, w \in L_{p'}(I^d)$$

and define a new linear functional $f(\cdot, \rho)$ on $L_{p'}(I^d)$ as

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

It would be reasonable to seek $u_{h,\bar{h},\delta} \in \mathscr{S}_h$ satisfying

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

However when $d < l$, this formulation leads to a linear system whose coefficient matrix contains entries that may not be computable. To see why, let us once again write

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

so that $\mathbf{u} = [\upsilon_1, \ldots, \upsilon_{n_h}]$ satisfies the linear system

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

where

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

and, for $1 \le i, j \le n_h$, we have

$$a_{i.j} = \langle s_{j,h}, s_{i,h} \rangle \qquad \text{and} \qquad b_{i,j} = \langle T_{\rho,k_\rho;h,\bar{h},\delta} s_{j,h}, s_{i,h} \rangle,$$

where

$$T_{\rho,k_\rho;h,\bar{h},\delta} v = \int_{I^d} (Q_{h,\bar{h},\delta} k_\rho)(\cdot, y) v(y) J(y; Q_{h,\delta}\rho)\, dy.$$

Hence

$$b_{i,j} = \sum_{i',j'=1}^{n_{\bar{h}}} \tilde{k}_{i,j,\delta} \left[\int_{I^d} s_{j',\bar{h}}(x) s_{j,h}(x)\, dx \right] \left[\int_{I^d} s_{i',\bar{h}}(y) s_{i,h}(y) J(y; Q_{h,\delta}\rho)\, dy \right].$$

If $d < l$, the integrands $s_{i',\bar{h}}(y) s_{i,h}(y) J(y; Q_{h,\delta}\rho)$ involve the square roots of piecewise polynomials. Hence these integrands may not have closed form antiderivatives. Thus the entries of $\mathbf{B}$ may not be computable, as claimed.

To deal with this problem, we use an approach found in [22, pg. 458] (and given in more detail in [21]), namely, replacing the square root appearing above by its Taylor expansion. For $\eta \in \mathbb{R}^{++}$ and any integer q, let $R_q(\cdot, \eta)$ denote the Taylor series of degree $q - 1$ for the square root at the point η, i.e.,

$$R_q(\xi, \eta) = \sqrt{\eta} + \sum_{i=1}^{q-1} \beta_q(\eta)(\xi - \eta)^i \qquad \forall\, \xi \in (\eta - 1, \eta + 1), \tag{4.4}$$

where

$$\beta_i(\eta) = \frac{1}{i!} \left(\frac{d}{d\xi} \right)^i \sqrt{\xi} \Bigg|_{\xi=\eta} = \frac{1}{\eta^{(2i-1)/2}} \binom{i - \frac{3}{2}}{i}.$$

Then

$$\left| \sqrt{\xi} - R_q(\xi, \eta) \right| \le |\beta_q|\, |\xi - \eta|^q \qquad \forall\, \xi \in (\eta - 1, \eta + 1). \tag{4.5}$$

We now define a modification $\tilde{T}_{\rho,k_\rho;h,\bar{h}\delta}$ of our operator $T_{\rho,k_\rho;h,\bar{h}\delta}$. First of all, if $d = l$, we simply take $\tilde{T}_{\rho,k_\rho;h,\bar{h},\delta} = T_{\rho,k_\rho;h,\bar{h},\delta}$. Now suppose that $d < l$. Let $\mathscr{Q}_h$ denote the set of h^{-d} cubes of side h into which I^d is partitioned when constructing $\mathscr{S}_h$. Then for $v \in L_p(I^d)$, we let

$$\tilde{T}_{\rho,k_\rho;h,\bar{h},\delta} v \Big|_K = \tilde{T}_{\rho,k_\rho;h,\bar{h},\delta;K} v \qquad \forall\, K \in \mathscr{Q}_h, \tag{4.6}$$

where

$$\tilde{T}_{\rho,k_\rho;h,\bar{h},\delta;K} v = \int_K (Q_{h,\bar{h},\delta} k_\rho)(\cdot, y)\, v(y)\, R_q\big(A(y; Q_{h,\delta}\rho), A(y^{(K)}; Q_{h,\delta}\rho)\big)\, dy \qquad \forall\, K \in \mathscr{Q}_h, \tag{4.7}$$

with $y^{(K)}$ a fixed evaluation point in K (such as the center or a specific corner) for each $K \in \mathscr{Q}_h$.

We are now ready to define our noisy modified Galerkin method. For $[\rho, k] \in \mathscr{R} \times \mathscr{K}$, we define a new bilinear form $\tilde{B}_{h,\bar{h},\delta}\big(\cdot, \cdot; \rho, k_\rho\big)$ on $L_p(I^d) \times L_{p'}(I^d)$ as

$$\tilde{B}_{h,\bar{h},\delta}(v, w; \rho, k_\rho) = \langle (I - \tilde{T}_{\rho,k_\rho;h,\bar{h},\delta})v, w\rangle \qquad \forall\, v \in L_p(I^d),\, w \in L_{p'}(I^d).$$

Then the *noisy modified Galerkin method* consists of finding $u_{h,\bar{h},\delta} \in \mathscr{S}_h$ satisfying

$$\tilde{B}_{h,\bar{h},\delta}(u_{h,\bar{h},\delta}, w; \rho, k_\rho) = f_{h,\delta}(w, \rho) \qquad \forall\, w \in \mathscr{S}_h.$$

If we write

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

then $\mathbf{u} = [\upsilon_1, \ldots, \upsilon_{n_h}]$ satisfies the linear system

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

where

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

and, for $1 \le i, j \le n_h$, we have

$$a_{i.j} = \langle s_{j,h}, s_{i,h}\rangle \qquad \text{and} \qquad b_{i,j} = \langle \tilde{T}_{\rho,k_\rho;h,\bar{h},\delta} s_{j,h}, s_{i,h}\rangle.$$

Note that the integrand appearing in each $b_{i,j}$ is piecewise polynomial. Hence the entries of $\mathbf{B}$ are computable, as required.

Let

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

where

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

and

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

with

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

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

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

where

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

5 Error analysis of the noisy modified Galerkin method

In this section, we establish an error bound for the noisy modified Galerkin method. As mentioned above, since the problem is unsolvable when $d < l$ and $r_1 = 1$, we only need to consider the case of $d = l$ or $r_1 \geq 2$. To derive our error bound, we first establish the uniform weak coercivity of the bilinear forms $B(\cdot, \cdot; \rho, k_\rho)$ for $[\rho, k] \in \mathscr{R} \times \mathscr{K}$. Once we know that the bilinear forms are uniformly weakly coercive, we can obtain an abstract error estimate, as a variant of the First Strang Lemma (see, e.g., [4, pg. 186]). The remaining task is then to estimate the various terms appearing in this abstract error estimate.

So, the first task is to establish uniform weak coercivity. Before doing so, we lay some groundwork.

The first thing we need to do is to recall approximation properties of the quasi-interpolation operators introduced in the previous section:

Lemma 5.1. *Let $\mathscr{S}_h$ and $\mathscr{S}_{\bar h} \otimes \mathscr{S}_{\bar h}$ be the spline spaces of degree m described in the previous section. For any $p \in [1, \infty]$ and $q \in \mathbb{Z}^{++}$, there exists $M_1 > 0$ (independent of h and $\bar h$) such that for any $r \in \{0, \dots, \min\{m, q, 2\}\}$, the following hold:*

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

$$\|w - Q_h w\|_{W^{r,p}(I^d)} \leq M_1 h^{\min\{m+1,q\}-r} \|w\|_{W^{q,p}(I^d)}.$$

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

$$\|w - Q_{\bar h \otimes \bar h} w\|_{W^{r,p}(I^{2d})} \leq M_1 \bar h^{\min\{m+1,q\}-r} \|w\|_{W^{q,p}(I^{2d})}.$$

Proof. See, e.g., [16]. ☐

Next, we need to establish an auxiliary lemma, which shows that the inverses of certain operators are uniformly bounded. By [21, Lemma 3.1], there exists $C_\circ > 0$ such that

$$\|k_\rho\|_{C^{\min\{r_1,r_2\}}(I^{2d})} \leq C_\circ \qquad \forall\, [\rho, k] \in \mathscr{R} \times \mathscr{K}.$$

Let

$$h_0 = \left(\frac{1}{2M_1 C_\circ c_4}\right)^{1/\min\{m+1,r_1,r_2\}}.$$

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 $\rho \in \mathscr{R}$ and any $g \in C(I^{2d})$, we have

$$T^*_{\rho,g} w = J(\cdot, \rho) \int_{I^d} g(x, \cdot) w(x)\, dx \qquad \forall\, w \in L_{p'}(I^d).$$

Lemma 5.2. *Let $h \in (0, h_0]$ and $k \in \mathscr{K}$. Then $I - T^*_{\rho, Q_{h\otimes h} k_\rho}$ is invertible on $L_{p'}(I^d)$, with*

$$\|(I - T^*_{\rho, Q_{h\otimes h} k_\rho})^{-1}\|_{\mathrm{Lin}[L_{p'}(I^d)]} \leq 2\, c_4.$$

Proof. Let $\rho \in \mathscr{K}$. Then

$$\|T^*_{\rho,g}\|_{\text{Lin}[L_{p'}(I^d)]} \le \|J(\cdot,\rho)\|_{C(I^d)}\|g\|_{C(I^{2d})} \le \kappa_{d,l}\|g\|_{C(I^{2d})} \qquad \forall\, g \in C(I^{2d}),$$

where $\kappa_{d,l}$ is defined in (2.4). For $[\rho,k] \in \mathscr{R} \times \mathscr{K}$ and $h \in (0,h_0]$, we may use Lemma 5.1 to find that

$$\begin{aligned}\|T^*_{\rho,(I-Q_{\bar h\otimes\bar h})k_\rho}\|_{\text{Lin}[L_{p'}(I^d)]} &\le \kappa_{d,l}\|(I-Q_{\bar h\otimes\bar h})k_\rho\|_{L_\infty(I^{2d})} \le \kappa_{d,l}M_1 h^{\min\{m+1,r_1,r_2\}}\|k_\rho\|_{W^{\min\{r_1,r_2\}}(I^{2d})}\\ &\le \kappa_{d,l}M_1 h_0^{\min\{m+1,r_1,r_2\}}C_\circ \le \frac{1}{2c_4}.\end{aligned}$$

Now follow the proof of [20, Lemma 8], replacing "k" by "k_ρ", "c_2" by "c_4," and "M_1" by "$\kappa_{d,l}M_1$." ☐

We now establish uniform weak coercivity.

Lemma 5.3. *There exist $h_1 > 0$ and $\gamma > 0$ such that the following holds: for any $[\rho,k] \in \mathscr{R}\times\mathscr{K}$, any $h \in (0,h_1]$, and any $v \in \mathscr{S}_h$, there exists nonzero $w \in \mathscr{S}_h$ such that*

$$B(v,w;\rho,k_\rho) \ge \gamma\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.$$

Proof. Let $[\rho,k] \in \mathscr{R}\times\mathscr{K}$ 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 $v \ne 0$.

By [20, Lemma 10], there exists nonzero $g \in L_{p'}(I^d)$ such that

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

Choose

$$w = (I - T^*_{Q_{\bar h\otimes\bar h}k_\rho})^{-1}\mathscr{P}_h g.$$

Since $T^*_{Q_{\bar h\otimes\bar h}k_\rho} : \mathscr{S}_h \to \mathscr{S}_h$, we may use (4.2) and Lemma 5.2 to see that w is a well-defined element of $\mathscr{S}_h$, and that

$$\|w\|_{L_{p'}(I^d)} \le 2\pi_{p'}c_4\|g\|_{L_{p'}(I^d)}.$$

Hence

$$\langle (I - T_{\rho,Q_{\bar h\otimes\bar h}k_\rho})v, w\rangle \ge \tfrac{1}{2}\|v\|_{L_p(I^d)}\|g\|_{L_{p'}(I^d)} \ge \frac{1}{4\pi_{p'}c_4}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)},$$

from which we see that $w \ne 0$.

Using the Minkowski inequality (as in the proof of [20, Lemma 11]), we find

$$\begin{aligned}|\langle T_{\rho,(I-Q_{\bar h\otimes\bar h})k_\rho}v, w\rangle| &\le \|(I-Q_{\bar h\otimes\bar h})k_\rho\|_{L_\infty(I^{2d})}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}\\ &\le M_1 C_\circ h_0^{\min\{m+1,r_1,r_2\}}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.\end{aligned}$$

Combining the last two inequalities and setting

$$h_1 = \min\left\{\frac{1}{8\pi_{p'}c_4M_1C_\circ}, h_0\right\} \qquad \text{and} \qquad \gamma = \frac{1}{8\pi_{p'}c_4},$$

the lemma follows. ☐

Since the bilinear forms $B(\cdot,\cdot;\rho,k)$ are uniformly weakly coercive for $k \in \mathscr{K}$, we have the following variant of the First Strang Lemma found in [4, pg. 186]:

Lemma 5.4. *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 $h, \bar{h} \in (0,h_2]$, any $[\rho,k] \in \mathscr{R} \times \mathscr{K}$, and any $v, w \in \mathscr{S}_h$, we have*

$$|B(v,w;\rho,k_\rho) - \tilde{B}_{h,\bar{h},\delta}(v,w;\rho,k_\rho)| \le \tfrac{1}{2}\gamma\, \|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}$$

where γ is as in Lemma 5.3*. 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 Galerkin method is well-defined. That is, there exists a unique $u_{h,\bar{h},\delta} \in \mathscr{S}_h$ such that*
$$\tilde{B}_{h,\bar{h},\delta}(u_{h,\bar{h},\delta}, w;\rho,k_\rho) = f_{h,\delta}(w;\rho) \qquad \forall\, w \in \mathscr{S}_h.$$
2. *Let $u_\rho = S([f,k])$. Then*

$$\|u_\rho - u_{h,\bar{h},\delta}\|_{L_{p'}(I^d)} \le M_2 \inf_{v\in\mathscr{S}_h}\left[\|u_\rho - v\|_{L_p(I^d)} + \sup_{w\in\mathscr{S}_h}\left(\frac{|B(v,w;\rho,k_\rho) - \tilde{B}_{h,\bar{h},\delta}(v,w;\rho,k_\rho)|}{\|w\|_{L_{p'}(I^d)}} + \frac{|\langle f,w\rangle - f_{h,\delta}(w;\rho)|}{\|w\|_{L_{p'}(I^d)}}\right)\right].$$

Proof. See, e.g., [19, pp. 310–312] for the proof of a version having slightly more restrictive conditions. ☐

We now estimate the quantities appearing in the second part of Lemma 5.4. First, we estimate the difference between the bilinear forms $B(\cdot,\cdot;\rho,k_\rho)$ and $\tilde{B}_{h,\bar{h},\delta}(\cdot,\cdot;\rho,k_\rho)$.

Lemma 5.5. *Suppose that $d = l$ or $r_1 \ge 2$. Let $m \ge \max\{r_1,r_2\} - 1$ and let $q \ge r_1 - \nu$ in* (4.6) *and* (4.7)*, where ν is given by* (1.4)*. There exists $M_3 > 0$ such that for any positive h, $\bar{h}$, and δ, for any $[\rho,k] \in \mathscr{K}$, and for any $v, w \in \mathscr{S}_h$, we have*

$$|B(v,w;\rho,k_\rho) - \tilde{B}_{h,\bar{h},\delta}(v,w;\rho,k_\rho)| \le M_3(h^{r_1-\nu} + \bar{h}^{\,r_2} + \delta)\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.$$

Proof. Given h, $\bar{h}$, δ, ρ, k, v, and w as in the statement of the lemma, define

$$\begin{aligned} A_1 &= \langle (T_{\rho,k_\rho} - T_{\rho,k_\rho;\bar{h}})v, w\rangle,\\ A_2 &= \langle (T_{\rho,k_\rho;\bar{h}} - T_{\rho,k_\rho;h,\bar{h}})v, w\rangle,\\ A_3 &= \langle (T_{\rho,k_\rho;h,\bar{h}} - T_{\rho,k_\rho;h,\bar{h},\delta})v, w\rangle,\\ A_4 &= \langle (T_{\rho,k_\rho;h,\bar{h},\delta} - \tilde{T}_{\rho,k_\rho;h,\bar{h},\delta})v, w\rangle, \end{aligned}$$

where

$$T_{\rho,k_\rho;\bar{h}}v = \int_{I^d} (Q_{\bar{h}\otimes\bar{h}}k_\rho)(\cdot,y)v(y)J(y;\rho)\,dy$$

and

$$T_{\rho,k_\rho;h,\bar{h}}v = \int_{I^d} (Q_{\bar{h}\otimes\bar{h}}k_\rho)(\cdot,y)v(y)J(y;Q_h\rho)\,dy.$$

Then

$$|B(v,w;\rho,k_\rho) - B_{h,\bar{h},\delta}(v,w;\rho,k_\rho)| \le |A_1| + |A_2| + |A_3| + |A_4|. \tag{5.1}$$

We first estimate $|A_1|$. From (2.4), we see that

$$\|J(\cdot,\rho)\|_{L_1(I^d)} \le \kappa_{d,l}.$$

Using Lemma 5.1, we obtain

$$\begin{aligned}|A_1| &\le \|(T_{\rho,k_\rho} - T_{\rho,k_\rho;\bar h})v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \le \kappa_{d,l}\|(I - Q_{\bar h\otimes\bar h})k_\rho\|_{L_\infty(I^{2d})}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \\ &\preccurlyeq \bar h^{\,r_2}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)}.\end{aligned} \tag{5.2}$$

Next, we estimate $|A_2|$. We have

$$\begin{aligned}A_2 &= \int_{I^d}\left[\int_{I^d}(Q_{\bar h\otimes\bar h}k_\rho)(x,y)v(y)[J(y;\rho) - J(y;Q_h\rho)]\,dy\right]w(x)\,dx \\ &= \int_{I^d}\int_{I^d}\omega(x,y)[\det A(y;\rho) - \det A(y;Q_h\rho)]\,dy\,dx,\end{aligned}$$

where

$$\omega(x,y) = \frac{(Q_{\bar h\otimes\bar h}k_\rho)(x,y)}{J(y;\rho) + J(y;Q_h\rho)}.$$

Let Π_d denote the set of all permutations of $\{1,\ldots,d\}$. Define

$$b_{\boldsymbol\pi,i,j}(x,y) = \omega(x,y)\bar a_{\pi_1,1}(y)\ldots\bar a_{\pi_{i-1},i-1}(y)a_{\pi_{i+1},i+1}(y)\ldots a_{\pi_d,d}(y)(\partial_i\rho_j)(y)$$

and

$$\bar b_{\boldsymbol\pi,i,j}(x,y) = \omega(x,y)\bar a_{\pi_1,1}(y)\ldots\bar a_{\pi_{i-1},i-1}(y)a_{\pi_{i+1},i+1}(y)\ldots a_{\pi_d,d}(y)(\partial_{\pi_i}\bar\rho_j)(y),$$

where $\partial_i = (\partial/\partial y_i)$ and $a_{i,j}$ and $\bar a_{i,j}$ respectively denote the (i,j)th components of $A(\cdot,\rho)$ and $A(\cdot,Q_h\rho)$.

As on [22, pp. 455-456], we find

$$A_2 = \sum_{\boldsymbol\pi\in\Pi_d}(-1)^{|\boldsymbol\pi|}\sum_{i=1}^d\sum_{j=1}^d\theta_{\boldsymbol\pi,i,j}, \tag{5.3}$$

where $|\boldsymbol\pi|$ denote the sign of $\boldsymbol\pi\in\Pi_d$ and we define

$$\begin{aligned}\theta_{\boldsymbol\pi,i,j} = \int_{I^d}&\left[\int_{I^d}b_{\boldsymbol\pi,i,j}(x,y)v(y)\partial_{\pi_i}\big(\rho_j(y) - \bar\rho_j(y)\big)\,dy\right]w(x)\,dx \\ &+ \int_{I^d}\left[\int_{I^d}\bar b_{\boldsymbol\pi,i,j}(x,y)v(y)\partial_i\big(\rho_j(y) - \bar\rho_j(y)\big)\,dy\right]w(x)\,dx\end{aligned}$$

for $\boldsymbol\pi\in\Pi_d$ and $i,j\in\{1,\ldots,d\}$.

Suppose we only have $r_1\ge 1$ or $r_2\ge 1$. Then

$$|\theta_{\boldsymbol\pi,i,j}| \le \|b_{\boldsymbol\pi,i,j}\|_{L_\infty(I^{2d})}\|\rho_j - \bar\rho_j\|_{W^{1,\infty}(I^d)}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \preccurlyeq h^{r_1-1}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)},$$

and hence

$$|A_2| \preccurlyeq h^{r_1-1}\|v\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \qquad \text{if } r_1\ge 1 \text{ or } r_2\ge 1. \tag{5.4}$$

Now suppose that $r_1\ge 2$ and $r_2\ge 2$. Fix $\boldsymbol\pi\in\Pi_d$ and $i,j\in\{1,\ldots,d\}$. Let I_j^{d-1} denote the $(d-1)$-dimensional cube in the variables $y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_d$ and let $dy_j^{d-1} = dy_1\,\ldots\,dy_{j-1}\,dy_{j+1}\,\ldots\,dy_d$.

Since $\rho_j, \bar\rho_j \in W^{2,\infty}(I^d)$, we have $b_{\boldsymbol{\pi},i,j}(x,\cdot), \bar b_{\boldsymbol{\pi},i,j}(x,\cdot) \in W^{1,\infty}(I^d)$ for all $x \in I^d$. Hence we can integrate by parts, obtaining

$$\theta_{\boldsymbol{\pi},i,j} = \theta_{\boldsymbol{\pi},i,j;1} - \theta_{\boldsymbol{\pi},i,j;2} + \theta_{\boldsymbol{\pi},i,j;3} - \theta_{\boldsymbol{\pi},i,j;4} \tag{5.5}$$

where

$$\begin{aligned}
\theta_{\boldsymbol{\pi},i,j;1} &= \int_{I^d} \left\{ \int_{I^{d-1}_{\pi_i}} \Big[b_{\boldsymbol{\pi},i,j}(x,y)v(y)\big(\rho_j(y) - \bar\rho_j(y)\big)\Big]_{y_{\pi_i}=0}^{y_{\pi_i}=1} dy^{d-1}_{\pi_i} \right\} w(x)\,dx \\
\theta_{\boldsymbol{\pi},i,j;2} &= \int_{I^d} \left\{ \int_{I^d} \partial_{\pi_i}[b_{\boldsymbol{\pi},i,j}(x,y)v(y)]\big(\rho_j(y) - \bar\rho_j(y)\big)\,dy \right\} w(x)\,dx \\
\theta_{\boldsymbol{\pi},i,j;3} &= \int_{I^d} \left\{ \int_{I^{d-1}_{i}} \Big[\bar b_{\boldsymbol{\pi},i,j}(x,y)v(y)\big(\rho_j(y) - \bar\rho_j(y)\big)\Big]_{y_{i}=0}^{y_{i}=1} dy^{d-1}_{i} \right\} w(x)\,dx \\
\theta_{\boldsymbol{\pi},i,j;4} &= \int_{I^d} \left\{ \int_{I^d} \partial_{i}[\bar b_{\boldsymbol{\pi},i,j}(x,y)v(y)]\big(\rho_j(y) - \bar\rho_j(y)\big)\,dy \right\} w(x)\,dx\,.
\end{aligned}$$

Let us estimate $\theta_{\boldsymbol{\pi},i,j;1}$. We clearly have

$$|\theta_{\boldsymbol{\pi},i,j;1}| \le I_1 \|w\|_{L_{p'}(I^d)},$$

where

$$\begin{aligned}
I_1 &= \left\{ \int_{I^d} \left| \int_{I^{d-1}_{\pi_i}} \Big[b_{\boldsymbol{\pi},i,j}(x,y)v(y)\big(\rho_j(y) - \bar\rho_j(y)\big)\Big]_{y_{\pi_i}=0}^{y_{\pi_i}=1} dy^{d-1}_{\pi_i} \right|^p dx \right\}^{1/p} \\
&\le \|b_{\boldsymbol{\pi},i,j}\|_{C(I^{2d})} \Big[\|v\|_{L_1(I^{d-1}_{\pi_i,0})} + \|v\|_{L_1(I^{d-1}_{\pi_i,1})} \Big] \|\rho_j - \bar\rho_j\|_{L_\infty(I^d)},
\end{aligned}$$

with $I^{d-1}_{i,a} = \{\, y \in I^{d-1} : y_i = a \,\}$. Using Lemma 5.1, along with the conditions defining $\mathscr{F}$, we find that

$$I_1 \preccurlyeq h^{r_1} \Big[\|v\|_{L_1(I^{d-1}_{\pi_i,0})} + \|v\|_{L_1(I^{d-1}_{\pi_i,1})} \Big].$$

Using the inverse theorem [4, Thm. 3.2.6], we have

$$\|v\|_{L_1(I^{d-1}_{\pi_i,a})} \le \|v\|_{L_\infty(I^d)} \preccurlyeq h^{-d/p} \|v\|_{L_p(I^d)} \qquad \text{for } a = 0 \text{ and } a = 1.$$

Hence

$$|\theta_{\boldsymbol{\pi},i,j;1}| \preccurlyeq h^{r_1 - d/p} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{5.6}$$

Similarly, we find

$$|\theta_{\boldsymbol{\pi},i,j;3}| \preccurlyeq h^{r_1 - d/p} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{5.7}$$

Let us estimate $\theta_{\boldsymbol{\pi},i,j;2}$. Using the product rule of differential calculus, we find

$$|\theta_{\boldsymbol{\pi},i,j;2}| \le (I_2 + I_3) \|w\|_{L_{p'}(I^d)},$$

where

$$I_2 = \left\{ \int_{I^d} \left| \int_{I^d} [\partial_{\pi_i} b_{\boldsymbol{\pi},i,j}(x,y)v(y)[\rho_j(y) - \bar\rho_j(y)]\,dy \right|^p dx \right\}^{1/p}$$

and

$$I_3 = \left\{ \int_{I^d} \left| \int_{I^d} b_{\boldsymbol{\pi},i,j}(x,y)\partial_{\pi_i} v(y)[\rho_j(y) - \bar\rho_j(y)]\,dy \right|^p dx \right\}^{1/p}.$$

Using Lemma 5.1, along with the conditions defining $\mathscr{F}$, we find

$$|I_2| \le \|b_{\boldsymbol{\pi},i,j}\|_{W^{1,\infty}(I^{2d})} \|v\|_{L_p(I^d)} \|\rho_j - \bar{\rho}_j\|_{L_\infty(I^d)} \preccurlyeq h^{r_1} \|v\|_{L_p(I^d)}$$

and

$$|I_3| \le \|b_{\boldsymbol{\pi},i,j}\|_{L_\infty(I^{2d})} |v|_{W^{1,\infty}(I^d)} \preccurlyeq h^{r_1} |v|_{W^{1,\infty}(I^d)}.$$

Our estimate of $|I_2|$ is satisfactory, but we need to do some work on the estimate of $|I_3|$. Once again using the inverse theorem [4, Thm. 3.2.6], we have

$$|v|_{W^{1,\infty}(I^d)} \preccurlyeq h^{d(1-1/p)} \cdot h^{-1} \|v\|_{L_p(I^d)} = h^{d/p'-1} \|v\|_{L_p(I^d)},$$

and so

$$|I_3| \preccurlyeq h^{r_1+d/p'-1} \|v\|_{L_p(I^d)}.$$

Since $d/p + d/p' = d \ge 1$, we have $d/p' - 1 \ge -d/p$, and thus $h^{d/p'-1} \le h^{-d/p}$. Thus we find

$$|I_3| \preccurlyeq h^{r_1-d/p} \|v\|_{L_p(I^d)}.$$

Using our estimates of $|I_2|$ and $|I_3|$, we have

$$|\theta_{\boldsymbol{\pi},i,j;2}| \preccurlyeq h^{r_1-d/p} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{5.8}$$

Similarly, we find

$$|\theta_{\boldsymbol{\pi},i,j;4}| \preccurlyeq h^{r_1-d/p} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{5.9}$$

Combining (5.3)–(5.9), we obtain

$$|A_2| \preccurlyeq h^{r_1-\min\{1,d/p\}} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)} \qquad \text{for } r_1 \ge 2 \text{ and } r_2 \ge 2. \tag{5.10}$$

We next estimate $|A_3|$. Let

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

Then

$$|A_3| \le \|\zeta\|_{L_\infty(I^{2d})} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}.$$

For $x \in I^d$, define

$$\operatorname{supp}_{\bar{h}} x = \left\{ i \in \{1, \ldots, n_{\bar{h}}\} : x \text{ is in the support of } s_{i,\bar{h}} \right\}.$$

Then there exist positive constants σ_1 and σ_2, independent of x, j, and $\bar{h}$, such that

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

and hence

$$|\zeta(x, y)| \le \sum_{\substack{i \in \operatorname{supp}_{\bar{h}} x \\ j \in \operatorname{supp}_{\bar{h}} y}} |k_\rho(x_{i,\bar{h}}, x_{j,\bar{h}}) - \tilde{k}_{i,j,\delta}| \, |s_{j,\bar{h}}(y)| \, |s_{i,\bar{h}}(x)| \le \sigma_1 \sigma_2^2 \, \delta,$$

so that

$$|A_3| \preccurlyeq \delta \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \tag{5.11}$$

We now estimate $|A_4|$. Of course, $A_4 = 0$ when $d = l$, so we only need to consider the case $d < l$. For a cube $K \in \mathscr{Q}_h$, let

$$\theta_K = \int_{I^d} \left\{ \int_K (Q_{h,\bar{h},\delta} k_\rho)(x, y)\big[J(y; Q_{h,\delta}\rho) - R_q\big(A(y; Q_{h,\delta}\rho), A(y^{(K)}; Q_{h,\delta}\rho)\big)\big]v(y)\, dy \right\} w(x)\, dx.$$

Recalling the definition (4.7), along with the error estimate (4.5), we find that

$$|\theta_K| \preccurlyeq h^{r_1-\nu} \int_{I^d} \left[\int_K |v(y)|\, dy \right] |w(x)|\, dx.$$

Hence

$$\begin{aligned} |A_4| &\le \sum_{K \in \mathscr{Q}_h} |\theta_K| \preccurlyeq h^{r_1-\nu} \int_{I^d} \left[\int_{I^d} |v(y)|\, dy \right] |w(x)|\, dx = h^{r_1-\nu} \|v\|_{L_1(I^d)} \|w\|_{L_1(I^d)} \\ &\le h^{r_1-\nu} \|v\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}. \end{aligned} \tag{5.12}$$

Finally, substituting (5.2), (5.4), (5.10), (5.11), and (5.12) into (5.1), we get the estimate in the statement of our lemma. ☐

Next, we need to estimate the difference between the linear forms $\langle f_\rho, \cdot \rangle$ and $f_{h,\delta}(\cdot, \rho)$. Before doing this, we will need to do prove a result concerning the Sobolev smoothness of composite functions.

Lemma 5.6. *Let $\rho \in \mathscr{R}$ and $v \in W^{r_3,p}(I^l)$. There exists $M_4 > 0$, independent of ρ and v, such that*

$$\|v_\rho\|_{W^{\min\{r_1, r_3-(l-d)/p\}}(I^d)} \le M_4 \|v\|_{W^{r_3,p}(I^d)}.$$

Proof. Let ρ and v be as given. Let α be a multi-index with d entries. The multivariate Faa di Bruno formula [5, Theorem 2.1] states that

$$(D^\alpha v_\rho)(x) = \sum_{1 \le |\beta| \le |\alpha|} (D^\beta v)\big(\rho(x)\big) \zeta_{\alpha,\beta}(x)$$

with

$$\zeta_{\alpha,\beta}(x) = \sum_{i=1}^{|\alpha|} \sum_{p_i(\alpha,\beta)} \alpha! \prod_{j=1}^{i} \frac{[D^{\ell_j} \rho(x)]^{k_j}}{k_j!(\ell_j!)^{|k_j|}}.$$

The set $p_i(\alpha, \beta)$ mentioned in the inner sum consists of all $(k_1, \ldots, k_i, \ell_1, \ldots, \ell_i) \in \big((\mathbb{Z}^{++})^l\big)^i \times \big((\mathbb{Z}^{++})^d\big)^i$, such that $\ell_1 < \cdots < \ell_i$ (lexicographically) and

$$\sum_{j=1}^{i} k_j = \beta \in (\mathbb{Z}^{++})^l \qquad \text{and} \qquad \sum_{j=1}^{i} |k_j| \ell_j = \alpha \in (\mathbb{Z}^{++})^d.$$

Hence

$$\|D^\alpha v_\rho\|_{L_p(I^d)} \le \alpha! \max_{1 \le |\beta| \le |\alpha|} \|\zeta_{\alpha,\beta}\|_{L_\infty(I^d)} \left[\int_{I^d} \big|(D^\beta v)\big(\rho(x)\big)\big|^p\, dx \right]^{1/p}.$$

If $|\alpha| \le r_1$, the conditions defining $\mathscr{R}$ guarantee that $\max_{1 \le |\beta| \le |\alpha|} \|\zeta_{\alpha,\beta}\|_{L_\infty(I^d)}$ is bounded, independently of ρ. Moreover

$$\int_{I^d} \big|(D^\beta v)\big(\rho(x)\big)\big|^p\, dx \le \frac{1}{c_2} \int_{I^d} \big|(D^\beta v)\big(\rho(x)\big)\big|^p J(x; \rho)\, dx = \frac{1}{c_2} \int_{\Omega_\rho} |(D^\beta v)(s)|^p\, ds.$$

Hence

$$\|D^\alpha v_\rho\|_{L_p(I^d)} \preccurlyeq \|D^\alpha v\|_{L_p(\Omega_\rho)},$$

provided the right-hand side is bounded. For any $r \le r_1$, we thus have

$$\|v_\rho\|_{W^{r,p}(I^d)} \preccurlyeq \|v\|_{W^{r,p}(\Omega_\rho)}.$$

Suppose that $d = l$. Setting $r = \min\{r_1, r_3\} = \min\{r_1, (l-d)/p\}$, we have

$$\|v_\rho\|_{W^{r,p}(I^d)} \preccurlyeq \|v\|_{W^{r,p}(\Omega_\rho)} \le \|v\|_{W^{r_3,p}(I^l)}, \tag{5.13}$$

and we are done.

Otherwise, let $\Omega^{[d]} = \Omega_\rho$ and choose $\Omega^{[d+1]}$ to be a $(d+1)$-dimensional C^{R_1}-manifold in I^l such that $\Omega^{[d]} \subseteq \partial\Omega^{[d+1]}$. Let $F\colon I^{d+1} \to \Omega^{[d+1]}$ be a bijection. Then for any $r \le r_1$, we have

$$\begin{aligned}
\|v_\rho\|_{W^{r,p}(\Omega^{[d]})} &\le \|v_\rho\|_{W^{r,p}(\partial\Omega^{[d+1]})} && \text{(since } \Omega^{[d]} \subseteq \partial\Omega^{[d+1]}\text{)}\\
&\preccurlyeq \|v_\rho \circ F\|_{W^{r,p}(\partial I^{d+1})} && \text{as in [4, Thm. 4.3.2]}\\
&\preccurlyeq \|v_\rho \circ F\|_{W^{r+1/p}(I^{d+1})} && \text{using the trace theorem [1, § 7.56]}\\
&\preccurlyeq \|v_\rho\|_{W^{r,p}(\Omega^{[d+1]})} && \text{as in [4, Thm. 4.3.2].}
\end{aligned}$$

If $l = d+1$, we stop. Otherwise we repeat this process, getting a sequence $\Omega^{[d]}, \dots, \Omega^{[l]}$ such that

$$\|v_\rho\|_{W^{r,p}(\Omega_\rho)} = \|v_\rho\|_{W^{r,p}(\Omega^{[d]})} \preccurlyeq \|v_\rho\|_{W^{r+1/p,p}(\Omega^{[d+1]})} \preccurlyeq \cdots \preccurlyeq \|v_\rho\|_{W^{r+(l-d)/p,p}(\Omega^{[l]})} \le \|v_\rho\|_{W^{r+(l-d)/p,p}(I^l)}$$

for any $r \le r_1$. Taking $r = \min\{r_1, r_3 = (l-d)/p\}$, the desired result follows from the previous inequality and the fi rst inequality in (5.13). $\square$

Using this lemma, we are now able to estimate the difference between the linear forms $\langle f_\rho, \cdot\rangle$ and $f_{h,\delta}(\cdot, \rho)$.

Lemma 5.7. *Let $m \ge r_3 - 1$. There exists $M_5 > 0$ such that*

$$|\langle f_\rho, w\rangle - f_{h,\delta}(w;\rho)| \le M_5(h^{\min\{r_1, r_3-(l-d)/p} + \delta)\|f\|_{W^{r_3,p}(I^l)}\|w\|_{L_{p'}(I^d)}$$

for all $\rho \in \mathscr{R}$, $f \in W^{r_3,p}(I^l)$, $h > 0$, $\delta \ge 0$, and $w \in \mathscr{S}_h$.

Proof. Choose $\rho \in \mathscr{R}$, $f \in W^{r_3,p}(I^l)$, $h > 0$, $\delta \ge 0$, and $w \in \mathscr{S}_h$. Then

$$|\langle f_\rho, w\rangle - f_{h,\delta}(w;\rho)| \le |A_1| + |A_2|, \tag{5.14}$$

where

$$A_1 = \langle f_\rho - Q_h f_\rho, w\rangle$$

and

$$A_2 = \langle Q_h f_\rho, w\rangle - f_{h,\delta}(w;\rho).$$

Using Lemmas 5.1 and 5.6, we have

$$\begin{aligned}
|A_1| &\le \|f_\rho - Q_h f_\rho\|_{L_p(I^d)}\|w\|_{L_{p'}(I^d)} \le M_1 h^{\min\{r_1, r_3-(l-d)/p\}}\|f_\rho\|_{W^{\min\{r_1, r_3-(l-d)/p, p\}}(I^d)}\|w\|_{L_{p'}(I^d)}\\
&\le M_1 M_4 h^{\min\{r_1, r_3-(l-d)/p}\|f\|_{W^{r_3,p}(I^l)}\|w\|_{L_{p'}(I^d)}.
\end{aligned}$$

Recalling the definitions of σ_1 and σ_2 in Lemma 5.5, we have

$$|A_2| \le \Bigl\| \sum_{j=1}^{n_h} [f_\rho(x_{j,h}) - \tilde{f}_{j,h,\delta}] s_{j,h} \Bigr\|_{L_{p'}(I^d)} \|w\|_{L_{p'}(I^d)} \le \delta \Bigl\| \sum_{j=1}^{n_h} |s_{j,h}| \Bigr\|_{L_p(I^d)} \|w\|_{L_{p'}(I^d)}$$
$$\le \delta \Bigl\| \sum_{j=1}^{n_h} |s_{j,h}| \Bigr\|_{L_\infty(I^d)} \|w\|_{L_{p'}(I^d)} \le \sigma_1 \sigma_2 \, \delta \|w\|_{L_{p'}(I^d)}.$$

Our lemma follows from these last two inequalities, along with (5.14). □

Our final preparatory step is to establish a "shift theorem" relating the smoothness of $(I - T_{\rho,k_\rho})^{-1} f$ to the smoothnesses of ρ, k, and f.

Lemma 5.8. *Let* $[\rho, k] \in \mathscr{R} \times \mathscr{K}$ *and* $f \in W^{r_3,p}(I^l)$.. *Then*

$$\|(I - T_{\rho,k_\rho})^{-1} f_\rho\|_{W^{\min\{r_1,r_2,r_3-(l-d)/p\},p}(I^d)} \le M_6 \|f\|_{W^{r_3,p}(I^l)},$$

where $M_6 > 0$ *is independent of* ρ, k, *and* f.

Proof. Given such ρ, k, and f, we have

$$\|f_\rho\|_{W^{\min\{r_1,r_2,r_3-(l-d)/p\},p}(I^d)} \le M_4 \|f\|_{W^{r_3,p}(I^l)}$$

by Lemma 5.6. Following the proof of [20, Lemma 16], we find that

$$\|(I - T_{\rho,k_\rho})^{-1} f_\rho\|_{W^{\min\{r_1,r_2,r_3-(l-d)/p\},p}(I^d)} \preccurlyeq \|f_\rho\|_{W^{\min\{r_1,r_2,r_3-(l-d)/p\},p}(I^d)}.$$

The desired result follows from these two bounds. □

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

Theorem 5.1. *Suppose that* $d = l$ *or* $r_1 \ge 2$. *Let* $m \ge \max\{r_1, r_2\} - 1$ *and let* $q \ge r_1 - \nu$ *in* (4.6) *and* (4.7), *and let* ν *be given by* (1.4). *Choose* $h_2 > 0$ *and* $\delta_0 > 0$ *such that*

$$M_3(h_1^{r_1-\nu} + h_2^{r_2} + \delta_0) \le \tfrac{1}{2}\gamma,$$

where h_1 *and* γ *are as in Lemma* 5.3. *There exists* $M_7 > 0$ *such that for any* $h \in (0, h_1]$, $\bar{h} \in (0, h_2]$, *and* $\delta \in [0, \delta_0]$:

1. *The noisy modified Galerkin method is well-defined.*

2. *We have the error bound*

$$e(\phi_{h,\bar{h},\delta}, \mathbb{N}_{h,\bar{h},\delta}) \le M_7(h^{\min\{r_1-\nu,r_2,r_3-(l-d)/p\}} + \bar{h}^{\,r_2} + \delta).$$

Proof. Let h, $\bar{h}$, and δ be as described. Choose $[\rho, k, f] \in \mathscr{F}$, and let $u_\rho = S([\rho, k, f])$. Using Lemmas 5.4 and 5.5, we immediately see that $u_{h,\bar{h},\delta} = \phi_{h,\bar{h},\delta}\bigl(\mathbb{N}_{h,\bar{h},\delta}([\rho, k, f])\bigr)$ is well-defined. It only remains to to prove the error bound. Let $r = \min\{r_1 - \nu, r_2, r_3 - (l - d)/p\}$, and set $v = Q_h u_\rho$. Using Lemmas 5.1 and 5.8, along with the conditions defining the class $\mathscr{F}$, we have

$$\|u_\rho - v\|_{L_p(I^d)} \le M_1 h^r \|u_\rho\|_{W^{r,p}(I^d)} \le M_1 h^r \|u_\rho\|_{W^{\min\{r_1,r_2,r_3-(l-d)/p\},p}(I^d)} \le M_1 M_6 h^r \|f\|_{W^{r_3,p}(I^l)}$$
$$\le M_1 M_6 h^r.$$

The desired result follows once we substitute this inequality, along with the results of Lemmas 5.5 and 5.7, into the error bound of Lemma 5.4. □

6 Minimizing the error of the noisy modified Galerkin method

Let $n \in \mathbb{Z}^+$, and consider noisy modified Galerkin methods using at most n noisy function evaluations. How can we choose the parameters h and $\bar{h}$ that will minimize the error of the noisy modified Galerkin method?

Recall that

$$\operatorname{card} \mathbb{N}_{h,\bar{h},\delta} \asymp n_{\bar{h}}^2 + n_h,$$

where

$$n_h = \left(\frac{m+1}{h}\right) \qquad \text{and} \qquad n_{\bar{h}} = \left(\frac{m+1}{\bar{h}}\right)^d.$$

It will be useful to rewrite this bound in terms of a proportionality constant, so that we have

$$\operatorname{card} \mathbb{N}_{h,\bar{h},\delta} \leq C_{\text{card}} \left(n_{\bar{h}}^2 + n_h\right).$$

As in the proof of Theorem 5.1, let

$$r = \min\left\{r_1 - \nu, r_2, r_3 - \frac{l-d}{p}\right\},$$

where ν is given by (1.4). Let

$$\tau = \frac{\max\{r_1, r_2\}}{\min\{r_1, r_2\}}.$$

We define parameters κ and $\bar{\kappa}$ as follows:

1. Suppose that $r_2 < 2r$, so that $r_2 < 2\min\{r_1, r_2\}$. Take

$$\kappa = \left(\frac{n}{\tau^{2d} C_{\text{card}}}\right)^{r_2/(2\min\{r_1,r_2\})} \qquad \text{and} \qquad \bar{\kappa} = \sqrt{\frac{n}{\theta 2d C_{\text{card}}} - \kappa}.$$

2. Suppose that $r_2 = 2r$. Take

$$\kappa = \frac{n}{2\theta^{2d} C_{\text{card}}} \qquad \text{and} \qquad \bar{\kappa} = \sqrt{\frac{n}{2\theta^{2d} C_{\text{card}}}}.$$

3. Suppose that $r_2 > 2r$. Take

$$\bar{\kappa} = \left(\frac{n}{\theta^{2d} C_{\text{card}}}\right)^{r/r_2} \qquad \text{and} \qquad \kappa = \frac{n}{\theta^{2d} C_{\text{card}}} - \bar{\kappa}^2.$$

With these definitions of κ and $\bar{\kappa}$, define meshsizes

$$h = \frac{\min\{r_1, r_2\}}{\kappa^{1/d}} \qquad \text{and} \qquad \bar{h} = \frac{\min\{r_1, r_2\}}{\bar{\kappa}^{1/d}}. \tag{6.1}$$

Since the degree of the spline space satisfies

$$m = \max\{r_1, r_2\} - 1,$$

we find that

$$n_h = \tau^d \kappa \qquad \text{and} \qquad n_{\bar{h}} = \tau^d \bar{\kappa}.$$

In the sequel, we shall assume without loss of generality that h and $\bar{h}$ have been chosen so that n_h and $n_{\bar{h}}$ are positive integers. With these choices of h and $\bar{h}$, let

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

Then

$$\operatorname{card} \mathbb{N}_{n,\delta} \le C_{\text{card}} \left(n_{\bar{h}}^2 + n_h \right) \le C_{\text{card}} \theta^{2d} (\bar{\kappa}^2 + \kappa) \le n.$$

We now have

Theorem 6.1. *Suppose that $d = l$ or $r_1 \ge 2$. Let $m = \max\{r_1, r_2\} - 1$ and let $q \ge r_1 - \nu$ in* (4.6) *and* (4.7), *where ν is given by* (1.4). *Then there exists $n_0 \in \mathbb{Z}^{++}$ and $\delta_0 > 0$ such that $\phi_{n,\delta}$ is well-defined for $n \ge n_0$ and $\delta \in [0, \delta_0]$. Furthermore, there exists a positive constant M_8 such that*

$$e(\phi_{n,\delta}, \mathbb{N}_{n,\delta}) \le M_8 (n^{-\mu_2} + \delta) \qquad \forall\, n \ge n_0,\, \delta \in [0, \delta_0],$$

where μ_2 is defined by (1.3).

Proof. The proof is the same as that of [20, Thm. 18], with the obvious minor notational changes. ☐

Comparing Theorems 3.1 and 6.1, we find the following bounds on the nth minimal error of noisy information:

Corollary 6.1.

1. *Let $d < l$ and $r_1 = 1$. Then*
$$r(n, \delta) \asymp 1.$$
2. *Let $d = l$ or $r_1 \ge 2$. Then*
$$\left(\frac{1}{n}\right)^{\mu_1} + \delta \preccurlyeq r(n, \delta) \preccurlyeq \left(\frac{1}{n}\right)^{\mu_2} + \delta,$$
where μ_1, μ_2, and ν are defined by (1.2)–(1.4). ☐

Using Corollary 6.1, we see that for some values of the parameters r_1, r_2, r_3, d, l, and p, we can obtain the tight bounds on the minimal noisy error that are given in the Introduction. However, tight bounds for the remaining cases remain an open problem.

7 Two-grid implementation of the noisy modified Galerkin method

We have just developed error estimates for the the noisy modified Galerkin method $\phi_{n,\delta}$. This algorithm has information cost $\mathbf{c}(\delta)\, n$. Unfortunately, its combinatory cost is much worse than $\Theta(n)$, since it involves the solution of a full $n_h \times n_h$ linear system, where

$$n_h \asymp \begin{cases} n^{r_2/(2\min\{r_1,r_2\})} & \text{if } r_2 < 2r, \\ n & \text{if } r_2 \ge 2r. \end{cases}$$

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

$$a = \begin{cases} \dfrac{3s}{2\min\{r_1, r_2\}} & \text{if } r_2 < 2r, \\ 3 & \text{if } r_2 \ge 2r. \end{cases}$$

Since $a \in [\frac{3}{2}, 3]$, the combinatory cost of the noisy modified Galerkin method overwhelms the informational cost.

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 $\mathbf{u}$. Our approach is that of [20], which (in turn) closely follows that of [10].

For given n, we shall define κ, $\bar{\kappa}$, h, and $\bar{h}$ as at the beginning of Section 6. This will give us a linear system $(\mathbf{A} - \mathbf{B})\mathbf{u} = \mathbf{f}$ whose solution we wish to approximate. Following an idea that can be traced back to [8], 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 Galerkin method using information of cardinality n^*, we would get an $n_{h^*} \times n_{h^*}$ linear system $(\tilde{\mathbf{A}} - \tilde{\mathbf{B}})\tilde{\mathbf{u}} = \tilde{\mathbf{f}}$. Here, h^* is h of Section 6, but defined for n^* rather than for n, and n_{h^*} is defined for h^* via relationship (6.1), but with h replaced by h^*.

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 [10]. Let $X = L_p(I^d)$, $X_{n_h} = (\mathbb{R}^{n_h}, \|\cdot\|_{\ell_p})$, and $X_{n_{h^*}} = (\mathbb{R}^{n_{h^*}}, \|\cdot\|_{\ell_p})$. We define the *canonical prolongation* $P_h \colon X_{n_h} \to X$ and

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

The *canonical restriction* $R_h \colon X \to X_{n_h}$ is defined as

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

We then define the *intergrid prolongation operator* $\mathfrak{p} \colon X_{n_{h^*}} \to X_{n_h}$ and the *intergrid restriction operator* $\mathfrak{r} \colon X_{n_h} \to X_{n_{h^*}}$ 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 $\mathfrak{p}^* \colon X_{n_h} \to X_{n_{h^*}}$, defined as

$$\mathfrak{p}^*\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot \mathfrak{p}\mathbf{w} \qquad \forall\, \mathbf{v} \in X_{n_h}, \mathbf{w} \in X_{n_{h^*}}.$$

Our two-grid algorithm is defined in Figure 2.75 This is essentially the same algorithm as we used in [20, Sect. 8], which (in turn) is the variant ZGM$'$ of the two-grid method found on [10, pg. 179].

Let us write

$$\check{u}_{n,\delta} = P_h[\mathrm{TG}(n, \mathbf{A}, \mathbf{B}, \mathbf{f})] = \sum_{j=1}^{n_h} v_j s_{j,h}, \tag{7.1}$$

Suppose that we define *two-grid information* of cardinality at most n as

$$\check{\mathbb{N}}_{n,\delta} = [\mathbb{N}_{n,\delta}, \mathbb{N}_{n^*,\delta}].$$

Then $\check{u}_{n,\delta}$ depends on $[\rho, k, f] \in \mathscr{F}$ only through the information $\check{\mathbb{N}}_{n,\delta}([\rho, k, f])$, and so we may write $\check{u}_{n,\delta} = \check{\phi}_{n,\delta}\big(\check{\mathbb{N}}_{n,\delta}([\rho, k, f])\big)$, 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 main result is then

Theorem 7.1. *Suppose that $d = l$ or $r_1 \geq 2$. Let $m = \max\{r_1, r_2\} - 1$ and let $q \geq r_1 - \nu$ in* (4.6) *and* (4.7)*, where ν is given by* (1.4)*. There exist positive constants M_9 and M_{10}, along with $n_0^* \in \mathbb{Z}^{++}$, such that for any $n \geq n_0^*$ and $\delta \in [0, \delta_0]$, we have*

$$e(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \leq M_9(n^{-\mu_2} + \delta),$$

```
function TG(n : Z^+;  A, B : R^{n_h×n_h}; f : R^{n_h}) : R^{n_h};
begin
  if n* ≤ n*_0 then
    compute u ∈ R^{n_h} such that (A − B)u = f
  else
    begin
      u := 0;
      for i := 1 to 3 do
        begin
          Solve the linear system Aũ = f + Bu;          {Picard iteration}
          d := p*(Aũ − f − Bũ);                          {compute defect}
          solve the system (Ã − B̃)δ = d;                 {coarse-grid solution}
          u := u − pδ                                    {coarse-grid correction}
        end
    end;
  TG := u
end;
```

Figure 2: The two-grid algorithm

where μ_1 *is given by* (1.2)*, with*

$$\mathrm{cost}(\check{\phi}_{n,\delta}, \check{\mathbb{N}}_{n,\delta}) \le M_{10}\, \mathbf{c}(\delta)\, n.$$

Proof. This theorem is the same as [20, Theorem 26], with the more-or-less obvious textual substitutions. Now the proof of [20, Theorem 26] is based on the estimates in the lemmas of [20, Sect. 6]. Hence, we need only replicate the proof of [20, Theorem 26], replacing the estimates in the lemmas of [20, Sect. 6] by those appearing in the corresponding lemmas of Section 5 in this paper, to get a proof of our theorem. ☐

8 Complexity

In this section, we determine the ε-complexity of the noisy Fredholm problem. Recalling the definitions of μ_1, μ_2, and ν from (1.2)–(1.4), our main result is

Theorem 8.1. *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. *If* $d < l$ *and* $r_1 = 1$*, then there exists* $\varepsilon_0 > 0$ *such that*

$$\mathrm{comp}(\varepsilon) = \infty \qquad \forall\, \varepsilon \in [0, \varepsilon_0].$$

2. *Suppose that* $d = l$ *or* $r_1 \ge 2$*.*

   (a) The problem complexity is bounded from below by

$$\mathrm{comp}(\varepsilon) \ge \inf_{0<\delta<C_1\varepsilon} \mathbf{c}(\delta) \left\lceil \left(\frac{1}{C_1\varepsilon - \delta}\right)^{1/\mu_1} \right\rceil.$$

(b) *The problem complexity is bounded from above by*

$$\mathrm{comp}(\varepsilon) \le C_2 \inf_{0<\delta<C_3\varepsilon} \mathbf{c}(\delta) \left\lceil \left(\frac{1}{C_3\varepsilon - \delta} \right)^{1/\mu_2} \right\rceil. \tag{8.1}$$

$$n = \left\lceil \left(\frac{1}{C_3\varepsilon - \delta} \right)^{1/\mu_2} \right\rceil,$$

with $C_3 = M_9^{-1}$ *from Theorem* 7.1 *and where* δ *is chosen to minimize the right hand side of appearing in* (8.1).

Proof. The first part follows from Theorem 3.1. The proof of the second part is identical to that of [20, Thm. 27]. □

For any particular cost function $\mathbf{c}\colon \mathbb{R}^+ \to \mathbb{R}^{++} \cup \{\infty\}$, we can use use the lower and upper bounds in Theorem 8.1 to get specific complexity bounds corresponding to that cost function.

For a cost function c, an error level ε, an exponent a, and a constant C, define $g_{\varepsilon,a,C}\colon \mathbb{R}^{++} \to \mathbb{R}^{++}$ as

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

and set

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

By Theorem 8.1, we see that

$$g^*_{\varepsilon,1/\mu_1,C_1} \le \mathrm{comp}(\varepsilon) \le C_2\, g^*_{\varepsilon,1/\mu_2,C_3}. \tag{8.2}$$

This inequality allows us to determine complexity bounds for any particular cost function $\mathbf{c}$. In particular, if $\mathbf{c}$ is differentiable, then the optimal δ minimizing the left-hand or right-hand sides of (8.2) must satisfy $g'_{\varepsilon,a,C}(\delta) = 0$, i.e., we must have

$$-\frac{\mathbf{c}(\delta)}{\mathbf{c}'(\delta)} = \frac{C\varepsilon - \delta}{a}, \tag{8.3}$$

for $a = 1/\mu_1$ and $a = 1/\mu_2$.

As a specific example, consider the cost function $\mathbf{c}(\delta) = \delta^{-b}$, where $b > 0$. We find that for $\varepsilon > 0$, the optimal δ satisfying (8.3) is

$$\delta^*_{\varepsilon,a} = \frac{C\varepsilon}{a+b}, \tag{8.4}$$

so that

$$g^*_{\varepsilon,a,C} = g_{\varepsilon,a,C}(\delta^*_{\varepsilon,a}) \asymp \frac{(a+b)^{a+b}}{C^{a+b}a^a b^b} \left(\frac{1}{\varepsilon} \right)^{a+b}.$$

Thus we see that the optimal $\delta^*_{\varepsilon,a}$ is proportional to ε, and that

$$\left(\frac{1}{\varepsilon} \right)^{b+1/\mu_1} \preccurlyeq \mathrm{comp}(\varepsilon) \preccurlyeq \left(\frac{1}{\varepsilon} \right)^{b+1/\mu_2}.$$

In particular, if $\mu_1 = \mu_2 = \mu$, then we have tight bounds

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

on the ε-complexity of our Fredholm problem.

Acknowledgments

References

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

[2] A. K. Aziz, M. R. Dorr, and R. B. Kellogg. A new approximation method for the Helmholtz equation in an exterior domain. *SIAM J. Numer. Anal.*, 19(5):899–908, 1982.

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

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

[5] G. M. Constantine and T. H. Savits. A multivariate Faa di Bruno formula with applications. *Trans. Amer. Math. Soc.*, 348:503–520, 1996.

[6] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 2. Wiley, 1989.

[7] C. H. Edwards, Jr. *Advanced Calculus of Several Variables*. Academic Press, New York, 1973.

[8] 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.

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

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

[11] 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.

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

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

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

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

[16] M. H. Schultz. L^∞-multivariate approximation theory. *SIAM J. Numer. Anal*, 6:161–183, 1969.

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

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

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

[20] A. G. Werschulz. Where does smoothness count the most for Fredholm equations of the second kind with noisy information? *J. Complexity*, 19, 2003 (to appear). Also available as `http://www.cs.columbia.edu/˜library/TR-repository/reports/reports-2001/cucs-017-01.pdf`.

[21] A. G. Werschulz and H. Woźniakowsk. Surface approximation is sometimes easier than surface integration. *Constructive Approximation*, 19, 2003 (to appear). Also available as `http://www.cs.columbia.edu/˜library/TR-repository/reports/reports-2001/cucs-018-01.pdf`.

[22] A. G. Werschulz and H. Woźniakowski. What is the complexity of surface integration? *Journal of Complexity*, 17:442–466, 2001.