M. Griebel, T. Jahn

Efficient solution of ill-posed integral equations through averaging

Efficient solution of ill-posed integral equations through averaging

MICHAEL GRIEBEL

Institut für Numerische Simulation, Universität Bonn, Friedrich-Hirzebruch-Allee 7, 53115 Bonn, Germany

Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53754 Sankt Augustin, Germany

AND

TIM JAHN*

Institut für Mathematik, TU Berlin, Straße des 17. Juni 135, 10623 Berlin , Germany
*Corresponding author: jahn@tu-berlin.de

This paper discusses the error and cost aspects of ill-posed integral equations when given discrete noisy point evaluations on a fine grid. Standard solution methods usually employ discretization schemes that are directly induced by the measurement points. Thus, they may scale unfavorably with the number of evaluation points, which can result in computational inefficiency. To address this issue, we propose an algorithm that achieves the same level of accuracy while significantly reducing computational costs. Our approach involves an initial averaging procedure to sparsify the underlying grid. To keep the exposition simple, we focus only on one-dimensional ill-posed integral equations that have sufficient smoothness. However, the approach can be generalized to more complicated two- and three-dimensional problems with appropriate modifications.

Keywords: integral equations; random noise; complexity; error estimates.

1. Introduction

This article discusses integral equations, which have numerous applications. Examples include various types of tomography, such as x-ray, electrical impedance, or magnetic resonance tomography in medical contexts. In geophysical applications and the oil industry, seismic waves are used to detect underground structures, resulting in an inverse problem. Additional examples are the determination of the implicit volatility that drives stock prices in financial mathematics or the identification of the atmospheric state required for weather predictions. In many cases, the desired object of determination is the distribution of a physical parameter, such as density, conductivity, or volatility, for spatially extended objects that are described by numerous parameters. Therefore, the discretization of the integral equation often results in a high number of measured data and degrees of freedom, making the cost of approximative solution algorithms a significant factor. The integral equation must be solved for a function that is not directly accessible but is measured in some way. The measurement model distinguishes between linear information, where Fourier modes or similar functionals of the data are accessible, and point-wise (standard) information, where only point evaluations of the data are available. This article focuses on point-wise information. Finding an appropriate discretization scheme can be challenging and often requires additional a priori information about the problem and the unknown solution. If the required information is unavailable, a common approach is to use the measurement grid as the discretization grid. The main idea of this article is to decrease the size of the initially given fine discretization by averaging. Averaging

data is a common engineering practice with many applications, see [17]. It has been successfully applied in the closely related field of numerical differentiation by finite differences, as shown in [1]. Numerical differentiation is often a good starting point for understanding new approaches to inverse problems, as it captures essential features and challenges while being analytically simple, see [8]. Here, we first introduce our new numerical method for such a model differentiation problem. Unlike [1], we address the problem directly within the framework of regularization theory, which enables the use of various regularization schemes and generalizes to other integral equations. Additionally, we focus solely on the one-dimensional case in this article This approach allows us to present the fundamental concepts and specific properties of our method in a clear and concise manner, without the complications that arise from a higher-dimensional setting. However, as we will discuss in the final section of this article, our approach can be applied to a wider range of problems.

The paper is structured as follows: Section 2 describes the general problem, while Section 3 introduces our novel approach and presents the main results. Specifically, we demonstrate that the optimal error rate can be achieved with significantly reduced cost complexity. We first illustrate this property using a simple example, for which we explicitly know the singular value decomposition of the discretized integral operator. The method is extended to more general settings where additionally quadrature methods are needed to derive approximative decompositions. The subject of adaptive regularization is also discussed. Section 4 presents numerical experiments, while Section 5 provides an outlook on how to generalize the results in various ways. The proofs of the findings are presented in detail in the appendix.

2. Our setting

We will first formulate and analyze our approach for the following exemplary integral equation

$$Kf = g, \tag{2.1}$$

where

$$g(x) = (Kf)(x) = \int_0^1 \kappa(x,y)f(y)\mathrm{d}x, \tag{2.2}$$

with the integral kernel

$$\kappa(x,y) := \min\left(x(1-y), y(1-x)\right). \tag{2.3}$$

The rule (2.2) defines a compact operator $K : L^2(0,1) \to L^2(0,1)$. Obviously, this operator K is not surjective, since $g = Kf$ is differentiable for any $f \in L^2(0,1)$. We assume that the exact g is corrupted by (irregular) measurement noise. More precisely, we deal with the associated inverse problem of reconstructing $f \in L^2$ from noisy point evaluations of $g = Kf$. These $m \in \mathbb{N}$ evaluations are taken on a set of points $\xi_{lm} \in (0,1), l = 1,...,m$ and are given by

$$\left(g_m^{\delta}\right)_{l=1}^m := \left(g(\xi_{lm}) + \delta Z_j\right)_{l=1}^m \in \mathbb{R}^m, \tag{2.4}$$

where $g = Kf$ is the exact (unknown) data, $\delta > 0$ is the noise level and Z_j are i.i.d random variables that are unbiased ($\mathbb{E}[Z_l] = 0$) with finite variance $\mathbb{E}[Z_l^2] = 1$. Moreover, the continuity of κ implies that Kf is continuous for all $f \in L^2[0,1]$, which justifies the use of point evaluations. For simplicity, we assume that the points are uniformly distributed for this example, i.e., $\xi_{lm} := l/(m+1)$, and thus form

a uniform grid. The central task is to solve the equation (1) from the noisy point evaluations of g. This can be formulated as follows:

Given noisy measurements $g_m^\delta \in \mathbb{R}^m$, find (approximate) f.

This is a classic example of an inverse problem. A characteristic challenge of inverse problems is that they are often ill-posed in the Hadamard sense. A problem is said to be well-posed when it admits a unique solution that is continuously dependent on the input data, and it is said to be ill-posed when it is not well-posed. The above infinite dimensional equation (2.1) is ill-posed. While it can be shown that K is injective, the domain of its inverse K^{-1} is strictly less than $L^2(0,1)$ (since K is not surjective), which implies that K^{-1} is not continuous on its domain. Since we have to solve (1) from noisy data, this non-continuity of K^{-1} implies that standard inversion methods are not feasible here, and we have to use regularization. Furthermore, since we have only partial data, we need to discretize. To construct an approximation for f, we proceed as follows: The measurement grid naturally induces a semi-discrete model of K, which we will denote by K_m. It is defined as

$$K_m : L^2(0,1) \to \mathbb{R}^m \tag{2.5}$$

$$f \mapsto ((Kf)(\xi_{lm}))_{l=1}^m = \left(\int \kappa(\xi_{lm}, y) f(y) \mathrm{d}y \right)_{l=1}^m. \tag{2.6}$$

This gives us the semi-discrete equation (with noisy measurements)

$$g_m^\delta = K_m f \tag{2.7}$$

which we have to solve. Obviously, even though K is injective, the semi-discrete operator K_m is not, i.e. solving (2.7) in all of L^2 is not enough to uniquely specify the solution. A natural further choice is to pick from the set of solutions of (2.7) the element of minimal norm. This is equivalent to restricting the solution space to $\mathcal{N}(K_m)^\perp = \mathcal{R}(K_m^*) \subset L^2(0,1)$. Now the ill-posedness of the infinite-dimensional operator K is inherited by the semi-discrete operator K_m in the sense that it is highly ill-conditioned. Thus, one has to resort to regularization methods, such as Tikhonov regularization, Landweber iteration, or spectral cut-off techniques. In this article we concentrate on spectral cut-off, but our results can be generalized to other methods as well.

Spectral cut-off is based on the singular value decomposition of K_m, which we denote by $(v_{jm}, u_{jm}, \sigma_{jm})_{j \le m}$. The singular values $(\sigma_{jm})_{j \le m}$ form a positive decreasing sequence $\sigma_{1m} \ge ... \ge \sigma_{mm} > 0$ and the singular functions $(\tilde{v}_{jm})_{j \le m}$ and vectors $(u_{jm})_{j \le m}$ are orthonormal bases in $\mathcal{N}(K_m)^\perp \subset L^2(0,1)$ and $\mathbb{R}^m$, respectively. In addition, the following two relations hold:

$$K_m v_{jm} = \sigma_{jm} u_{jm} \qquad \text{and} \qquad K_m^* u_{jm} = \sigma_{jm} v_{jm},$$

for all $j = 1, ..., m$, where $K_m^* : \mathbb{R}^m \to L^2(0,1)$ is the adjoint operator implicitly given by the relation $(f, K^*\lambda)_{L^2(0,1)} = (Kf, \lambda)_{\mathbb{R}^m}$, for $f \in L^2(0,1)$ and $\lambda \in \mathbb{R}^m$. We emphasize here that while every compact operator between Hilbert spaces has a singular value decomposition, the concrete singular values and vectors (or functions) for a particular setting are rarely known explicitly and usually have to be approximated numerically somehow. The specific setting we consider here is exceptional in the sense that we can derive the singular value decomposition explicitly. In the second part of the manuscript we study more general integral equations where we have to rely on numerical approximation methods.

Based on the singular value decomposition we now define an approximation to the unknown f via spectral cut-off by

$$f_{k,m}^{\delta} := \sum_{j=1}^{k} \frac{\left(g_m^{\delta}, u_{jm}\right)_{\mathbb{R}^m}}{\sigma_{jm}} v_{jm}, \tag{2.8}$$

where $k = 0, ..., m$ is the truncation index which has to be chosen depending on the measurement g_m^{δ} and the noise level $\delta > 0$. To determine the singular value decomposition needed for $f_{k,m}^{\delta}$ we rely on a result from [14]. There it is shown that the singular value decomposition of the semi-discrete operator $K_m : L^2(0,1) \to \mathbb{R}^m$ is closely related to the eigenvalue decomposition of a representation matrix. In fact, recall that we restricted the search space for the solution of (2.7) to $\mathcal{N}(K_m)^{\perp}$. This space is finite-dimensional and has the form

$$\mathcal{N}(K_m)^{\perp} := \left\{ \sum_{l=1}^{m} \alpha_l \kappa(\xi_{lm}, \cdot) \ : \ \alpha_l \in \mathbb{R}, l = 1, ..., m \right\}.$$

The basis functions $\kappa(\xi_{lm}, \cdot)$ are exactly the Riesz representers of the bounded linear functionals $f \mapsto (Kf)(\xi_{lm})$, since $(Kf)(\xi_{lm}) = (f, \kappa(\xi_{lm}, \cdot))$. Now, expressing the action of K_m in the basis $\{\kappa(\xi_{lm}, \cdot), l = 1, ..., m\}$ yields

$$K_m \left(\sum_{l=1}^{m} \alpha_l \kappa(\xi_{lm}, \cdot) \right) = \sum_{l=1}^{m} \alpha_l \begin{pmatrix} \int \kappa(\xi_{1m}, y) \kappa(\xi_{lm}, y) \mathrm{d}y \\ ... \\ \int \kappa(\xi_{mm}, y) \kappa(\xi_{lm}, y) \mathrm{d}y \end{pmatrix} = T_m \alpha,$$

with the matrix

$$T_m := \left(\int \kappa(\xi_{im}, y) \kappa(\xi_{jm}, y) \mathrm{d}y \right)_{ij} \in \mathbb{R}^{m \times m} \tag{2.9}$$

and the vector $\alpha = \begin{pmatrix} \alpha_1 & ... & \alpha_m \end{pmatrix}^T$. Note that due to the special form of the kernel κ we can explicitly evaluate the entries of the matrix T_m. Furthermore, T_m is symmetric by construction. The following relationship between the eigenvalue decomposition of T_m and the singular value decomposition of K_m is shown in [14]:

Proposition 1   *Let $(\lambda_{jm}, w_{jm})_{j=1}^{m}$ denote the eigenvalue decomposition of T_m, i.e. $T_m = W_m \Lambda_m W_m^T$, where w_{jm} indicates the j-th column of W_m and Λ_m is diagonal with the j-th diagonal entry λ_{jm}. Then, for the singular value decomposition of K_m it holds that*

$$\sigma_{jm} = \sqrt{\lambda_{jm}}, \qquad u_{jm} = w_{jm}, \qquad v_{jm}(\cdot) = \frac{1}{\sigma_{jm}} \sum_{l=1}^{m} (w_{jm})_l \kappa(\xi_{lm}, \cdot),$$

where $(w_{jm})_l$ is the l-th component of the vector w_{jm}.

Consequently, expressing the estimator $f_{k,m}^{\delta}$ in the basis $\{\kappa(\xi_{lm}, \cdot), l = 1, \ldots, m\}$ and using Proposition 1, we get

$$f_{k,m}^{\delta} = \sum_{l=1}^{m} \alpha_l \kappa(\xi_{lm}, \cdot) \qquad \text{with} \qquad \alpha_l := \sum_{j=1}^{k} \frac{(g_m^{\delta}, w_{jm})_{\mathbb{R}^m}}{\lambda_{jm}} (w_{jm})_l. \tag{2.10}$$

We note that Proposition 1 actually holds for any continuous kernel κ and not just for our specific choice of (2.3). The first main contribution of this article is the derivation of a precise bound for the

error $\|f_{k,m}^{\delta} - f\|$, in particular regarding its dependence on the initial measurement grid. This analysis relies critically on the fact that for our setting we have explicit representations of the singular value decompositions of both K and K_m.

Thus we deduce that the regularized approximation (2.8) to the solution of the continuous integral equation (2.1) from noisy point evaluations (2.4) can be constructed as follows: Determine the k leading eigenvalues and vectors of the matrix T_m from (2.9) and use them to compute the coefficients of $f_{k,m}^{\delta}$ in the basis $\{\kappa(\xi_{lm}, \cdot), l = 1, \ldots, m\}$ via (2.10). The concrete choice and the effect of the truncation index k will be discussed below. First, we will analyze the accuracy of $f_{k,m}^{\delta}$ according to (2.8). Then, in the next section, we will present our new modified approach based on averaging, which achieves the same order of accuracy. The main advantage of our new modified approach is that it is based on the matrix T_{m_o} in (2.9) for m_o less than m. This is important when we consider more general settings in Section 3.2, where the matrix T_m and its eigenvalue decomposition must be approximated numerically. We discuss the computational cost there and show that our modified approach achieves the same error rate at a much lower cost.

We will now look at the accuracy of $f_{k,m}^{\delta}$ in more detail. First, our preliminary example with the kernel (2.3) is particularly simple. Here we can not only directly determine T_m, but also its eigenvalue decomposition and the singular value decomposition of K and K_m, which allows us to compute the error $\|f_{k,m}^{\delta} - f\|$ exactly. We take the following result from [14]:

Lemma 2 *The singular values and left singular vectors of K from* (2.1) *are*

$$\sigma_j = (\pi j)^{-2}, \qquad v_j(x) = \sin(j\pi x), \quad j \in \mathbb{N}.$$

The singular values and left singular vectors of K_m are

$$\sigma_{jm} = \frac{1}{4(m+1)^{3/2} \sin^2\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)} \sqrt{1 - \frac{2}{3} \sin^2\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)}$$

and

$$v_{jm}(x) = \frac{1}{\sigma_{jm}^2} \sum_{l=1}^{m} \sin\left(j\pi\xi_{lm}\right) \kappa(\xi_{lm}, x)$$

with $j = 1, \ldots, m$.[1]

In order to derive explicit error bounds for $f_{k,m}^{\delta}$ for the given kernel (2.3), one must impose certain smoothness conditions on f. Usually one assumes that f belongs to some unknown source set $\mathscr{X}_{s,\rho}$ of

---

[1] In general, however, the situation is not so simple: For most practical integral kernels, we can no longer determine T_m exactly. In Section 3.2 below, we therefore consider the setting of a general integral kernel for which T_m is approximated by a quadrature rule. The computational cost of this approximation will also be discussed there.

the form

$$\mathscr{X}_{s,\rho} := \left\{ \varphi_s(K^*K)h \ : \ j \in L^2(0,1), \ \|h\| \le \rho \right\}$$
$$= \left\{ \sum_{i=1}^{\infty} \varphi_s\left(\sigma_i^2\right)(h,v_i)_{L^2} v_i \ : \ h \in L^2, \ \|h\| \le \rho \right\}$$

and $\varphi_s(t) := t^{\frac{s}{2}}$, where $s, \rho > 0$ are unknown parameters. For functions in $\mathscr{X}_{s,\rho}$ we can indeed give quantitative estimates for the approximation error. We discuss their relation to Sobolev smoothness at the beginning of the next section. To determine the estimator $f_{k,m}^{\delta}$, we must first choose the truncation index $k \le m$. The optimal choice for k minimizes the expected distance to $f \in \mathscr{X}_{s,\rho}$ and gives the error

$$\mathrm{err}(\delta,m,s,\rho) := \min_{k=1,\ldots,m} \sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E}\|f_{k,m}^{\delta} - f\|^2 \asymp \left(\frac{\delta^2}{m}\right)^{\frac{4s}{5+4s}} (\rho^2)^{\frac{5}{5+4s}} + \Delta_m, \tag{2.11}$$

where Δ_m denotes the discretization error $\|P_{\mathscr{N}(K_m)}f\|$, where $P_{\mathscr{N}(K_m)}$ denotes the projection onto the null space of K_m in (2.5). This result will be specified and proved below. The k that minimizes (2.11) depends on the smoothness parameters s and ρ, which are usually unknown. In Section 3.3 we discuss an adaptive choice for determining the optimal k using only the noisy point evaluations and the noise level δ. Finally, note that the first term in (2.11) is a universal optimal error bound. This means that no reconstruction method can yield a smaller error over $\mathscr{X}_{s,\rho}$ uniformly, at least if one ignores the multiplicative constant. This was proved for example in [18], where an asymptotically statistically equivalent functional white noise model (with variance δ^2/m) is considered. In this model one starts from noisy linear functionals of the exact data instead of noisy point evaluations, which allows a discretization-free analysis.

3. Main results

In this section we introduce our modification and state the main error bounds. We first consider the special case given in (2.3) in the introduction and then extend to general Fredholm equations.

3.1. *Results for the kernel with known spectral decomposition*

For the kernel given in (2.3), we have the following simple relation between the abstract smoothness defined by the source set $\mathscr{X}_{s,\rho}$ and the classical smoothness, see Proposition 3.9 in [14]. For $f \in \mathscr{X}_{s,\rho}$, if $s > \frac{3}{4}$, then f is differentiable and if $s > \frac{5}{4}$, then f is twice differentiable. Thus, to derive our main error estimates, we assume in this section that

$$f \in \mathscr{X}_{s,\rho} \quad \text{with} \quad s > 3/4. \tag{3.1}$$

Note at this point that this is a substantial smoothness assumption for f, which allows us to derive simple bounds on the discretization error, which measure how well the unknown f can be approximated in $\mathscr{N}(K_m)^{\perp}$. In the next section we will omit it for the treatment of general Fredholm integral equations.

The main idea of our method is that the initial discretization given by the m noisy point evaluations may be unnecessarily fine relative to the data noise δ and the unknown solution f. More precisely, $\mathbb{E}\|f_{k,m}^{\delta} - f\|^2$ is split into three terms, first a data propagation error coming from the noise on the point evaluations, then an approximation error of the projected unknown solution, and finally a discretization

error. While the first two terms depend, among other things, on the level of truncation k, the last term depends only on the parameter of the discretization dimension m. Roughly speaking, if for the minimizing k the first two terms dominate the last term, then we see that the initial discretization based on the design matrix of size $m \times m$ was unnecessarily large. Let us be precise: With $f_m := \sum_{j=1}^{m} (f, v_{jm}) v_{jm}$ denoting the projection of f onto $\mathcal{N}(K_m)^{\perp}$, the error decomposition is

$$
\begin{aligned}
f_{k,m}^{\delta} - f := & \sum_{j=1}^{k} \frac{\left(g_m^{\delta}, u_{jm}\right)}{\sigma_{jm}} v_{jm} - \sum_{j=1}^{k} (f, v_{jm}) v_{jm} - \sum_{j=k+1}^{m} (f, v_{jm}) v_{jm} + f_m - f \\
= & \sum_{j=1}^{k} \frac{\left(g_m^{\delta} - K_m f, u_{jm}\right)}{\sigma_{jm}} v_{jm} - \sum_{j=k+1}^{m} (f, v_{jm}) v_{jm} + f_m - f.
\end{aligned}
$$

From orthogonality we get

$$
\mathbb{E}[\|f_{k,m}^{\delta} - f\|^2] = \delta^2 \sum_{j=1}^{k} \frac{1}{\sigma_{jm}^2} + \sum_{j=k+1}^{m} (f, v_{jm})^2 + \|f_m - f\|^2, \tag{3.2}
$$

and with the source condition (3.1) and Lemma 2 we get the explicit upper bound

$$
\mathbb{E}[\|f_{k,m}^{\delta} - f\|^2] \leq C \left(\frac{k^5}{m} \delta^2 + k^{-4s} \rho^2 + \frac{\|f'\|^2}{m^2} \right) \tag{3.3}
$$

for the error, where $C > 0$ is a constant. This will be shown in the proof of Theorem 3 below. It will also be shown there that the right-hand side is also a lower bound in many cases, up to a constant factor.

In our error bound (3.3) we identify the variance, also called the data propagation error, and the bias part, consisting of an approximation error and a discretization error. As k increases, the first term increases while the second term decreases. As m increases, the first and third terms decrease.

As explained above, our approach is based on the observation that k only affects the first two terms in the error decomposition. It is easy to see that the minimizing k roughly balances the first two terms (and thus satisfies $k^{5+8s} \asymp \frac{m\rho^2}{\delta^2}$). If the contribution of the third term (the discretization error) is much smaller, then the initial discretization dimension m (for the design matrix), i.e. the number of point evaluations, was unnecessarily high. Thus, there is hope to somehow reduce m without spoiling the overall error rate. But directly reducing the parameter m (which would mean simply discarding some of the measurements) increases the first and third terms (with the effect being more pronounced for the third term).

Our proposed strategy is based on averaging components of the measured data, which can keep the size of the first two terms constant while decreasing m. The idea of averaging point evaluations from a fine grid to obtain (approximations of) point evaluations on a coarser grid is borrowed from the recent preprints [12, 15], where a combination of a data-driven regularization method and an adaptive choice of the discretization dimension, called discretization-adaptive regularization [11, 13], was numerically tested and performed well. In the present article we shed light on this observed phenomenon and give a rigorous justification. The reasoning is that if not too many components are averaged, then due to the smoothness of the uncorrupted data g, a reduction of the stochastic noise on the point evaluations is obtained. Of course, this depends on the concrete relationship between δ, m and the unknown data g.

We explore this approach explicitly. To do so, we define for $o \in \mathbb{N}$, so that $m_o := m/o \in \mathbb{N}$, the averaged data

$$\overline{g}_{m_o}^\delta = \left(\frac{\sum_{j=1}^l \left(g_m^\delta \right)_{(i-1)l+j}}{l} \right)_{i=1,\dots,m_o} \in \mathbb{R}^{m_o}.$$

Similarly, we define our averaged spectral cut-off estimator as

$$\overline{f}_{k,m_o}^\delta := \sum_{j=1}^k \frac{\left(\overline{g}_{m_o}^\delta, u_{jm_o} \right)_{\mathbb{R}^{m_o}}}{\sigma_{jm_o}} v_{jm_o} = \sum_{l=1}^{m_o} \overline{\alpha}_l \kappa(\xi_{lm_o}, \cdot) \qquad \text{with} \qquad \overline{\alpha}_l := \sum_{j=1}^k \frac{\left(\overline{g}_{m_o}^\delta, w_{jm_o} \right)_{\mathbb{R}^{m_o}}}{\lambda_{jm_o}} (w_{jm_o})_l.$$

We briefly mention here that the calculation of the averaged data $\overline{g}_{m_o}^\delta$ costs exactly m (regardless of the value of l), since each entry of the initial data g_m^δ is touched once. This is cheap compared to the cost we face when we have to approximate T_m and its eigenvalue decomposition for general kernels in the next section. We define the minimax error of the averaged estimator as

$$\overline{\mathrm{err}}(\delta, m_o, s, \rho) := \min_{k=1,\dots,m} \sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E} \| \overline{f}_{k,m_o}^\delta - f \|^2.$$

The following theorem tells us how much data we can average without spoiling the overall error rate.

Theorem 3   *For $o_m := \lfloor \sqrt{\frac{m\delta^2}{\rho}} \rfloor \geq 1$ and $m_o = m/o_m$ there holds*

$$c \cdot \mathrm{err}(\delta, m, x, \rho) \leq \overline{\mathrm{err}}(\delta, m_o, s, \rho) \leq C \cdot \mathrm{err}(\delta, m, s, \rho) \tag{3.4}$$

with constants

$$c := \frac{\frac{16}{5} + \frac{1}{3\pi^{4(s+1)}}}{\frac{3\pi^4}{5} + \frac{3^{1+s}}{2^{4s-1}} + \frac{1}{2}} \qquad \text{and} \qquad C := \frac{6\pi^4 + \frac{3^{s+1}}{2^{4s-1}}}{\frac{16}{5} + \frac{1}{3\pi^{4(s+1)}}}.$$

Note that finding the number of components $o_m = \sqrt{m}\delta^2/\rho^2$ to average is a problem in practice, since ρ is typically unknown. As demonstrated and explained below in the numerical section, an alternative reasonable choice for o_m would be $o_m = \left(\frac{m^2\delta^2}{\|g'\|^2} \right)^{\frac{1}{3}}$. Note that determining $\|g'\|$ is itself an inverse problem, but a fairly simple one.

3.2. Extension to general Fredholm integral equations

We now explain how our approach can be applied to various Fredholm integral equations. Consider the general integral equation

$$Kf = g, \tag{3.5}$$

where

$$g(x) = (Kf)(x) = \int_0^1 \kappa(x,y) f(y) \mathrm{d}y. \tag{3.6}$$

The integral kernel κ is assumed to be 2-times differentiable. The adjoint of K is defined as $(K^*g)(y) = \int_0^1 k(x,y) g(x) \mathrm{d}x$. Again, discretization on a grid given by the points $\xi_{1m}, \dots, \xi_{mm}$ yields a semi-discrete

operator, which we normalize this time. We then have

$$K_m : L^2(0,1) \to \mathbb{R}^m \tag{3.7}$$

$$f \mapsto ((Kf)(\xi_{jm}))_{j=1}^m = \frac{1}{\sqrt{m}} \left(\int \kappa(\xi_{jm}, y) f(y) \mathrm{d}y \right)_{j=1}^m.$$

We emphasize that, in general, we can no longer evaluate the integrals exactly. Instead, we use midpoint collocation as a quadrature rule, i.e. we have weights $q_{1m} = \ldots = q_{mm} = m^{-1}$ and knots $\xi_{jm} := \frac{2j-1}{2m}$, $j = 1, \ldots, m$. We define $A_m \in \mathbb{R}^m$ by $(A_m)_{ij} = q_{jm} \kappa(\xi_{im}, \xi_{jm}) = \frac{1}{m} \kappa(\xi_{im}, \xi_{jm})$. This method is taken from the popular open source toolbox [9] and is used in the numerical examples in the following section.[2] The midpoint collocation method has degree of exactness two, that is, for any $h \in \mathscr{C}^2([0,1])$, there holds

$$\left| \int h(x) \mathrm{d}x - \frac{1}{m} \sum_{i=1}^m h(\xi_{im}) \right| \leq \frac{1}{24} \frac{\|h^{(2)}\|_\infty}{m^2}.$$

Therefore, we make the following assumption about the smoothness of the integral kernel κ and the unknown solution f.

Assumption 1 *We assume that that $f \in \mathscr{C}^2([0,1])$ and $\kappa \in \mathscr{C}^2([0,1]^2)$ and we set*

$$C_f := \sup_x \max_{n \leq 2} |f^{(n)}(x)| \qquad and \qquad C_K := \sup_{x,y} \max_{\substack{n,n' \leq 2 \\ n+n' \leq 2}} |\partial_x^n \partial_y^{n'} k(x,y)|. \tag{3.8}$$

Consequently, for T_m the design matrix as defined in (2.9) the previous section, but with a general kernel κ satisfying the above assumption 1, there holds

$$\left(A_m^T A_m\right)_{ij} - \frac{1}{m}\left(T_m\right)_{ij} = \frac{1}{m} \left(\frac{1}{m} \sum_{l=1}^m \kappa(\xi_{im}, \xi_{lm}) \kappa(\xi_{jm}, \xi_{lm}) - \int \kappa(\xi_{im}, y) \kappa(\xi_{lm}, y) \mathrm{d}y \right) = \mathscr{O}\left(\frac{1}{m^3}\right).$$

In light of the above results, we base our approximation on the discrete singular value decomposition of $A_m \in \mathbb{R}^{m \times m}$ (which is effectively the eigenvalue decomposition of $A_m^T A_m \approx T_m$), denoted by $(\tilde{z}_{jm}, \tilde{w}_{jm}, \tilde{\sigma}_{jm})$. Note that unlike in the previous section, where we could give closed-form expressions for the eigenvalues and eigenvectors of T_m, here we have to rely on numerical algorithms that only approximate the singular value decomposition of A_m. Since there are numerical algorithms that approximate the singular value decomposition to machine precision, we will not distinguish between the exact singular value decomposition of A_m and its numerically approximated counterpart. However, we will discuss the computational cost of the numerical approximation in detail below. We define

$$\tilde{v}_{jm} := \frac{1}{\tilde{\sigma}_{jm}\sqrt{m}} \sum_{l=1}^m (\tilde{w}_{jm})_l \kappa(\xi_{lm}, \cdot) \in L^2(0,1), \tag{3.9}$$

which approximate the singular function v_{jm} of (3.7) (for j small enough). We mention here that, in contrast to the exact singular functions v_{jm}, the functions $\tilde{v}_{jm}$ form an orthonormal basis only approximately and for j small, see Lemma 6 below. We now use (3.9) to denote our (discrete) spectral cutoff

---

[2] Clearly, the results can be generalized, and we will discuss later the possible advantages of higher-order quadrature rules.

estimator for the unknown f as

$$\tilde{f}_{k,m}^{\delta} := \sum_{j=1}^{k} \frac{(g_m^{\delta}, \tilde{w}_{jm})_{\mathbb{R}^m}}{\sqrt{m} \tilde{\sigma}_{jm}} \tilde{v}_{jm}. \tag{3.10}$$

Similar to the previous section, expressing the estimator $\tilde{f}_{k,m}^{\delta}$ in the basis $\{\kappa(\xi_{lm}, \cdot), \ l = 1, ..., m\}$ yields

$$\tilde{f}_{k,m}^{\delta} = \sum_{l=1}^{m} \tilde{\alpha}_l \kappa(\xi_{lm}, \cdot) \qquad \text{with} \qquad \tilde{\alpha}_l = \sum_{j=1}^{k} \frac{(g_m^{\delta}, \tilde{w}_{jm})}{m \tilde{\sigma}_{jm}^2} (\tilde{w}_{jm})_l. \tag{3.11}$$

Before analyzing the accuracy of the estimator $\tilde{f}_{k,m}^{\delta}$, we discuss its computational cost. If the k leading singular values and vectors of A_m can be derived explicitly, the k quantities $(g_m^{\delta}, \tilde{w}_{jm})_{\mathbb{R}^m}/m\tilde{\sigma}_{jm}^2$ cost $2m$ operations each, and the m coefficients $\tilde{\alpha}_l$ cost $2km$ operations altogether. Next, we discuss the cost of computing the leading singular values and vectors of A_m numerically. Highly stable and accurate methods for doing this first transform the matrix A_m into a tridiagonal matrix, then compute the full singular value decomposition of this matrix, and finally keep only the leading k values and vectors, see e.g. [3]. These algorithms typically require $\mathcal{O}(m^3)$ operations and are therefore very expensive. Since we are only interested in the k leading values and vectors, it is natural to consider methods that determine them directly. Typically, these methods are much cheaper, but also less accurate. We discuss some of them below.

If matrix-vector multiplications can be computed quickly, Krylov subspace methods such as the Lanczos or Arnoldi algorithm are commonly used. However, summarizing the computational costs with corresponding accuracy guarantees is challenging because these methods are inherently numerically unstable. The specific costs depend heavily on the matrix properties and the effort spent to stabilize the routines. For more information, refer to the survey article [7]. As a (usually overly optimistic) rule of thumb, the typical cost of such methods is proportional to $kC_m^{\text{mult}} + mk^2$, where C_m^{mult} denotes the cost of an exact or at least approximate matrix-vector multiplication with the initial matrix A_m. Apart from the setting with sparse matrix A_m, fast multipole expansions [4] or H (and H^2) matrix arithmetic [6] can reduce C_m^{mul} to the order of $\log(m)m$ or even m. However, this requires the kernel to be asymptotically smooth. Also, as mentioned above, additional work is required to stabilize the resulting methods, and exact error guarantees are rare. Therefore, we focus our discussion on the use of provably stable two-step procedures to determine the truncated spectral decomposition, see [7] and [16] for overviews. These procedures work as follows: First, an approximate basis of size $n < m$ for the range of A_m is computed and used to determine a rank-n approximation of A_m, which we call $A_m^{(n)}$. Second, the full singular value decomposition of this approximation $A_m^{(n)}$ is computed by established methods [3] and the leading k values and vectors are used to approximate $\tilde{\sigma}_{jm}$ and $\tilde{w}_{jm}$ for $j = 1, \ldots, k$. To do this, the auxiliary parameter n must be chosen much larger than the target rank k to ensure that the leading k singular values and vectors of $A_m^{(n)}$ approximate the leading eigenvalues and vectors of A_m with sufficient accuracy, see (3.12) below[3]. Several methods are available to perform the first step, e.g. based on rank-revealing QR decompositions [5], random projections [19], or cross/skeleton approximations [2, 20]. Using rank-revealing QR-pivoting or random projections for the first step requires operations on the order of nm^2. In the case of the cross/skeleton approximation, this can go down to $\mathcal{O}(nm)$, but it requires that the integral kernel be approximately smooth. The second step, the determination

---

[3] We emphasize that this is in sharp contrast to many classical settings where the goal is simply to find a near-optimal rank-k approximation of A_m. Here typically $n = k + p$ where p is a small oversampling parameter.

of the full singular value decomposition of $A_m^{(n)}$, costs $\mathcal{O}(n^2 m)$ and thus the total cost is of the order of $\mathcal{O}\left(nC_m + mn^2\right)$, where C_m depends on the approximation approach and the regularity property of the specific kernel. Overall, depending on the actual kernel κ, the computational cost for (2.10) is somewhere between $n^2 m$ and nm^2. This brings us to the choice of n. Let $\tilde{w}_{jm}^{\text{approx}}$ and $\tilde{\sigma}_{jm}^{\text{approx}}$ denote the j-th singular value and right singular vector of $A_m^{(n)}$, and let $\tilde{f}_{k,m}^{\delta,\text{approx}}$ and $\tilde{\alpha}^{\text{approx}} \in \mathbb{R}^m$ be defined as in (3.11). Theorem 4.2 from [10] shows that for k not too large

$$\frac{\left\|\tilde{f}_{k,m}^{\delta,\text{approx}} - \tilde{f}_{k,m}^{\delta}\right\|}{\left\|\tilde{f}_{k,m}^{\delta}\right\|} = \mathcal{O}\left(\frac{\|A_m - A_m^{(n)}\|}{\tilde{\sigma}_{km}^2}\right). \tag{3.12}$$

Furthermore, for the two-step procedure, the total approximation error is usually about[4]

$$\|A_m - A_m^{(n)}\| = \mathcal{O}\left(\sqrt{nm}\,\tilde{\sigma}_{n+1m}\right). \tag{3.13}$$

So the value n must be chosen so that at least $\tilde{\sigma}_{n+1m} = o\left(\tilde{\sigma}_{km}^2 (mn)^{-\frac{1}{2}}\right)$. For example, suppose[5] $\tilde{\sigma}_{jm}^2 \sim j^{-q}$ with $q > 1$, and thus approximating $\tilde{f}_{k,m}^{\delta}$ up to the relative error ε requires $\tilde{\sigma}_{n+1m} \sim \varepsilon \tilde{\sigma}_{km}^2 (nm)^{-\frac{1}{2}}$, which implies $n \sim k^{\frac{2q}{q-1}} \varepsilon^{-\frac{2}{q-1}} m^{\frac{1}{q-1}}$. Therefore, between $k^{\frac{4q}{q-1}} \varepsilon^{-\frac{4}{q-1}} m^{1+\frac{2}{q-1}}$ and $k^{\frac{2q}{q-1}} \varepsilon^{-\frac{2}{q-1}} m^{2+\frac{1}{q-1}}$ operations are necessary to determine $\tilde{f}_{k,m}^{\delta}$ up to the relative error ε. In the following, we derive the optimal choice for k that minimizes the error $\|\tilde{f}_{km}^{\delta} - f\|$. This choice depends on m, the noise level δ, and the regularity of the exact solution. [6]

As a next main result, we compute the error of our estimator. Since we have explicitly taken into account the discretization by a quadrature rule, it is natural that additional assumptions, somewhat uncommon in the framework of classical infinite-dimensional regularization theory, have to be imposed on the unknown f in order to perform a meaningful error analysis, see Assumption 1.

For this purpose, the following auxiliary lemma states that the first m singular values σ_j of the continuous operator K are close to the singular values $\tilde{\sigma}_{jm}$ of the quadrature matrix A_m.

Lemma 4 *Let $(\tilde{\sigma}_{jm})_{j \leq m}$ be the singular values of $(A_m)_{ij \leq m} = \frac{1}{m}\kappa(\xi_{im}, \xi_{jm})$. Then, it holds that*

$$\left|\sigma_j^2 - \tilde{\sigma}_{jm}^2\right| \leq \frac{C_K^2}{3m^2}$$

for $j = 1, ..., m$.

Before formulating the main results of this paper, we introduce two important auxiliary lemmata. The first relates approximate eigenvectors, i.e. vectors v satisfying $\|K^*Kv - \lambda v\| \approx 0$ for some $\lambda > 0$, to exact ones.

---

[4] Depending on the decay and gaps of the spectrum, in some cases the first factor can be reduced to logarithmic terms.

[5] In contrast to the previous section, there should be no significant dependence on m, since K_m and A_m are normalized here.

[6] We will then present a modified spectral cut-off estimator with drastically reduced computational cost. The cost reduction is due to the fact that the modified estimator is based on $A_{\tilde{m}}$ with $\tilde{m} < m$.

Lemma 5 *Let $K : \mathscr{X} \to \mathscr{X}$ be an endomorphism with orthornomal eigenbasis $(v_i)_{i\in\mathbb{N}}$ and corresponding eigenvalues $(\lambda_i)_{i\in\mathbb{N}}$. Suppose there are $v \in \mathscr{X}$ and $\lambda, \varepsilon \in \mathbb{R}$ such that $\|Kv - \lambda v\| \leq \varepsilon$. Then we have*

$$\min_{i\in\mathbb{N}} |\lambda_i - \lambda| \leq \frac{\varepsilon}{\|v\|}. \tag{3.14}$$

Furthermore, let $j \in \arg\min_{i\in\mathbb{N}} |\lambda_i - \lambda|$ and assume that $\min_{i\neq j} |\lambda_i - \lambda| \geq c > 0$ for some $c \in \mathbb{R}$. Then we have

$$\|v\|^2 \geq (v, v_j)^2 \geq \|v\|^2 - \frac{\varepsilon^2}{c^2}. \tag{3.15}$$

The next lemma shows that the constructed functions $\tilde{v}_{jm}$ in (3.9), based on the (discrete) eigenvectors of $A_m^T A_m$, are indeed approximate eigenfunctions of the continuous operator K^*K, if j is not too large. We define $c_j := \min_{i\neq j} |\sigma_i^2 - \sigma_j^2|$ and set

$$J_m := \max\left\{ l \geq 1 \,:\, c_l \geq \frac{2C_K^2}{3m^2} \right\}. \tag{3.16}$$

Since $\sigma_l^2 \geq c_l$, it is clear by Lemma 5 for $j \leq J_m$ that

$$\tilde{\sigma}_{jm}^2 \geq \sigma_j^2 - |\tilde{\sigma}_{jm}^2 - \sigma_j^2| \geq \sigma_j^2 - \frac{C_K^2}{3m^2} \geq \frac{\sigma_j^2}{2}.$$

We also have the following result:

Lemma 6 *Recall that $(A_m)_{ij} = \frac{1}{m}\kappa(\xi_{im}, \xi_{jm})$ and $\tilde{v}_{jm} := \frac{1}{\tilde{\sigma}_{jm}\sqrt{m}} \sum_{l=1}^m (\tilde{w}_{jm})_l \kappa(\xi_{lm}, \cdot)$ where $(\tilde{\sigma}_{jm}, \tilde{z}_{jm}, \tilde{w}_{jm})$ is the singular value decomposition of A_m. Then*

$$|(\tilde{v}_{jm}, \tilde{v}_{i,m}) - \delta_{ij}| \leq \frac{C_K^2}{3\sigma_j\sigma_i m^2} \tag{3.17}$$

and

$$\|\tilde{v}_{jm}\|^2 \geq (\tilde{v}_{jm}, v_j)^2 \geq \|\tilde{v}_{jm}\|^2 - \frac{C_K^4}{c_j^2\sigma_j^2 m^4}. \tag{3.18}$$

In particular, it holds

$$1 + \frac{C_K^2}{\sigma_j^2 m^2} \geq (\tilde{v}_{jm}, v_j)^2 \geq 1 - \frac{2C_K^2}{3\sigma_j^2 m^2} - \frac{C_K^4}{c_j^2\sigma_j^2 m^4}. \tag{3.19}$$

We are now ready to give a bound on the total error.

Theorem 7 *Assuming that Assumption 1 is fulfilled, we set $c_j := \min_{i \neq j} |\sigma_i - \sigma_j|$ and*

$$J_m := \max \left\{ j \geq 1 \, : \, m \geq \max \left(\sqrt{\frac{2C_K^2}{3c_j}}, \sqrt{\frac{j}{c_j \sigma_j^2}} \right) \right\}.$$

Then

$$\sqrt{\mathbb{E} \| f_{k,m}^{\delta} - f \|^2} \leq \frac{2\delta}{\sqrt{m}} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty} (f, v_j)^2} + \frac{C_1}{m^2} \sqrt{\sum_{j=1}^{k} \frac{1}{c_j^2 \sigma_j^2}}$$

for all $k \leq J_m$, with

$$C_1 := \sqrt{6} C_f^2 + \sqrt{2} C_K^3 \|f\| + \sqrt{3} C_K^2 \|f\| \tag{3.20}$$

and the constants from (3.8).

Let us comment on the result. First, only low-frequency singular vectors can be approximated with sufficient accuracy by the quadrature rule, since the high-frequency ones oscillate too much. This gives the constraint $k \leq J_m$. Second, the smoothness condition for f ensures that the error of the quadrature rule for the coefficients (f, v_j) can be bounded. Such conditions were not needed for the special case treated in Section 2, since there we could compute the matrix T_m exactly and thus did not need to use a quadrature rule. Note again that a source condition could be included to control $\sum (f, v_j)^2$. Finally, the dependence on the quantity c_j has technical reasons, see Lemma 5, which relates the discrete singular vectors to the continuous ones. In particular, we see that the result becomes useless when the operator K has singular values with multiplicity greater than one. Probably the proofs could be modified to cover the latter cases as well.

We now have the error for our computational estimator $\tilde{f}_{k,m}^{\delta}$ from (3.10) based on the m initial measurements g_m^{δ} and the quadrature matrix $A_m \in \mathbb{R}^{m \times m}$. As in the previous section, we identify a variance term, an approximation term, and a discretization term. Compared to Theorem 3, however, the situation here is more delicate, since the discretization error now depends on the truncation level k. This is due to the fact that we have taken the error from the quadrature rule into account. Nevertheless, we may still encounter the situation where the initial discretization was too fine, in the sense that if we consider the k, such that the (square root of the) variance $\frac{2\delta}{\sqrt{m}} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2}}$ is balanced by the approximation error $\sqrt{\sum_{j=k+1}^{m} (f, v_j)^2}$, this k can lead to a discretization error $\frac{C_1}{m^2} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2 c_j^2}}$ of smaller order. Therefore, we again average the m-th initial measurements to obtain m_o new measurements, and base the estimator on the singular value decomposition $(\tilde{\sigma}_{j m_o}, \tilde{z}_{j m_o}, \tilde{w}_{j m_o})$ of $A_{m_o} \in \mathbb{R}^{m_o \times m_o}$ and the $\tilde{v}_{j,m_o}$ from (3.9). So we consider the estimator

$$\overline{\tilde{f}}_{k,m_o}^{\delta} := \sum_{j=1}^{k} \frac{(\overline{g}_{m_o}^{\delta}, \tilde{w}_{j m_o})_{\mathbb{R}^m}}{\sqrt{m} \tilde{\sigma}_{j m_o}} \tilde{v}_{j,m_o} = \sum_{l=1}^{m_o} \overline{\tilde{\alpha}}_l \kappa(\xi_{l m_o}, \cdot) \qquad \text{where} \qquad \overline{\tilde{\alpha}}_l = \sum_{j=1}^{k} \frac{(\overline{g}_{m_o}^{\delta}, \tilde{w}_{j m_o})}{m_o \tilde{\sigma}_{j m_o}^2} (\tilde{w}_{j m_o})_l, \tag{3.21}$$

with averaged data

$$\overline{g}_{m_o}^{\delta} := \left(\frac{1}{o} \sum_{i=1}^{o} \left(g_m^{\delta} \right)_{o(j-1)+i} \right)_{j=1}^{m_o} \in \mathbb{R}^{m_o}.$$

As will be seen in the proofs, the main difference in the error analysis compared to the unaveraged estimator $\tilde{f}_{k,m}$ is the systematic data error introduced by averaging. In the special case with kernel (2.3) treated in Section 3.1, this error was roughly $\|g'\|/m$. Now we will obtain an even faster decay of this systematic component due to the additional smoothness assumptions for the solution f and the kernel κ and the special geometry of the discretization grid.

Theorem 8   *For $m_o = \frac{m}{o} \in \mathbb{N}$, assume that Assumption 1 is satisfied, set $c_j := \min_{i \neq j} |\sigma_i - \sigma_j|$ and set*

$$J_{m_o} := \max \left\{ j \geq 1 \ : \ m_o \geq \max \left(\sqrt{\frac{2C_K^2}{3c_j}}, \sqrt{\frac{j}{c_j \sigma_j^2}} \right) \right\}.$$

Then,

$$\sqrt{\mathbb{E} \left\| \overline{\tilde{f}}_{k,m_o}^{\delta} - f \right\|^2} \leq \frac{2\delta}{\sqrt{m}} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2}} + \frac{\|g''\|_\infty}{12 m_o^2 \sigma_k} + \frac{C_K \|g''\|_\infty}{6\sqrt{6} m_o^3} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^4}} + \frac{C_1}{m_o^2} \sqrt{\sum_{j=1}^{k} \frac{1}{c_j^2 \sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty} (f, v_j)^2}$$

for all $k \leq J_{m_o}$, where C_1 is given in (3.20).

We compare our upper bound in Theorem 8 for the averaged estimator with the upper bound in Theorem 7 for the non-averaged one. First, we observe that for $m_o, k \to \infty$ holds

$$\frac{\|g''\|_\infty}{12 m_o^2 \sigma_k} + \frac{C_K \|g''\|_\infty}{6\sqrt{6} m_o^3} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^4}} \ll \frac{C_1}{m_o^2} \sqrt{\sum_{j=1}^{k} \frac{1}{c_j^2 \sigma_j^2}}.$$

Consequently, if the initial discretization was too fine, i.e,

$$\frac{2\delta}{\sqrt{m}} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty} (f, v_j)^2} \gg \frac{C_1}{m^2} \sqrt{\sum_{j=1}^{k} \frac{1}{c_j^2 \sigma_j^2}},$$

and if m_o is not too small, we see that the upper bounds in Theorem 7 and Theorem 8 are asymptotically the same. In principle, it should be possible to obtain a stronger result here, comparable to Theorem 3. However, we leave this as future work, since clearly some additional restrictions are needed to obtain an exact lower bound for terms like

$$\sup_{\substack{f \in \mathscr{X}_{s,\rho} \\ \|f^{(q)}\|_\infty \leq \rho'}} \mathbb{E} \|f_{k,m}^{\delta} - f\|^2.$$

Looking again at the example where $\tilde{\sigma}_{jm}^2 \sim j^{-q}$, we see that approximating $\tilde{f}_{k,m}^{\delta}$ to a relative error of $\varepsilon > 0$ is more expensive by a factor between $o^{1 + \frac{2}{q-1}}$ and $o^{2 + \frac{1}{q-1}}$ than approximating $\overline{\tilde{f}}_{k,m_o}^{\delta}$ to the same relative accuracy.

3.3. *Adaptivity*

Finally, we consider the question of adaptivity, which concerns the concrete choice of the truncation level k and the discretization dimension m_o. Since properties of the unknown solution f, such as the particular smoothness φ_s or the norm of the source element ρ, are usually unknown, strategies are needed that depend only on the measurements g_m^δ and possibly on the noise level $\delta > 0$. Obviously, one is interested in adaptively finding k and m_o such that the error $\|\overline{f}_{k,m_o}^\delta - f\|$ reaches the infeasible optimal choice $\arg\min_k \|f_{k,m}^\delta - f\|$ by at least a multiplicative factor (for $\overline{\tilde{f}}_{k,m_o}^\delta$ and $\tilde{f}_{k,m}^\delta$, respectively). A popular method to determine a regularization parameter in a data-driven manner is the discrepancy principle. This principle follows the paradigm that the regularization parameter should be chosen such that the residual norm of the candidate approximation is approximately the same size as the data error. In contrast to classical settings where the discretization dimension is fixed and usually only one regularization parameter has to be chosen, here we need a strategy to determine both the truncation level k and the dimension m_o. In the following, we propose a modified discrepancy principle, which can be seen as a multiscale method. For simplicity, we restrict ourselves to the case where $m = a^n$ for some $a, n \in \mathbb{N}$. The expected squared data error is

$$\mathbb{E}\left[\|\overline{g}_{m_o}^\delta - g_{m_o}\|_{\mathbb{R}^{m_o}}^2\right] = \mathbb{E}\left[\left\|\overline{g}_{m_o}^\delta - \overline{g}_{m_o}\right\|_{\mathbb{R}^{m_o}}^2\right] + \left\|\overline{g}_{m_o} - g_{m_o}\right\|_{\mathbb{R}^{m_o}}^2 = \delta^2 \mathbb{E}\left\|\begin{pmatrix}\bar{Z}_1\\\vdots\\\bar{Z}_{m_o}\end{pmatrix}\right\|_{\mathbb{R}^{m_o}}^2 + \left\|\overline{g}_{m_o} - g_{m_o}\right\|_{\mathbb{R}^{m_o}}^2$$

$$= \frac{m_o}{o}\delta^2 + \left\|\begin{pmatrix}\frac{\sum_{t=1}^l g(\xi_{tm})}{o} - g(\xi_{1m_o})\\\vdots\\\frac{\sum_{t=1}^l g(\xi_{(m-1)o+t,m})}{o} - g(\xi_{m_o m_o})\end{pmatrix}\right\|_{\mathbb{R}^{m_o}}^2, \tag{3.22}$$

and the residual norm is $\|K_m \overline{f}_{k,m_o}^\delta - \overline{g}_{m_o}^\delta\|_{\mathbb{R}^{m_o}}$. First, note that for $\kappa(x,y) = \min(x(1-y), y(1-x))$ and a uniform grid, we can compute the residual norm exactly as

$$\|K_m \overline{f}_{k,m_o}^\delta - \overline{g}_{m_o}^\delta\|_{\mathbb{R}^{m_o}}^2 = \sum_{j=k+1}^{m_o} (\overline{g}_{m_o}^\delta, u_{jm_o})_{\mathbb{R}^{m_o}}^2.$$

It is monotonically decreasing in k and is zero for $k = m_o$. The systematic error (the second term in (3.22)) can be estimated as in (A.7) below, and we get

$$\|\overline{g}_{m_o} - g_{m_o}\|_{\mathbb{R}^{m_o}}^2 \leq \frac{o}{m}\|g'\|^2.$$

Consequently, for $\kappa(x,y) = \min(x(1-y), y(1-x))$, for a uniform grid, for a fixed dimension m_o and for the parameter $\tau > 1$, we define the truncation level determined by the discrepancy principle as

$$k_{\text{dp}}^\delta(m_o) := \max\left\{k \geq 0 \,:\, \|K_m \overline{f}_{k,m_o}^\delta - \overline{g}_{m_o}^\delta\|_{\mathbb{R}^{m_o}} > \tau\sqrt{\frac{\|g'\|^2}{m_o} + \frac{m_o}{o}\delta^2}\right\}. \tag{3.23}$$

This choice is intended to balance the residual norm and the (expected) data error. In the case of a general kernel κ (fulfilling assumption 1) and the discretization by midpoint collocation (we assume

that $\dim(\mathscr{R}(A_{m_o}) = m_o$, which implies that $(\tilde{w}_{jm_o})_j$ forms an orthonormal basis of $\mathbb{R}^{m_o}$). Then we have

$$
\left(K_m\overline{\tilde{f}}^{\delta}_{k,m_o}\right)_i = \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{\sqrt{m}\tilde{\sigma}_{jm_o}} (K_m\tilde{v}_{jm_o})_i
$$

$$
= \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{\sqrt{m}\tilde{\sigma}_{jm_o}} \frac{1}{\sqrt{m}\tilde{\sigma}_{jm_o}} \sum_{t=1}^{m_o} (\tilde{w}_{jm_o})_t \int \kappa(\xi_{tm_o},y)\kappa(\xi_{im_o},y)\mathrm{d}y
$$

and if we also estimate the integral by the midpoint collocation quadrature, we get

$$
\sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{\sqrt{m}\tilde{\sigma}_{jm_o}} \frac{1}{\sqrt{m}\tilde{\sigma}_{jm_o}} \sum_{t=1}^{m_o} (\tilde{w}_{jm_o})_t \frac{1}{m} \sum_{s=1}^{m_o} \kappa(\xi_{tm_o},\xi_{sm_o})\kappa(\xi_{im_o},\xi_{sm_o})
$$

$$
= \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{\tilde{\sigma}_{jm_o}} \frac{1}{m} \sum_{s=1}^{m_o} (\tilde{z}_{jm_o})_s \kappa(\xi_{im_o},\xi_{sm_o}) = \sum_{j=1}^{k} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}} (\tilde{w}_{jm_o})_i,
$$

with quadrature error

$$
\left| (K_{m_o}\overline{\tilde{f}}^{\delta}_{k,m_o})_i - \sum_{j=1}^{k} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}} (\tilde{w}_{jm_o})_i, \right| \le \left| \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{m_o\tilde{\sigma}^2_{jm_o}} \sum_{t=1}^{m_o} (\tilde{w}_{jm_o})_t \left[\int \kappa(\xi_{tm_o},y)\kappa(\xi_{im_o},y)\mathrm{d}y \right. \right.
$$

$$
\left. \left. - \frac{1}{m} \sum_{s=1}^{m_o} \kappa(\xi_{tm_o},\xi_{sm_o})\kappa(\xi_{im_o},\xi_{sm_o}) \right] \right|
$$

$$
\le \frac{C_K^2}{6m_o^2} \left| \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}}}{m_o\tilde{\sigma}^2_{jm_o}} \sum_{t=1}^{m_o} (\tilde{w}_{jm_o})_t \right| \le \frac{C_K^2}{6m_o^3\sigma^2_{k,m_o}} \left| \sum_{j=1}^{k} (\overline{g}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}} \right| \left| \sum_{t=1}^{m_o} (\tilde{w}_{jm_o})_t \right|
$$

$$
\le \frac{C_K^2\|\overline{g}_{m_o}\|_{\mathbb{R}^{m_o}}}{3m_o^2\sigma^2_k}.
$$

Now, since $(\tilde{w}_{jm_o})_j$ is an orthonormal basis in $\mathbb{R}^{m_o}$, we can estimate the residual norm as

$$
\left\| K_m\overline{\tilde{f}}^{\delta}_{k,m_o} - \overline{g}^{\delta}_{m_o} \right\|_{\mathbb{R}^{m_o}}
$$

$$
= \left\| \sum_{j=1}^{k} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}} \tilde{w}_{jm_o} - \overline{g}^{\delta}_{m_o} \right\|_{\mathbb{R}^{m_o}} + \sqrt{\sum_{i=1}^{m_o} \left((K_{m_o}\overline{\tilde{f}}^{\delta}_{k,m_o})_i - \sum_{j=1}^{k} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})_{\mathbb{R}^{m_o}} (\tilde{w}_{jm_o})_i \right)^2}
$$

$$
\le \sqrt{\sum_{j=k+1}^{m_o} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})^2_{\mathbb{R}^{m_o}}} + \frac{C_K^2\|\overline{g}^{\delta}_{m_o}\|_{\mathbb{R}^{m_o}}}{3\sigma^2_k m_o^{\frac{3}{2}}}.
$$

Finally, we define the discrepancy principle using

$$
k^{\delta}_{\mathrm{dp}}(m_o) := \max \left\{ k \ge 0 \; : \; \sqrt{\sum_{j=k+1}^{m_o} (\overline{g}^{\delta}_{m_o},\tilde{w}_{jm_o})^2_{\mathbb{R}^{m_o}}} > \tau \sqrt{\mathrm{err}^2_{\mathrm{sys}}(m_o) + \frac{m_o}{o}\delta^2} \right\},
$$

where we estimate the systematic error either as in Section 3.1 for the special kernel (2.3) by

$$\text{err}^2_{\text{sys}}(m_o) := \frac{\|g'\|^2}{m_o},$$

or as in Section 3.2 for the general kernel and the discretization grid $\xi_{jm} := \frac{2j-1}{2m}, j = 1, ..., m$ by

$$\text{err}^2_{\text{sys}}(m_o) := \frac{\|g''\|^2_\infty}{9 \cdot 64 m_o^3},$$

which is usually a much better bound, see (A.15) below. In any case, we emphasize that knowledge of δ and either $\|g'\|$ or $\|g''\|_\infty$ is required here. Usually these two quantities can be estimated from the initial data g_m^δ. Note that the determination of $\|g'\|$ or $\|g''\|_\infty$ is itself an inverse problem, but a rather mild one. Obviously, it would be of interest to derive a method that does not require this additional knowledge. We will address this issue in Section 5. Our adaptive method iteratively applies the discrepancy principle on a scale of different discretization levels $m_{l_0} \le m_{l_1} \le ... \le m$, starting with a very coarse discretization, i.e. some $m_{l_0} = a^{n_0} \le m$. For each discretization level m_o the classical discrepancy principle (3.23) is applied. The method stops when the truncation index determined by the discrepancy principle decreases for the first time. See Algorithm 1 for the numerical implementation. In the following section, we will apply Algorithm 1 to some examples and compare its performance to that of the infeasible optimal choice $k_{\text{opt}}^\delta(m) = \arg\min_{k \le m} \|f_{k,m}^\delta - f\|$.

Algorithm 1 Discrepancy principle + dimension reduction

1: Given noisy point evalutations $g_m^\delta \in \mathbb{R}^m$ and $o_0 \in \mathbb{N}$.;
2: *Initialisation*
3: $i = 0$ and $o_{i+1} = a o_i$
4: Determine $k_{\text{dp}}^\delta(m_{o_i})$ and $k_{\text{dp}}^\delta(m_{o_{i+1}})$
5: *Check stopping criterion*
6: **while** $k_{\text{dp}}^\delta(m_{o_{i+1}}) \ge k_{\text{dp}}^\delta(m_{o_i})$ **do**
7:     $i = i + 1$;
8:     $o_{i+1} = 2 o_i$;
9:     Determine $k_{\text{dp}}^\delta(m_{o_{i+1}})$
10: **end while**
11: Final choice $\overline{f}^\delta_{k_{\text{dp}}^\delta, m_{o_i}}$

---

4. Numerical Experiments

In this section we present the results of our numerical experiments. The initial number of point evaluations is $m = 4^6 = 4096$ and for the reduced data we use $m_o \in \{4^6, ..., 4^2\}$. We perturb the evaluations with i.i.d. Gaussian noise, but note that the results hold for any other centered i.i.d. white noise with finite

variance. We choose the noise level δ implicitly via the signal-to-noise ratio (SNR), which is defined as

$$\text{SNR} := \frac{\|g\|_{L^2}}{\sqrt{\mathbb{E}\left[\|g_m^\delta - g\|^2\right]}} = \frac{\|g\|_{L^2}}{\sqrt{m}\delta}.$$

We will calculate the following quantities:

- The optimal truncation level $k_{\text{opt}} := \arg\min_{k \leq m} \|\overline{f}_{k,m_o}^\delta - f\|$ with the corresponding optimal (relative) error $e_{\text{opt}} := \min_{k \leq m} \|\overline{f}_{k,m_o}^\delta - f\|/\|f\|$.
- The data-driven truncation level k_{dp} determined by Algorithm 1 with the corresponding (relative) error $e_{\text{dp}} := \|\overline{f}_{k_{\text{dp}},m_o}^\delta - f\|/\|f\|$, with parameter $\tau = 1.5$.

Here we take the average of 50 independent runs. We start with the integral equation (2.2) with kernel $\kappa(x,y) = \min(x(1-y), y(1-x))$, for which we know the exact singular value decomposition for both the continuous and the discretized case. In the following we will denote it by deriv2 (because of its relation to the Poisson equation). We define the exact unknown solution over the singular vectors and set $f_i := \sum_{j=1}^{D} \varphi_{s_i}(\sigma_j^2) v_j$ with $D = 5000$ and smoothness parameter $s_i = \frac{1+2i}{8}$, $i = 0, 1, 2$. The noise is varied by the signal-to-noise ratio $\text{SNR} \in \{8^3, 8^2, 8, 1\}$. Note that here all errors are computed exactly, for the formulas we refer to [14]. In view of Theorem 3 it would be interesting to determine $m_{o_{\max}}$, since for this discretization level we expect roughly the same optimal error as for the initial dimension m. However, even with simulated data, $m_{o_{\max}}$ is difficult to obtain because the source element (and hence its norm) for f is not unique. As an alternative, we compute here another reasonable a priori choice for the optimal dimension, which arises naturally in particular when considering the discrepancy principle. We define it implicitly by balancing the dimensions of the two contributions of the data error, i.e., the variance due to random noise on the measurements and the bias due to averaging, in

$$\mathbb{E}\|\overline{g}_{m_o}^\delta - g_{m_o}\|^2 = \frac{\|g'\|^2}{m_o} + \frac{m_o}{o}\delta^2.$$

This way we get

$$o = \frac{m_o^2\delta^2}{\|g'\|^2} = \frac{m^2\delta^2}{l^2\|g'\|^2} \quad \Longrightarrow \quad o = \left(\frac{m^2\delta^2}{\|g'\|^2}\right)^{\frac{1}{3}}.$$

In the following tables, we print the corresponding column number $\left[\frac{1}{3}\log_4\left(\frac{m^2\delta^2}{\|g'\|^2}\right)\right]$ in bold. The results are shown in Tables 1-3 below. Each table consists of four blocks corresponding to the four different SNRs. The five columns are indexed by the decreasing discretization dimension m_o. The four rows in each block show the values of the optimal error, the discretization error, and the respective truncation levels. In addition, the column corresponding to the a priori choice of m_o mentioned above is shown in bold. We observe that for small noise (i.e., large SNR), the optimal error grows as m_o decreases (i.e., as more and more measurements are averaged). However, for larger noise (smaller SNR), the optimal error remains nearly constant for larger m_o. Similarly, for larger noise, the optimal truncation index remains almost constant for larger m_o. This confirms that for larger noise, the reduced data gives the same accuracy. Moreover, we observe that the dimension where the discretization error starts to dominate (i.e., where the optimal error for further reducing m_o grows) is close to the one determined a priori. The error obtained by the discretization principle is larger, in some cases substantially, than the optimal one. This

TABLE 1 `deriv2` *with rough solution* $(s = 1/8)$

m_o		4^6	4^5	4^4	4^3	4^2
SNR $= 8^3$	e_{opt}	**7.0e-1**	7.3e-1	8.0e-1	8.5e-1	8.9e-1
	e_{dp}	**8.0e-1**	8.3e-1	8.7e-1	9.1e-1	9.4e-1
	k_{opt}	**65**	56	23	10	4
	k_{dp}	**15**	9	5	2	1
SNR $= 8^2$	e_{opt}	7.6e-1	**7.7e-1**	7.9e-1	8.5e-1	8.9e-1
	e_{dp}	8.6e-1	**8.5e-1**	8.7e-1	9.1e-1	9.4e-1
	k_{opt}	29	**29**	22	10	4
	k_{dp}	5	**7**	5	2	1
SNR $= 8$	e_{opt}	8.2e-1	8.2e-1	**8.2e-1**	8.5e-1	8.9e-1
	e_{dp}	9.1e-1	9.0e-1	**8.9e-1**	9.1e-1	9.4e-1
	k_{opt}	13	13	**13**	9	4
	k_{dp}	2	3	**3**	2	1
SNR $= 1$	e_{opt}	8.7e-1	8.8e-1	8.7e-1	**8.8e-1**	9.0e-1
	e_{dp}	1.0e0	9.5e-1	9.4e-1	**9.2e-1**	9.4e-1
	k_{opt}	5	5	5	**5**	4
	k_{dp}	0	1	1	**2**	1

TABLE 2 `deriv2` *with medium smooth solution* $(s = 3/8)$

m_o		4^6	4^5	4^4	4^3	4^2
SNR $= 8^3$	e_{opt}	**3.7e-2**	5.6-2	1.0e-1	1.7e-1	2.6e-1
	e_{dp}	**9.9e-2**	1.4e-1	1.8e-1	2.5e-1	4.1e-1
	k_{opt}	**20**	14	8	4	2
	k_{dp}	**6**	4	3	2	1
SNR $= 8^2$	e_{opt}	6.6e-2	**7.0-2**	1.0e-1	1.7e-1	2.6e-1
	e_{dp}	1.8e-1	**1.8e-1**	1.8e-1	2.5e-1	4.1e-1
	k_{opt}	11	**11**	8	4	2
	k_{dp}	3	**3**	3	2	1
SNR $= 8$	e_{opt}	1.2e-1	1.2-1	**1.3e-1**	1.8e-1	2.6e-1
	e_{dp}	4.1e-1	2.5e-1	**2.5e-1**	2.5e-1	4.1e-1
	k_{opt}	6	6	**6**	4	2
	k_{dp}	1	2	**2**	2	1
SNR $= 1$	e_{opt}	2.1e-1	2.1-1	2.1e-1	**2.1e-1**	2.6e-1
	e_{dp}	1e0	4.1e-1	4.1e-1	**4.1e-1**	4.1e-1
	k_{opt}	3	3	3	**3**	2
	k_{dp}	0	1	1	**1**	1

is to be expected due to the rough upper bound on the data noise. Interestingly, we observe that the a priori determined discretization level is similar to the one where the discrepancy principle gives a maximum truncation level, which is mostly the level determined by Algorithm 1 (in fact, they agree whenever the maximum is strict). These results hold regardless of the smoothness of the solution f. For general integral equations we use the open source software package ReguTools from [9]. Here we choose the

TABLE 3   deriv2 *with smooth solution (s = 5/8)*

m_o		4^6	4^5	4^4	4^3	4^2
SNR $= 8^3$	e_{opt}	**6.0e-3**	1.5-2	3.4e-2	7.4e-2	8.1e-2
	e_{dp}	**3.9e-2**	3.9e-2	7.4e-2	1.9e-1	1.9e-1
	k_{opt}	**10**	6	4	2	2
	k_{dp}	**3**	3	2	1	1
SNR $= 8^2$	e_{opt}	1.4e-2	**1.8-2**	3.4e-2	7.4e-2	8.1e-2
	e_{dp}	7.4e-2	**7.4e-2**	7.4e-2	1.9e-1	1.9e-1
	k_{opt}	7	**6**	4	2	2
	k_{dp}	2	**2**	2	1	1
SNR $= 8$	e_{opt}	3.3e-2	3.3-2	**4.1e-2**	7.5e-2	8.1e-2
	e_{dp}	1.9e-1	1.9e-1	**7.7e-2**	1.9e-1	1.9e-1
	k_{opt}	4	4	**4**	2	2
	k_{dp}	1	1	**2**	1	1
SNR $= 1$	e_{opt}	8.4e-2	8.3-2	8.3-2	**8.7e-2**	9.6e-2
	e_{dp}	1e0	1.9e-1	1.9e-1	**1.9e-1**	1.9e-1
	k_{opt}	2	2	2	**2**	2
	k_{dp}	0	1	1	**1**	1

TABLE 4   *gravity*

m_o		4^6	4^5	4^4	4^3	4^2
SNR $= 8^3$	e_{opt}	1.7e-2	**1.7-2**	2.0e-2	5.0e-2	1.9e-1
	e_{dp}	6.1e-2	**6.1e-2**	6.2e-2	1.6e-1	2.8e-1
	k_{opt}	11	**11**	11	11	9
	k_{dp}	5	**5**	5	3	1
SNR $= 8^2$	e_{opt}	3.3e-2	3.3-2	**3.5e-2**	5.7e-2	1.9e-1
	e_{dp}	1.2e-1	8.3e-2	**1.2e-1**	1.6e-1	2.8e-1
	k_{opt}	9	9	**9**	9	8
	k_{dp}	4	5	**4**	3	1
SNR $= 8$	e_{opt}	6.0e-2	6.0e-2	6.2e-2	**7.7e-2**	2.0e-1
	e_{dp}	2.1e-1	2.1e-1	1.5e-1	**1.6e-1**	2.8e-1
	k_{opt}	6	6	6	**6**	6
	k_{dp}	1	1	3	**3**	1
SNR $= 1$	e_{opt}	1.3e-1	1.3-1	1.3e-1	1.3e-1	**2.2e-1**
	e_{dp}	1e0	2.1e-1	2.1e-1	2.2e-1	**2.8e-1**
	k_{opt}	4	4	4	4	**4**
	k_{dp}	0	1	1	1	**1**

two test problems gravity and heat. We use the default parameters and example= 2 in gravity. These are discretizations of a gravity survey problem and the inverse heat equation using midpoint quadrature collocation.   For the dimension m_o we get a system of linear equations $A_{m_o} x_{m_o} = b_{m_o}$ with $(A_{m_o})_{ij} = \frac{1}{m_o} k(\xi_{im_o}, \xi_{j,m_o})$ and $(x_{m_o})_i = f(\xi_{im_o})$ and $b_{m_o} = A_{m_o} x_{m_o}$, with the grid $\xi_{im_o} = \frac{1+2i}{2m_o}$,

TABLE 5 *heat*

m_o	4^6	4^5	4^4	4^3	4^2
SNR $= 8^3$ e_{opt}	**2.2e-2**	2.7-2	8.6e-2	3.4e-1	9.2e-1
e_{dp}	**1.1e-1**	1.8e-1	3.1e-1	6.5e-1	1.0e0
k_{opt}	**32**	32	31	17	3
k_{dp}	**16**	14	9	6	2
SNR $= 8^2$ e_{opt}	5.7e-2	**5.9e-2**	1.0e-1	3.4e-1	9.2e-1
e_{dp}	3.0e-1	**2.6e-1**	3.1e-1	6.5e-1	1.0e0
k_{opt}	24	**24**	24	17	3
k_{dp}	9	**10**	9	6	2
SNR $= 8$ e_{opt}	1.5e-1	1.5-1	**1.6e-1**	3.5e-1	9.2e-1
e_{dp}	6.8e-1	6.0e-1	**4.0e-1**	6.5e-1	1.0e0
k_{opt}	16	16	**17**	16	3
k_{dp}	3	6	**8**	6	2
SNR $= 1$ e_{opt}	3.5e-1	3.5e-1	3.6e-1	**4.3-1**	9.2e0
e_{dp}	1.0e0	8.9e-1	7.5e-1	**6.8e-1**	1.0e0
k_{opt}	10	10	10	**10**	3
k_{dp}	0	1	2	**3**	2

$i, j = 1, ..., m_o$. We choose the same initial numbers of function evaluations and discretization dimensions as for the previous test problem `deriv2`. However, unlike the example above, we do not have explicit formulas for the singular value decompositions. Also, due to the fact that the integral kernels are more complicated, we can no longer evaluate $\|\tilde{f}^{\delta}_{k,m}\|$ exactly. Therefore, instead of constructing $\overline{f}_{k,m_o}$ in L^2 using the formula (3.21), we now avoid the anchored kernel altogether and solve the linear equations directly, i.e, we set $\overline{\tilde{f}}_{k,m_o} := \sum_{j=1}^{k} \frac{(\overline{g}^{\delta}_{m_o}, \tilde{w}_{jm})_{\mathbb{R}^m}}{m \tilde{\sigma}_{jm}} \tilde{z}_{jm} \in \mathbb{R}^{m_o}$ (remember that $(\tilde{\sigma}_{jm_o}, \tilde{z}_{jm_o}, \tilde{w}_{jm_o})$ is the singular value decomposition of A_{m_o}). The total error is then computed as follows: First, we have $e_{k,m_o} := \|\overline{\tilde{f}}^{\delta}_{k,m_o} - x\|^2_{\mathbb{R}^{m_o}}/m_o$, which gives the error (of the coefficients) of the estimator on the projected solution. Then we set $e_{m_o} := \|P_{m_o,D} x_{m_o} - x_D\|^2_D/D$, where $D = 2m = 4096$ and $P_{m_o,D} x_{m_o}$ interpolates and extrapolates $x_{m_o} = \left(f(\xi_{1m_o}) \quad ... \quad f(\xi_{m_o m_o})\right)^T \in \mathbb{R}^{m_o}$ linearly to $\mathbb{R}^D$ (more precisely to a piecewise linear function on the grid $\xi_{1,D}, ... \xi_{D,D}$) and thus gives the error of the projected solution to the exact one. Consequently, the total error $\|\overline{\tilde{f}}^{\delta}_{k,m_o} - f\|_{L^2} = \sqrt{\|\overline{\tilde{f}}^{\delta}_{k,m_o} - f_m\|^2 + \|f_m - f\|^2}$ is approximated by $\sqrt{e_{k,m_o} + e_{m_o}}$. The choice of linear interpolation is reasonable since the integral kernels are smooth. Note that we use $\text{err}^2_{\text{sys}}(m_o) := \frac{\|g'\|^2}{m_o}$ instead of the estimate $\text{err}^2_{\text{sys}}(m_o) := \|g''\|^2_{\infty}/(9 \cdot 64 m_o^3)$. The reason is that in the implementation of the problems from the Toolbox, the right hand side is computed as an application of the quadrature discretization to the coefficient vector of the exact solution. The exact solution is continuous, but only piecewise smooth. Therefore, the latter estimate $\text{err}^2_{\text{sys}}(m_o) := \|g''\|^2_{\infty}/(9 \cdot 64 m_o^3)$ falls below the modeling error and is replaced by the more conservative choice $\text{err}^2_{\text{sys}}(m_o) := \frac{\|g'\|^2}{m_o}$. The numerical results are shown in Tables 4 and 5. They are very similar to the observations we made for `deriv2`.

Overall, our numerical results clearly confirm the potential to reduce the computational cost of ill-posed integral equations, as long as the noise is not too small.

5. Concluding Remarks

In this paper, we introduced and analyzed a novel approach for solving ill-posed integral equations, with a focus on reducing the necessary computational cost. We obtained rigorous error bounds, and designed and implemented an adaptive method that performed promisingly and stably. Note that in further experiments, where we tested the setup with an asymmetric heavy-tailed distribution instead of a Gaussian one, we observed no significant differences and obtained similar results.

We point out three important issues for further research: First, the important task of constructing adaptive data-driven methods has only been touched upon briefly. In particular, it would be advantageous to have a method that does not require knowledge of the noise level δ and the norm of g'. A possible candidate would be the (modified) heuristic discrepancy principle as introduced in [15], for which good results have been demonstrated. We have already tested this method numerically. While for fixed m_o the achieved error was smaller than that of the discrepancy principle, it was not possible to identify the optimal discretization dimension as the maximum. A possible reason could be that the simple heuristic discrepancy principle is not suitable for the systematic error resulting from the averaging of the true data, and further modification might be needed. Another promising approach would be to integrate the averaging into the forward operator itself, as was already done in [12]. This has the advantage that the data error of the averaged data no longer has a systematic error component, since we now have to consider the error $\overline{g}_{m_o}^{\delta} - \overline{g}_{m_o}$. Consequently, knowledge of $\|g'\|$ or $\|g''\|$ is no longer required. However, a rigorous analysis is more complicated, since the effect of the averaging step on the approximation error must be treated carefully, i.e. the effect of the averaging of the operator on the (discretized) singular value decomposition must be well controlled.

In this article, we deliberately focused on the one-dimensional situation only. This allowed us to explain our new averaging method without too much technical detail, to derive precise error estimates, and to demonstrate its superior numerical properties in a straightforward way. Of course, the next step is to extend our method to the more practically relevant higher-dimensional problems that arise, for example, in imaging science. Finally, note that instead of spectral cut-off, other regularization methods could be used, such as standard ones (Tikhonov regularization or Landweber iteration) and in particular more advanced techniques such as conjugate gradient or stochastic gradient descent. This will be future work.

A. Appendix

In this section, we collect the proofs of our theoretical results.

A.1. *Proofs of Section 3.1*

We begin with the proofs for Section 3.1.

Proof of Theorem 3 Note that $\mathcal{N}(K_m)^{\perp}$ is spanned by linear splines on the grid defined by $\xi_{1m},...,\xi_{mm}$. As mentioned above, $s > 3/4$ implies that f is differentiable and we get the bound

$$\Delta_m = \|f_m - f\| \leq \frac{\|f'\|}{\sqrt{2}m}, \tag{A.1}$$

see (3.12) in [14]. From the assumption $o_m \geq 1$ follows that $m\delta^2 \geq \rho^2$. Thus

$$\frac{\|f'\|^2}{m^2} \leq \frac{\rho^2}{m^2} \leq \frac{\delta^2}{m} \leq \frac{k^5\delta^2}{m} \tag{A.2}$$

for all $k = 1, ..., m$ and we see that the discretization error Δ_m is dominated by the other contributions. We now make this precise. We start by showing that

$$c' \left(\frac{k^5}{m} \delta^2 + k^{-8s} \rho^2 \right) \leq \sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E} \| f_{k,m}^\delta - f \|^2 \leq C' \left(\frac{k^5}{m} \delta^2 + k^{-8s} \rho^2 \right) \tag{A.3}$$

with

$$c' := \frac{16}{5} + \frac{1}{3\pi^{4(s+1)}} \qquad C' := \frac{3\pi^4}{5} + \frac{3^{1+s}}{2^{4s-1}} + \frac{1}{2}.$$

For the variance term $\frac{\delta^2}{m+1} \sum_{j=1}^k \frac{1}{\tilde{\sigma}_{jm}^2}$, Lemma 4 and the elementary estimate $2x/\pi \leq \sin(x) \leq x$ for $x \leq \pi/2$ gives

$$\frac{16 k^5}{5m} \leq \sum_{j=1}^k \frac{1}{\tilde{\sigma}_{jm}^2} \leq \frac{3\pi^4 k^5}{5m} \tag{A.4}$$

For the approximation error $\sum_{j=k+1}^m (f, v_{jm})^2$ we have by (3.6) in [14] that

$$\sup_{f \in \mathscr{X}_{s,\rho}} \sum_{j=k+1}^m (f, v_{jm})^2 \leq \frac{3^{s+1}}{2^{4s-1}} \frac{\rho^2}{k^{4s}}, \tag{A.5}$$

which together with (A.1), (A.2) and (A.4) gives the upper bound in (A.3). For the lower bound we use the specific instance $\tilde{f} := \varphi(\sigma_{k+1}^2) \rho v_{k+1}$. Since $f - f_m$ is orthogonal to the range of $v_{1,m}, ..., v_{m,m}$, we have

$$\sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E} \| f_{k,m}^\delta - f \|^2 \geq \delta^2 \sum_{j=1}^k \frac{1}{\tilde{\sigma}_{jm}^2} + \sum_{j=k+1}^m (\tilde{f}, v_{jm})^2. \tag{A.6}$$

All that remains is to bound the second term from below. Due to the special choice of $\tilde{f}$ and position 3.8 from [14] we get

$$\sum_{j=k+1}^m (\tilde{f}, v_{jm})^2 = \varphi_s(\sigma_{k+1}^2)^2 \rho^2 \frac{(m+1)\sigma_{k+1}^2}{\sigma_{k+1,m}^2} \geq \frac{\rho^2}{3\pi^{4(s+1)} k^{4s}}$$

which shows the assertion (A.3).

Next, we trace the effect of modifying $\overline{g}_{m_o}^\delta$ from g_m^δ, taking advantage of the fact that the components of $\overline{g}_{m_o}^\delta$ are approximations of $Kf = g$ on the coarser grid with meshwidth $1/(m_o + 1)$ instead of $1/(m + 1)$. First, it holds that

$$\left(\overline{g}_{m_o}^\delta \right)_i = \frac{1}{o} \sum_{j=1}^o \left(g \left(\xi_{(i-1)l+j,m} \right) + \delta Z_{(i-1)o+j} \right) = \frac{1}{o} \sum_{j=1}^o g \left(\frac{(i-1)o+j}{m+1} \pi \right) + \frac{\delta}{\sqrt{o}} \overline{Z}_i,$$

where we defined $\overline{Z}_i := \frac{\sum_{j=1}^o Z_{(i-1)o+j}}{\sqrt{o}}$. Note that $\overline{Z}_i \sqrt{o}$ and Z_1 have the same distribution. Furthermore, the $\overline{Z}_i$ are independent and identically distributed. Consequently, the variance of the measurement is

reduced by a factor of $1/o$. However, the bias has changed. Now, using the Cauchy-Schwartz inequality, we obtain the upper bound

$$\left| \frac{\sum_{j=1}^{o} g(\xi_{(i-1)l+j,m})}{o} - g(\xi_{im_o}) \right| \le \frac{1}{o} \left| \sum_{j=1}^{o} g(\xi_{(i-1)o+j,m}) - g(\xi_{im_o}) \right| \le \sup_{t \in (\xi_{(i-1)o,m}, \xi_{i,m})} |g(t) - g(\xi_{im_o})|$$

$$\le \left| \int_{\xi_{(i-1)o,m}}^{\xi_{io,m}} g'(t) \mathrm{d}t \right| \le \sqrt{\frac{l}{m}} \sqrt{\int_{\xi_{(i-1)o,m}}^{\xi_{io,m}} (g')^2(t) \mathrm{d}t}. \tag{A.7}$$

Consequently,

$$\sum_{j=1}^{k} \mathbb{E} \left(\bar{z}_{m_o}^{\delta} - K_{m_o} f, u_{jm_o} \right)^2$$

$$= \sum_{j=1}^{k} \left(\left(\begin{pmatrix} \frac{\sum_{s=1}^{o} g(\xi_{s,m})}{o} - g(\xi_{1m_o}) \\ \vdots \\ \frac{\sum_{s=1}^{o} g(\xi_{(m-1)o+s,m})}{o} - g(\xi_{m_o m_o}) \end{pmatrix}, u_{jm_o} \right) \right)^2 + \sum_{j=1}^{k} \delta^2 \mathbb{E} \left[\left(\begin{pmatrix} \bar{Z}_1 \\ \vdots \\ \bar{Z}_{m_o} \end{pmatrix}, u_{jm_o} \right)^2 \right]$$

$$\le \left\| \begin{pmatrix} \frac{\sum_{s=1}^{o} g(\xi_{s,m})}{o} - g(\xi_{1m_o}) \\ \vdots \\ \frac{\sum_{s=1}^{o} g(\xi_{(m-1)o+s,m})}{o} - g(\xi_{m_o m_o}) \end{pmatrix} \right\|^2 + k \frac{\delta^2}{o} \le \frac{l}{m} \sum_{i=1}^{m_o} \int_{\xi_{(i-1)o+1,m}}^{\xi_{io,m}} |g'(x)|^2 \mathrm{d}x + k \frac{\delta^2}{o}$$

$$\le \frac{l}{m} \|g'\|^2 + k \frac{\delta^2}{o}.$$

From [14] we know that

$$\sup_{f \in \mathcal{X}_{s,\rho}} \|f_m - f\|^2 \le \frac{\|f'\|^2}{2m^2} \le \frac{\rho^2}{2m^2}$$

and

$$\sup_{f \in \mathcal{X}_{s,\rho}} \sum_{j=k+1}^{m} (f, v_{jm})^2 \le C_s k^{-4s} \rho^2$$

with $C_s = \frac{3^{s+1}}{2^{4s-1}}$. Moreover, we have $\sup_{f \in \mathscr{X}_{s,\rho}} \|g'\| \leq \rho$. Thus, the error gets

$$\sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E}\|\overline{f}^{\delta}_{k,m_o} - f\|^2$$

$$\leq \sup_{f \in \mathscr{X}_{s,\rho}} \sum_{j=1}^{k} \frac{\mathbb{E}\left(\overline{z}^{\delta}_{m_o} - K_{m_o}f, u_{jm_o}\right)^2}{\tilde{\sigma}^2_{jm_o}} + \sup_{f \in \mathscr{X}_{s,\rho}} \sum_{j=k+1}^{m} (f, v_{jm_o})^2 + \sup_{f \in \mathscr{X}_{s,\rho}} \|f_{m_o} - f\|^2$$

$$\leq \sigma^{-2}_{k,m} \sup_{f \in \mathscr{X}_{s,\rho}} \sum_{j=1}^{k} \mathbb{E}(\overline{z}^{\delta}_{m_o} - K_{m_o}f, u_{jm_o})^2 + C_s \frac{\rho^2}{k^{4s}} + \frac{\rho^2}{2m_o^2}$$

$$\leq \frac{3\pi^4 k^4}{m_o}\left(\sup_{f \in \mathscr{X}_{s,\rho}} \frac{o}{m}\|g'\|^2 + k\frac{\delta^2}{o}\right) + C_2\frac{\rho^2}{k^{4s}} + \frac{\rho^2}{2m_o^2}$$

$$\leq 3\pi^4 k^4\left(\frac{o^2}{m^2}\rho^2 + k\frac{\delta^2}{m}\right) + C_2\frac{\rho^2}{k^{4s}} + \frac{\rho^2}{2m_o^2} \leq 3\pi^4 k^4\left(\frac{3o^2}{2m^2}\rho^2 + k\frac{\delta^2}{m}\right) + C_s\frac{\rho^2}{k^{4s}}.$$

For $o \leq \sqrt{\frac{m\delta^2}{\rho^2}} =: o_m$ holds that $\frac{o^2}{m^2}\rho^2 \leq \frac{\delta^2}{m}$. So for $o \leq o_m$,

$$\sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E}\|\overline{f}^{\delta}_{k,m_o} - f\|^2 \leq 6\pi^4 k^5 \frac{\delta^2}{m} + C_s\frac{\rho^2}{k^{4s}} \leq C \sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E}\|f^{\delta}_{k,m} - f\|^2$$

with $C = \frac{6\pi^4 + C_s}{c'}$, where we used (A.3). This gives us the upper bound in (3.4). For the lower bound we have

$$\sup_{f \in \mathscr{X}_{s,\rho}} \mathbb{E}\|\overline{f}^{\delta}_{k,m_o} - f\|^2$$

$$= \frac{\delta^2}{o}\sum_{j=1}^{k}\frac{1}{\tilde{\sigma}^2_{jm_o}} + \sup_{f \in \mathscr{X}_{s,\rho}}\left(\sum_{j=1}^{k}\frac{(\overline{g}_{m_o} - g_{m_o}, u_{jm_o})^2}{\tilde{\sigma}^2_{jm_o}} + \sum_{j=k+1}^{m_o}(f, v_{jm_o})^2 + \|f - f_{m_o}\|^2\right)$$

$$\geq \frac{\delta^2}{o}\sum_{j=1}^{k}\frac{1}{\tilde{\sigma}^2_{jm_o}} + \sup_{f \in \mathscr{X}_{s,\rho}}\sum_{j=k+1}^{m_o}(f, v_{jm_o})^2 \geq \frac{16}{5}\frac{k^5}{om_o} + \rho^2\frac{k^{-4s}}{3\pi^{4(s+1)}}$$

$$\geq c'\left(k^5\frac{\delta^2}{m} + \rho^2 k^{-4s}\right) \geq \frac{c'}{C'}\sup_{f \in \mathscr{X}_{s,\rho}}\mathbb{E}\|f^{\delta}_{k,m} - f\|^2 = c\sup_{f \in \mathscr{X}_{s,\rho}}\mathbb{E}\|f^{\delta}_{k,m} - f\|^2.$$

After taking the minimum over $k = 1, ..., m$ on both sides, the lower bound in (3.4) is obtained and Theorem 3 is finally proved. □

A.2. *Proofs of Section 3.2*

To prove the main theorems, we need some auxiliary results. The corresponding lemmata were given above, and we now give the proof here.

Proof of the Lemma 4 The proof is based on the Courant-Fischer principle, which states that for self-adjoint operators S, T in Hilbert spaces there exists $|\lambda_i(S) - \lambda_i(T)| \leq \|S - T\|$, where $\lambda_i(\cdot)$ is the i-th

largest eigenvalue. We first apply this bound to K^*K and $K_m^*K_m$, where K_m is from (3.7). For the singular vector v_j we have

$$K^*Kv_j - K_m^*K_mv_j = \sigma_j \left(K^*u_j - \frac{1}{\sqrt{m}}K_m^* \left(u_j(\xi_{1m}) \quad \dots \quad u_j(\xi_{mm}) \right)^T \right)$$

$$= \sigma_j \left(\int \kappa(y,\cdot)u_j(y)\mathrm{d}y - \frac{1}{m}\sum_{l=1}^m \kappa(\xi_{lm},\cdot)u_j(\xi_{lm}) \right).$$

Now note that the singular functions v_j and u_j of K are twice differentiable, since

$$u_j = \frac{Kv_j}{\sigma_j} = \frac{\int \kappa(x,\cdot)v_j(x)\mathrm{d}x}{\sigma_j},$$

and thus, for $\alpha \leq 2$, and after changing the order of differentiation and integration, we have

$$u_j^{(\alpha)} = \frac{\int \partial_y^\alpha \kappa(x,\cdot)v_j(x)\mathrm{d}x}{\sigma_j}.$$

This is justified by the fact that κ and also the function $y \mapsto v_j(y) = \frac{1}{\sigma_j}\int \kappa(x,y)u_j(x)\mathrm{d}x$ are continuous (and thus bounded), since

$$\left| \int \kappa(x,y)u_j(x)\mathrm{d}x - \int k(x,y')u_j(x)\mathrm{d}x \right| \leq \sqrt{\int (\kappa(x,y) - \kappa(x,y'))^2\,\mathrm{d}x \int u_j^2(x)\mathrm{d}x}$$

$$\leq \sup_{u,w}|\partial_w\kappa(u,w)||y-y'|.$$

Consequently,

$$\|u_j^{(\alpha)}\|_\infty \leq \frac{1}{\sigma_j}\sup_y \left| \int \partial_y^\alpha \kappa(x,y)v_j(x)\mathrm{d}x \right| \leq \frac{1}{\sigma_j}\sup_y \sqrt{\int (\partial_y^\alpha\kappa(x,y))^2\,\mathrm{d}x \int v_j^2(x)\mathrm{d}x} \leq \frac{C_K}{\sigma_j}$$

and furthermore, for any y,

$$\left\|\partial_x^2\left(\kappa(x,y)u_j(x)\right)\right\|_\infty \leq \sup_{x,y}\sum_{\alpha=0}^2 \binom{2}{\alpha}|\partial_x^\alpha\kappa(x,y)\partial_x^{2-\alpha}u_j(x)| \leq 4\frac{C_K^2}{\sigma_j}.$$

Thus

$$\left| \int \kappa(x,y)u_j(x)\mathrm{d}x - \frac{1}{m}\sum_{l=1}^m \kappa(\xi_{lm},y)u_j(\xi_{lm}) \right| \leq \frac{\left\|\partial_x^2\left(\kappa(x,y)u_j(x)\right)\right\|_\infty}{24m^2} \leq \frac{C_K^2}{6\sigma_jm^2}$$

and since the v_j's form a basis of $\mathcal{N}(K)^\perp$ and $K_m^*w = 0 \in \mathbb{R}^m$, for any $w \in \mathcal{N}(K)$ we get

$$\|K^*K - K_m^*K_m\| = \sup_j\|K^*Kv_j - K_m^*K_mv_j\| \leq \frac{C_K^2}{6m^2}. \tag{A.8}$$

Now recall the design matrix

$$(T_m)_{ij\leq m} := \frac{1}{m}\left(\int \kappa(x,\xi_{im})\kappa(x,\xi_{jm})\mathrm{d}x\right)_{ij\leq m} \in \mathbb{R}^m,$$

from (2.9), but now with the kernel κ satisfying the assumption 1. We have

$$\left(T_m\tilde{z}_{km} - A_m^T A_m\tilde{z}_{km}\right)_i$$

$$= \frac{1}{m}\sum_{j=1}^m \int \kappa(x,\xi_{im})\kappa(x,\xi_{jm})(\tilde{z}_{km})_j - \frac{1}{m^2}\sum_{j=1}^m\sum_{l=1}^m \kappa(\xi_{lm},\xi_{im})\kappa(\xi_{lm},\xi_{jm})(\tilde{z}_{km})_j$$

$$= \frac{1}{m}\sum_{j=1}^m (\tilde{z}_{km})_j\left(\int \kappa(x,\xi_{im})\kappa(x,\xi_{jm})\mathrm{d}x - \frac{1}{m}\sum_{l=1}^m \kappa(\xi_{lm},\xi_{im})\kappa(\xi_{lm},\xi_{jm})\right)$$

and thus

$$\|T_m - A_m^T A_m\| = \sup_{k\leq m}\|T_m\tilde{z}_{km} - A_m^T A_m\tilde{z}_{km}\| \tag{A.9}$$

$$= \sup_{k\leq m}\sqrt{\sum_{i=1}^m\left(\frac{1}{m}\sum_{j=1}^m(\tilde{z}_{km})_j\left(\int \kappa(x,\xi_{im})\kappa(x,\xi_{jm})\mathrm{d}x - \frac{1}{m}\sum_{l=1}^m \kappa(\xi_{lm},\xi_{im})\kappa(\xi_{lm},\xi_{jm})\right)\right)^2}$$

$$\leq \sup_{k\leq m}\sqrt{\sum_{i=1}^m\left(\frac{1}{m}\sum_{j=1}^m|(\tilde{z}_{km})_j|\frac{\|\partial_x^2\left(\kappa(x,\xi_{im})\kappa(x,\xi_{jm})\right)\|_\infty}{24m^2}\right)^2} \leq \sup_{k\leq m}\frac{C_K^2}{6m^2}\frac{1}{\sqrt{m}}\sum_{j=1}^m|(\tilde{z}_{km})_j| \leq \frac{C_K^2}{6m^2}.$$

Now recall Proposition 1. Thus, the singular values of the semi-discrete operator K_m are the square root of the eigenvalues of the matrix T_m, which we denote by η_{jm}. Therefore, using the Courant-Fischer principle and (A.9), we deduce that

$$\left|\eta_{jm} - \tilde{\sigma}_{jm}^2\right| \leq \|T_m - A_m^T A_m\| \leq \frac{C_K^2}{6m^2} \quad \text{for all } j = 1,...,m.$$

Furthermore, by (A.8) we also infer

$$\left|\sigma_j^2 - \eta_{jm}\right| \leq \|K^*K - K_m^*K_m\| \leq \frac{C_K^2}{6m^2} \quad \text{for all } j = 1,...,m.$$

With the triangle inequality we get

$$\left|\sigma_j^2 - \tilde{\sigma}_{jm}^2\right| \leq \frac{C_K^2}{3m^2},$$

and the proof of Lemma 4 is finished. $\square$

Proof of Lemma 5 The proof of (3.14) is straightforward: We have

$$\varepsilon^2 \geq \|Kv - \lambda v\|^2 = \|\sum_{i \in \mathbb{N}} \lambda_i(v, v_i)v_i - \lambda \sum_{i \in \mathbb{N}} (v, v_i)v_i\|^2 = \sum_{i \in N} (\lambda_i - \lambda)^2(v, v_i)^2$$

$$\geq \min_{i \in \mathbb{N}} |\lambda_i - \lambda|^2 \sum_{i \in \mathbb{N}} (v, v_i)^2 = \|v\|^2 \min_{i \in \mathbb{N}} |\lambda_i - \lambda|^2.$$

For (3.15), it holds that

$$(v, v_j)^2 = \|v\|^2 - \sum_{i \neq j} (v, v_i)^2 = \|v\|^2 - \sum_{i \neq j} \frac{(\lambda_i - \lambda)^2}{(\lambda_i - \lambda)^2}(v, v_i)^2 \geq \|v\|^2 - \frac{1}{c^2} \sum_{i \neq j} (\lambda_i - \lambda)^2(v, v_i)^2$$

$$\geq \|v\|^2 - \frac{1}{c^2} \sum_{i \in \mathbb{N}} (\lambda_i - \lambda)^2(v, v_i)^2 = \|v\|^2 - \frac{1}{c^2}\|Kv - \lambda v\|^2 \geq \|v\|^2 - \frac{\varepsilon^2}{c^2}.$$

□

Proof of Lemma 6 We start with the proof of (3.17). First,

$$(\tilde{v}_{jm}, \tilde{v}_{im}) = \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}m} \sum_{l,l'=1}^{m} (\tilde{w}_{jm})_l(\tilde{w}_{im})_{l'} \int \kappa(\xi_{lm}, y)\kappa(\xi_{l'm}, y)\mathrm{d}y$$

$$= \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}m} \sum_{l,l'=1}^{m} (\tilde{w}_{jm})_l(\tilde{w}_{im})_{l'} \left(\frac{1}{m} \sum_{t=1}^{m} \kappa(\xi_{lm}, \xi_{tm})\kappa(\xi_{l'm}, \xi_{tm}) + \mathcal{O}\left(\frac{1}{m^2}\right) \right)$$

$$= \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \sum_{t=1}^{m} \left(\frac{1}{m} \sum_{l=1}^{m} \kappa(\xi_{lm}, \xi_{tm})(\tilde{w}_{jm})_l \right) \left(\frac{1}{m} \sum_{l'=1}^{m} \kappa(\xi_{l'm}, \xi_{tm})(\tilde{w}_{im})_{l'} \right)$$

$$+ \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \sum_{l=1}^{m} \frac{|(\tilde{w}_{jm})_l|}{\sqrt{m}} \sum_{l'=1}^{m} \frac{|(\tilde{w}_{im})_{l'}|}{\sqrt{m}} \mathcal{O}\left(\frac{1}{m^2}\right)$$

$$= \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \sum_{t=1}^{m} \tilde{\sigma}_{jm}(\tilde{z}_{jm})_t \tilde{\sigma}_{im}(\tilde{z}_{im})_t + \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \mathcal{O}\left(\frac{1}{m^2}\right) = \delta_{ij} + \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \mathcal{O}\left(\frac{1}{m^2}\right).$$

Since $\|\partial_y^2(\kappa(\xi_{lm}, y)\kappa(\xi_{l'm}, y))\|_\infty \leq 4C_K^2$ and $\sum_{l=1}^{m} \frac{|(\tilde{w}_{jm})_l|}{\sqrt{m}} \leq 1$, we can write the constant exactly in $\mathcal{O}$ notation as $\mathcal{O}(m^{-2}) = 4C_K^2 m^{-2}$. Together with the fact that $\tilde{\sigma}_{jm}^2 \geq \sigma_j^2/2$ for $j \leq J_m$, we get

$$\left| (\tilde{v}_{jm}, \tilde{v}_{im}) - \delta_{ij} \right| \leq \frac{1}{\tilde{\sigma}_{jm}\tilde{\sigma}_{im}} \frac{4C_K^2}{24m^2} \leq \frac{2C_K^2}{3\sigma_j\sigma_im^2}.$$

This brings us to the proof of (3.18). For the upper bound, due to (3.17), we have

$$(\tilde{v}_{jm}, v_j)^2 \leq \|\tilde{v}_{jm}\|^2 \leq \left(1 + \frac{C_K^2}{3\sigma_j^2m^2} \right)^2 \leq 1 + \frac{5}{2}\frac{C_K^2}{3\sigma_j^2m^2} \leq 1 + \frac{C_K^2}{\sigma_j^2m^2},$$

since $(1+x)^2 \leq 1 + \frac{5}{2}x$ for all $0 \leq x \leq \frac{1}{2}$. For the lower bound we use Lemma 5. As a first step, we show that

$$\left\| K^*K\tilde{v}_{jm} - \tilde{\sigma}_{jm}\tilde{v}_{jm} \right\| \leq \frac{C_K^3}{2\sigma_j m^2}. \tag{A.10}$$

Indeed,

$$K^*K\tilde{v}_{jm} = \int k(y,\cdot) \int \kappa(y,z)\tilde{v}_{jm}(z)\mathrm{d}z\mathrm{d}y$$

$$= \int \kappa(y,\cdot) \left[\int \kappa(y,z)\tilde{v}_{jm}(z)\mathrm{d}z - \frac{1}{m}\sum_{l=1}^{m}\kappa(y,\xi_{lm})\tilde{v}_{jm}(\xi_{lm}) \right] \mathrm{d}y$$

$$+ \int \kappa(y,\cdot)\frac{1}{m}\sum_{l=1}^{m}\kappa(y,\xi_{lm})\tilde{v}_{jm}(\xi_{lm})\mathrm{d}y$$

$$= \int \kappa(y,\cdot) \left[\int \kappa(y,z)\tilde{v}_{jm}(z)\mathrm{d}z - \frac{1}{m}\sum_{l=1}^{m}\kappa(y,\xi_{lm})\tilde{v}_{jm}(\xi_{lm}) \right] \mathrm{d}y \tag{A.11}$$

$$+ \frac{1}{m}\sum_{l=1}^{m}\tilde{v}_{jm}(\xi_{lm}) \left[\int \kappa(y,\cdot)\kappa(y,\xi_{lm})\mathrm{d}y - \frac{1}{m}\sum_{l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_l) \right] \tag{A.12}$$

$$+ \frac{1}{m^2}\sum_{l,l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_l)\tilde{v}_{jm}(\xi_{lm}). \tag{A.13}$$

Now, for (A.13) we have

$$\frac{1}{m^2}\sum_{l,l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_l)\tilde{v}_{jm}(\xi_{lm}) = \frac{1}{m^2}\sum_{l,l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_l)\frac{1}{\tilde{\sigma}_{jm}\sqrt{m}}\sum_{i=1}^{m}(\tilde{w}_{jm})_i\kappa(\xi_{im},\xi_{lm})$$

$$= \frac{1}{m^{\frac{3}{2}}}\sum_{l,l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_{lm})(\tilde{z}_{jm})_l = \frac{\tilde{\sigma}_{jm}}{\sqrt{m}}\sum_{l'=1}^{m}(\tilde{w}_{jm})_{l'}\kappa(\xi_{l'm},\cdot) = \tilde{\sigma}_{jm}^2\tilde{v}_{jm}(\cdot).$$

Next we bound (A.12) from above and obtain

$$\left\| \frac{1}{m}\sum_{l=1}^{m}\tilde{v}_{jm}(\xi_{lm}) \left[\int \kappa(y,\cdot)\kappa(y,\xi_{lm})\mathrm{d}y - \frac{1}{m}\sum_{l'=1}^{m}\kappa(\xi_{l'm},\cdot)\kappa(\xi_{l'm},\xi_l) \right] \right\|$$

$$\leq \frac{C_K^2}{6m^2}\frac{1}{m}\sum_{l=1}^{m}|\tilde{v}_{jm}(\xi_{lm})| = \frac{C_K^2}{6m^3}\sum_{l=1}^{m}\left| \frac{1}{\tilde{\sigma}_{jm}\sqrt{m}}\sum_{i=1}^{m}(\tilde{w}_{jm})_i\kappa(\xi_{im},\xi_{lm}) \right|$$

$$= \frac{C_K^2}{6m^{\frac{5}{2}}}\sum_{l=1}^{m}\left| (\tilde{z}_{jm})_i \right| \leq \frac{C_K^2}{6m^2}.$$

To bound (A.11) we use for $\alpha \leq 2$ that

$$\|\partial^\alpha\tilde{v}_{jm}\|_\infty = \left\| \frac{1}{\tilde{\sigma}_{jm}\sqrt{m}}\sum_{l=1}^{m}\partial_y^\alpha\kappa(\xi_{lm},\cdot)(\tilde{w}_{jm})_l \right\|_\infty \leq \frac{1}{\tilde{\sigma}_{jm}}\frac{C_K}{\sqrt{m}}\sum_{l=1}^{m}|(\tilde{w}_{jm})_l| \leq \frac{C_K}{\tilde{\sigma}_{jm}}$$

and thus

$$\left\| \int \kappa(y,\cdot) \left[\int \kappa(y,z) \tilde{v}_{jm}(z) \mathrm{d}z - \frac{1}{m} \sum_{l=1}^{m} \kappa(y,\xi_{lm}) \tilde{v}_{jm}(\xi_{lm}) \right] \mathrm{d}y \right\|$$

$$\leq C_K \frac{\sup_y \|\partial_y^2 (\kappa(y,\cdot) \tilde{v}_{jm}(\cdot))\|_\infty}{24m^2} \leq \frac{C_K^3}{6\tilde{\sigma}_{jm} m^2} \leq \frac{\sqrt{2} C_K^3}{6\sigma_j m^2}.$$

Since $\sigma_j \leq C_K$ we get

$$\left\| K^* K \tilde{v}_{jm} - \tilde{\sigma}_{jm} \tilde{v}_{jm} \right\| \leq \frac{C_K^2}{6m^2} + \frac{\sqrt{2} C_K^3}{6\sigma_j m^2} \leq \frac{(1+\sqrt{2}) C_K^3}{6\sigma_j m^2} \leq \frac{C_K^3}{2\sigma_j m^2}$$

which shows (A.10). Now we use Lemma 5. By Lemma 4 we have that $|\sigma_j^2 - \tilde{\sigma}_{jm}^2| \leq \frac{C_K^2}{3m^2} \leq \frac{c_j}{2}$, so $\arg\min_{i \in \mathbb{N}} |\sigma_i^2 - \tilde{\sigma}_{jm}^2| = j$. Then by (A.12) of Lemma 5 with

$$\min_{i \neq j} |\sigma_i^2 - \tilde{\sigma}_{jm}^2| \geq \min_{i \neq j} |\sigma_i^2 - \sigma_j^2| - |\sigma_j^2 - \tilde{\sigma}_{jm}^2| \geq c_j - \frac{C_K^2}{3m^2} \geq \frac{c_j}{2} =: c$$

and

$$\| K^* K \tilde{v}_{jm} - \tilde{\sigma}_{jm}^2 \tilde{v}_{jm} \| \leq \frac{C_K^3}{2\sigma_j m^2} =: \varepsilon,$$

we deduce that

$$(v_j, \tilde{v}_{jm})^2 \geq \|\tilde{v}_{jm}\|^2 - \frac{\varepsilon^2}{c^2} \geq \left(1 - \frac{C_K^2}{3\sigma_j^2 m^2} \right)^2 - \frac{\frac{C_K^4}{4\sigma_j^2 m^4}}{\frac{c_j^2}{4}} \geq 1 - \frac{2C_K^2}{3\sigma_j^2 m^2} - \frac{C_K^4}{c_j^2 \sigma_j^2 m^4}.$$

□

Proof of Theorem 7 First, note that

$$\frac{(g_m, \tilde{w}_{jm})_{\mathbb{R}^m}}{\sqrt{m}\tilde{\sigma}_{jm}} = \frac{(K_m f, \tilde{w}_{jm})_{\mathbb{R}^m}}{\tilde{\sigma}_{jm}} = \frac{(f, K_m^* \tilde{w}_{jm})}{\tilde{\sigma}_{jm}} = \frac{(f, \sum_{l=1}^{m} (\tilde{w}_{jm})_l \kappa(\xi_{lm}, \cdot))}{\sqrt{m}\tilde{\sigma}_{jm}} = (f, \tilde{v}_{jm}).$$

We begin to decompose the error as

$$\tilde{f}_{k,m}^\delta - f := \sum_{j=1}^{k} \frac{(g_m^\delta, \tilde{w}_{jm})_{\mathbb{R}^m}}{\sqrt{m}\tilde{\sigma}_{jm}} \tilde{v}_{jm} - \sum_{j=1}^{\infty} (f, v_j) v_j = \sum_{j=1}^{k} \frac{(g_m^\delta - g_m, \tilde{w}_{jm})_{\mathbb{R}^m}}{\sqrt{m}\tilde{\sigma}_{jm}} \tilde{v}_{jm}$$

$$+ \sum_{j=1}^{k} (f, \tilde{v}_{jm} - v_j) \tilde{v}_{jm} + \sum_{j=1}^{k} (f, v_j)(\tilde{v}_{jm} - v_j) - \sum_{j=k+1}^{\infty} (f, v_j) v_j. \qquad (A.14)$$

First we analyze the first term, the random contribution. Due to the unbiasedness of the noise, we obtain

$$\mathbb{E}\left[\left\|\sum_{j=1}^{k}\frac{(g_m^\delta-g_m,\tilde{w}_{jm})_{\mathbb{R}^m}}{\sqrt{m}\tilde{\sigma}_{jm}}\tilde{v}_{jm}\right\|^2\right]=\sum_{j,j'=1}^{k}\frac{\mathbb{E}\left[(g_m^\delta-g_m,\tilde{w}_{jm})_{\mathbb{R}^m}(g_m^\delta-g_m,\tilde{w}_{j'm})_{\mathbb{R}^m}\right]}{m\tilde{\sigma}_{jm}\tilde{\sigma}_{j'm}}(\tilde{v}_{jm},\tilde{v}_{j'm})$$

$$=\frac{\delta^2}{m}\sum_{j=1}^{k}\frac{1}{\tilde{\sigma}_{jm}^2}\|\tilde{v}_{jm}\|^2\le\frac{4\delta^2}{m}\sum_{j=1}^{k}\frac{1}{\sigma_j^2}.$$

To treat the next terms, note that without loss of generality we can assume that $(\tilde{v}_{jm},v_j)\ge 0$ (otherwise we consider $-v_j$ instead of v_j). Then

$$\|\tilde{v}_{jm}-v_j\|^2=\left\|\sum_{l=1}^{\infty}(\tilde{v}_{jm},v_l)v_l-v_j\right\|^2=((\tilde{v}_{jm},v_j)-1)^2+\sum_{l\ne j}(v_l,\tilde{v}_{jm})^2$$

$$=((\tilde{v}_{jm},v_j)-1)^2+\|\tilde{v}_{jm}\|^2-(v_j,\tilde{v}_{jm})^2=(1-\|\tilde{v}_{jm}\|)^2+2(\|\tilde{v}_{jm}\|-(\tilde{v}_{jm},v_j))$$

$$=\left(\frac{1-\|\tilde{v}_{jm}\|^2}{1+\|\tilde{v}_{jm}\|}\right)^2+2\frac{\|\tilde{v}_{jm}\|^2-(\tilde{v}_{jm},v_j)^2}{\|\tilde{v}_{jm}\|+(\tilde{v}_{jm},v_j)}.$$

With Lemma 6 and $(\tilde{v}_{jm},v_j)\ge 0$ we get

$$\|\tilde{v}_{jm}-v_j\|^2\le\left(\frac{C_K^2}{3\sigma_j^2m^2}\right)^2+2\frac{C_K^4}{c_j^2\sigma_j^2m^4}\le\frac{3C_K^4}{c_j^2\sigma_j^2m^4},$$

since $c_j\le\sigma_j$. So

$$\left\|\sum_{j=1}^{k}(f,\tilde{v}_{jm}-v_j)\tilde{v}_{jm}\right\|^2=\sum_{j=1}^{k}(f,\tilde{v}_{jm}-v_j)^2\|\tilde{v}_{jm}\|^2+\sum_{\substack{j,j'=1\\j\ne j'}}^{k}(f,\tilde{v}_{jm}-v_j)(f,\tilde{v}_{j'm}-v_{j'})(\tilde{v}_{jm},\tilde{v}_{j'm})$$

$$\le\|f\|\sum_{j=1}^{k}\|\tilde{v}_{jm}-v_j\|^2\|\tilde{v}_{jm}\|^2+\sum_{\substack{j,j'=1\\j\ne j'}}^{k}|(f,\tilde{v}_{jm}-v_j)(f,\tilde{v}_{j'm}-v_{j'})|\frac{2C_K^2}{3\sigma_j\sigma_{j'}m^2}$$

$$\le 2\|f\|\sum_{j=1}^{k}\|\tilde{v}_{jm}-v_j\|^2+\frac{2C_K^2}{3m^2}\left(\sum_{j=1}^{k}\frac{|(f,\tilde{v}_{jm}-v_j)|}{\sigma_j}\right)^2$$

$$\le 2\|f\|\sum_{j=1}^{k}\|\tilde{v}_{jm}-v_j\|^2+\frac{2C_K^2\|f\|^2}{3m^2}\left(\sum_{j=1}^{k}\frac{\|\tilde{v}_{jm}-v_j\|}{\sigma_j}\right)^2$$

$$\le\frac{6C_K^4\|f\|^2}{m^4}\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}+\frac{6C_K^6\|f\|^2}{3m^6}\left(\sum_{j=1}^{k}\frac{1}{c_j\sigma_j^2}\right)^2.$$

Finally, using Cauchy's inequality, we get

$$\left\|\sum_{j=1}^{k}(f,v_j)(\tilde{v}_{jm}-v_j)\right\|^2 \le \left(\sum_{j=1}^{k}|(f,v_j)|\|\tilde{v}_{jm}-v_j\|\right)^2 \le \sum_{j=1}^{k}(f,v_j)^2 \sum_{j=1}^{k}\|\tilde{v}_{jm}-v_j\|^2$$

$$\le \frac{3C_K^4\|f\|^2}{m^4}\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}.$$

Combining all the previous estimates, we end up with

$$\sqrt{\mathbb{E}\|\tilde{f}_{k,m}^{\delta}-f\|^2}$$

$$\le \frac{2\delta}{\sqrt{m}}\sqrt{\sum_{j=1}^{k}\frac{1}{\sigma_j^2}} + \frac{\sqrt{6}C_f^2}{m^2}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}} + \frac{\sqrt{2}C_K^3\|f\|}{m^3}\sum_{j=1}^{k}\frac{1}{c_j\sigma_j^2} + \frac{\sqrt{3}C_k^2\|f\|}{m^2}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}}$$

$$+ \sqrt{\sum_{j=k+1}^{\infty}(f,v_j)^2} \le \frac{2\delta}{\sqrt{m}}\sqrt{\sum_{j=1}^{k}\frac{1}{\sigma_j^2}} + \frac{C_1}{m^2}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty}(f,v_j)^2},$$

with C_1 from (3.20). This results because with $k \le J_m$ we have

$$\frac{1}{m}\sum_{j=1}^{k}\frac{1}{\sigma_j^2} \le \frac{c_1^2}{m}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j\sigma_j^2}}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}} \le \frac{c_1^2}{m}\sqrt{\frac{k}{c_k\sigma_k^2}}\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}} \le c_1^2\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}}$$

and similar

$$\frac{1}{m}\sum_{j=1}^{k}\frac{1}{c_j\sigma_j^2} \le c_1\sqrt{\sum_{j=1}^{k}\frac{1}{c_j^2\sigma_j^2}}.$$

This completes the proof.   □

Proof of Theorem 8 As mentioned above, compared to the setting in Theorem 7, we need to carefully analyze the contribution of the systematic component in the data propagation error. First, we show that

$$\left\|\bar{g}_{m_o}-g_{m_o}\right\|^2 \le \frac{\|g''\|_\infty^2}{9\cdot 64m_o^3}. \tag{A.15}$$

To prove (A.15), we check that the right-hand side g of (2.1) is twice differentiable. In fact, we have

$$\frac{d^2}{dx^2}g(x) = \frac{d^2}{dx^2}(Kf)(x) = \frac{d^2}{dx^2}\int \kappa(x,y)f(y)\mathrm{d}y = \int \frac{\partial^2}{\partial x^2}\kappa(x,y)f(y)\mathrm{d}y.$$

This is due to the dominated convergence theorem, since both κ and f are bounded by the assumption. Thus, using the Cauchy-Schwarz inequality, we obtain

$$\|g''\|_\infty \le \sup_{x}\|\partial_x^2\kappa(x,\grave{)}\|\|f\| \le C_\kappa\|f\| \le C_\kappa C_f.$$

Now we use the Taylor expansion on ξ_{im_o} with the exact Peano remainder term. This gives for $\zeta_t \in [\xi_{o(i-1)+t,m}, \xi_{im_o}]$ the identity

$$\frac{1}{o} \sum_{t=1}^{o} g(\xi_{(i-1)o+t,m}) - g(\xi_{im_o})$$

$$= -g(\xi_{im_o}) + \frac{1}{o} \sum_{t=1}^{o} \left(g(\xi_{im_o}) + g'(\xi_{im_o})(\xi_{(i-1)o+t,m} - \xi_{im_o}) + \frac{g''(\zeta_t)}{2} \left(\xi_{(i-1)o+t,m} - \xi_{im_o} \right)^2 \right)$$

$$= \frac{1}{o} \sum_{t=1}^{o} \left(g'(\xi_{im_o})(\xi_{(i-1)o+t,m} - \xi_{im_o}) + \frac{g''(\zeta_t)}{2} \left(\xi_{(i-1)o+t,m} - \xi_{im_o} \right)^2 \right).$$

Next, we show that

$$\frac{1}{o} \sum_{t=1}^{o} g'(\xi_{im_o})(\xi_{(i-1)o+t,m} - \xi_{im_o}) = 0. \tag{A.16}$$

First,

$$\xi_{i,\frac{m}{o}} = \frac{1}{o} \sum_{t=1}^{o} \xi_{o(i-1)+t,m},$$

$$\xi_{(i-1)o+t,m} - \xi_{im_o} = \frac{2(o(i-1)+t)-1}{2m} - \frac{2i-1}{2\frac{m}{o}} = \frac{2t-o-1}{2m},$$

for $i = 1, ..., m/o$ and $t = 1, ..., o$. Furthermore, for o even, we have

$$\xi_{(i-1)o+t,m} - \xi_{im_o} = -\left(\xi_{(i-1)o+(o-t),m} - \xi_{im_o} \right)$$

for $i = 1, ..., m/o$ and $t = 1, ..., o/2$, while for o odd we have

$$\xi_{(i-1)o+t,m} - \xi_{im_o} = -\left(\xi_{(i-1)o+(o-t),m} - \xi_{im_o} \right)$$

for $i = 1, ..., m/o$ and $t = 1, ..., (o-1)/2$, and

$$\xi_{(i-1)o+\frac{o+1}{2},m} - \xi_{im_o} = 0.$$

This shows (A.16). Consequently,

$$\frac{1}{o} \sum_{t=1}^{o} g(\xi_{(i-1)o+t,m}) - g(\xi_{im_o}) = \frac{1}{o} \sum_{t=1}^{o} \frac{g''(\zeta_t)}{2} \frac{(2t-o-1)^2}{4m^2}$$

and we get

$$\|\bar{g}_{m_o} - g_{m_o]}\|^2 = \sum_{i=1}^{m_o} \left(\frac{1}{o} \sum_{t=1}^{o} g(\xi_{(i-1)o+t,m}) - g(\xi_{im_o}) \right)^2 \leq \frac{\|g''\|_{\infty}^2}{64 m^4 o^2} \sum_{i=1}^{m_o} \left(\sum_{t=1}^{o} (2t-o-1)^2 \right)^2$$

$$\leq \frac{\|g''\|_{\infty}^2}{64 m^4 o^2} m_o \left(\frac{1}{3}(o(o^2-1)) \right)^2 \leq \frac{\|g''\|_{\infty}^2}{9 \cdot 64 m_o^3},$$

which shows (A.15). Now, for the data propagation error (the variance), we derive

$$
\mathbb{E}\left[\left\|\sum_{j=1}^{k} \frac{(\overline{g}_{m_o}^{\delta} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}}{\sqrt{m_o}\tilde{\sigma}_{jm_o}} \tilde{v}_{jm_o}\right\|^2\right]
$$

$$
= \sum_{j=1}^{k} \frac{\mathbb{E}\left[(\overline{g}_{m_o}^{\delta} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}^2\right]}{m_o\tilde{\sigma}_{jm_o}^2} \|\tilde{v}_{jm_o}\|^2
$$

$$
+ \sum_{\substack{j,i \\ j \neq i}}^{k} \frac{\mathbb{E}\left[(\overline{g}_{m_o}^{\delta} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}\right] \mathbb{E}\left[(\overline{g}_{m_o}^{\delta} - g_{m_o}, \tilde{w}_{im_o})_{\mathbb{R}^m}\right]}{m_o\tilde{\sigma}_{jm_o}\sigma_{i,m_o}} (\tilde{v}_{jm_o}, \tilde{v}_{im_o})
$$

$$
\leq \frac{2}{m_o} \sum_{j=1}^{k} \frac{\frac{\delta^2}{o} + (\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}}{\tilde{\sigma}_{jm_o}^2} + \frac{2C_K^2}{3m_o^3} \sum_{\substack{i,j \\ i \neq j}}^{k} \frac{(\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}(\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{im_o})_{\mathbb{R}^m}}{\tilde{\sigma}_{jm_o}^2 \sigma_{i,m_o}^2}
$$

$$
\leq \frac{4\delta^2}{m} \sum_{j=1}^{k} \frac{1}{\sigma_j^2} + \frac{4}{m_o} \sum_{j=1}^{k} \frac{(\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}^2}{\sigma_j^2} + \frac{8C_K^2}{3m_o^3} \left(\sum_{j=1}^{k} \frac{(\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}}{\sigma_j^2}\right)^2
$$

$$
\leq \frac{4\delta^2}{m} \sum_{j=1}^{k} \frac{1}{\sigma_j^2} + \frac{4}{m_o\sigma_k^2} \sum_{j=1}^{k} (\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}^2
$$

$$
+ \frac{8C_K^2}{3m_o^3} \left(\sum_{j=1}^{k} \frac{1}{\sigma_j^4}\right)\left(\sum_{j=1}^{k} (\overline{g}_{m_o} - g_{m_o}, \tilde{w}_{jm_o})_{\mathbb{R}^m}^2\right)
$$

$$
\leq \frac{4\delta^2}{m} \sum_{j=1}^{k} \frac{1}{\sigma_j^2} + \frac{4}{m_o\sigma_k^2} \|\overline{g}_{m_o} - g_{m_o}\|_{\mathbb{R}^m}^2 + \frac{8C_K^2}{3m_o^3} \|\overline{g}_{m_o} - g_{m_o}\|_{\mathbb{R}^m}^2 \sum_{j=1}^{k} \frac{1}{\sigma_j^4}
$$

$$
\leq \frac{4\delta^2}{m} \sum_{j=1}^{k} \frac{1}{\sigma_j^2} + \frac{\|g''\|_{\infty}^2}{9 \cdot 16 m_o^4 \sigma_k^2} + \frac{C_K^2 \|g''\|_{\infty}^2}{9 \cdot 24 m_o^6} \sum_{j=1}^{k} \frac{1}{\sigma_j^4}.
$$

For the remaining error terms, we simply replace m with m_o and obtain

$$
\sqrt{\mathbb{E}\|\overline{\tilde{f}}_{k,m_o} - f\|^2}
$$

$$
\leq \frac{2\delta}{\sqrt{m}} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^2}} + \frac{\|g''\|_{\infty}}{12 m_o^2 \sigma_k} + \frac{C_K \|g''\|_{\infty}}{6\sqrt{6} m_o^3} \sqrt{\sum_{j=1}^{k} \frac{1}{\sigma_j^4}} + \frac{C_1}{m_o^2} \sqrt{\sum_{j=1}^{k} \frac{1}{c_j^2 \sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty} (f, v_j)^2}.
$$

This completes the proof. $\square$

Ackknowledgments

References

1. R. Anderssen and M. Hegland. For numerical differentiation, dimensionality can be a blessing! *Mathematics of Computation*, 68(227):1121–1141, 1999.
2. M. Bebendorf. Approximation of boundary element matrices. *Numerische Mathematik*, 86:565–589, 2000.
3. G. Golub and C. Van Loan. *Matrix computations*. JHU press, 2013.
4. L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73 (2):325–348, 1987.
5. M. Gu and S. Eisenstat. Efficient algorithms for computing a strong rank-revealing qr factorization. *SIAM Journal on Scientific Computing*, 17(4):848–869, 1996.
6. W. Hackbusch. A sparse matrix arithmetic based on H-matrices. Part I: Introduction to H-matrices. *Computing*, 62(2):89–108, 1999.
7. N. Halko, P.-G. Martinsson, and J. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.
8. M. Hanke and O. Scherzer. Inverse problems light: Numerical differentiation. *The American Mathematical Monthly*, 108(6):512–521, 2001.
9. P. Hansen. Regularization Tools version 4.0 for Matlab 7.3. *Numer. Algorithms*, 46(2), 2007.
10. K. Ito and B. Jin. Regularized linear inversion with randomized singular value decomposition. In *Centre International de Rencontres Mathématiques*, pages 45–72. Springer, 2019.
11. T. Jahn. Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise. *Numerical Functional Analysis and Optimization*, 43(2):145–167, 2022.
12. T. Jahn. Discretisation-adaptive regularisation of statistical inverse problems. *arXiv preprint arXiv:2204.14037*, 2022.
13. T. Jahn. A probabilistic oracle inequality and quantification of uncertainty of a modified discrepancy principle for statistical inverse problems. *Electronic Transactions on Numerical Analysis*, 57:35–56, 2022.
14. T. Jahn. Convergence of generalized cross-validation for an ill-posed integral equation. *Institut für Numerische Simulation*, 2303, 2023.
15. T. Jahn. Noise level free regularization of general linear inverse problems under unconstrained white noise. *SIAM/ASA Journal on Uncertainty Quantification*, 11(2):591–615, 2023.
16. N. Kumar and J. Schneider. Literature survey on low rank approximation of matrices. *Linear and Multilinear Algebra*, 65(11):2212–2244, 2017.
17. R. Lyons. *Understanding Digital Sgnal Processing*. Pearson Education India, 2004.
18. P. Mathé and S. Pereverzev. Complexity of linear ill-posed problems in Hilbert space. *Journal of Complexity*, 38:50–67, 2017.
19. V. Rokhlin, A. Szlam, and M. Tygert. A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1100–1124, 2010.
20. E. Tyrtyshnikov. Incomplete cross approximation in the mosaic-skeleton method. *Computing*, 64:367–380, 2000.