Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Friedrich-Hirzebruch-Allee 7 • 53115 Bonn • Germany
phone +49 228 73-69828 • fax +49 228 73-69847
www.ins.uni-bonn.de

T. Jahn

Convergence of generalized cross-validation for an ill-posed integral equation

INS Preprint No. 2303

December 2023

Convergence of generalized cross-validation for an ill-posed integral equation*

Tim Jahn†

Abstract. In this article we rigorously show consistency of generalized cross-validation applied to an exemplary ill-posed integral equation, given a finite number of noisy point evaluations. In particular, we present non-asymptotic order-optimal error estimates in probability. Hereby it is remarkable that the unknown true solution is not required to fulfill a self-similarity condition, which is generally needed for other heuristic parameter choice rules.

Key words. statistical inverse problems, generalized cross-validation, consistency, error estimates

MSC codes.

1. Introduction.

Generalized cross validation (GCV) is a popular parameter choice rule for regularized solution of ill-posed inverse problems. It is based on dividing the data into two parts, where the first fraction is used to construct a solution candidate for the task, while the second fraction is used to validate the performance of the candidate, see e.g. Stone [19] for a classic reference or, more recently, Hastie et. al [8] and Arlot & Celisse [1]. The generalized cross-validation technique analyzed here goes back to Wahba & Craven [6], who used it for spline smoothing of noisy point evaluations of a function. One distinct feature of the rule is that neither knowledge of the noise level nor knowledge of the smoothness of the unknown function is required. In its original form, 'leaving-one-out', one tries to fit a spline to all but one datum, and takes the error of the unused datum as the quality criteria, where one varies a so called smoothing parameter to balance how well the candidate fits the data points with the norm of the candidate. Ultimately, this results in a minimization problem over the smoothing parameter. In the similar framework of inverse integral equations the same method has been applied for choosing the regularization parameter by Wahba [21], Vogel [20], Lukas [15] and others. Extending the original fields of application, GCV and its variants have established themselves as some of the main re-sampling methods in high-dimensional statistics, data science and machine learning, see Witten & Frank [22], Kuhn & Johnson [11] or Giraud [7] for an overview. Given the importance of GCV as a practical rule in these areas, in this article we aim to shed some light on the theoretical properties of the original method.

In general one differs between two types of convergence results for cross-validation. The vast majority is of weak type. This means that not properties of the minimizer of the (random) data-driven functional are investigated, but properties of the minimizer of the population counterpart of that functional. While convergence results for minimizes of the expected value give valuable insight into the problem, from a statistical perspective, they do not even guarantee consistency of the original method. For inverse integral equations there are yet no strong

---

*Submitted to the editors DATE.

Funding: Funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy - GZ 2047/1, Projekt-ID 390685813.

†Institut für Numerische Simulation and Hausdorff Center for Mathematics, University of Bonn, Germany (jahn@ins.uni-bonn.de).

convergence results for GCV. Given the inherent instability of inverse integral equations, this is clearly unsatisfactory. The major contribution of this manuscript is a convergence analysis for GCV applied to inverse integral equations of strong type, that is where properties of the minimizer of the random data-driven functional are studied.

Such strong results have been obtained in some other settings, as e.g. spline smoothing or model selection, by Speckman [18] and Li [12, 13]. Moreover, there exists a consistency result in the framework of semi-supervised statistical learning from Caponnetto & Yao [5]. However, we will not follow the approach from Li, which is based on comparison to Stein-estimators. Consequently, our result will not be a straightforward generalization of the approach from Li and takes a different form. For example, Li showed that generalized cross-validation is asymptotically optimal for model selection, as the number of point evaluations tends to infinity, while the noise level δ and the smoothness of the exact solution are kept fixed. As a preliminary result in Corollary 3.6 below, we show that generalized cross-validation is order-optimal (that is optimal up to a constant, which is weaker than asymptotic optimality), however this bound is guaranteed to hold also in the non-asymptotic regime.

Apart from showing the consistency of GCV, we also carefully analyze the discretization error, which is often not taken into account. While the integral equation is formulated in an inherently infinite-dimensional setting, through the finite number of measurement points a discrete model is induced. Moreover, the cross-validation method can only be formulated in the finite-dimensional setting, and in most works no error estimates of the constructed estimator to the continuous solution are given. Here we will give the complete picture, that is we give a strong consistency result for our cross-validation estimator and show convergence to the continuous solution, when the number of point evaluations tend to infinity. We do this for a concrete explicit yet not trivial example and also show paths how to extend the results to more general settings.

2. Setting and main result.

We will analyze the following integral equation

$$(Kf)(x) = \int_0^1 \kappa(x,y)f(y)\mathrm{d}x, \tag{2.1}$$

with $\kappa(x,y) := \min(x(1-y), y(1-x))$. Note that several results developed in this article will hold for general continuous κ also. We have access to noisy point evaluations

$$g^\delta_{j,m} := g^\dagger(\xi_{j,m}) + \delta\varepsilon_j, \quad j = 1, ...m, \tag{2.2}$$

where $g^\dagger = Kf^\dagger$ is the unknown exact data, $\xi_{j,m} := j/(m+1) \in (0,1)$ are the evaluation points, $\delta > 0$ is the noise level and ε_j are unbiased i.i.d random variables with unit variance. The goal is to reconstruct the exact solution $f^\dagger$. Through (2.1) a compact operator $K : L^2(0,1) \to L^2(0,1)$ is defined. Moreover, continuity of κ implies that Kf is continuous even if f is only square-integrable. The above equation (2.1) is ill-posed and hence needs to be regularized. For that we rely on spectral methods using the spectral decomposition of the induced discretization of K, which we will denote by K_m and define as follows:

$$\begin{aligned} K_m &: L^2(0,1) \to \mathbb{R}^m \\ f &\mapsto ((Kf)(\xi_{j,m}))_{j=1}^m = \left(\int \kappa(\xi_{j,m},y)f(y)\mathrm{d}y\right)_{j=1}^m, \end{aligned}$$

with $j = 1, ..., m$. We will assume from now a uniform discretization, i.e., $\xi_{j,m} := j/(m+1)$. The setting here is particularly simple, since we can give the exact singular value decomposition of K and K_m:

Lemma 2.1. *For $\lambda_k := \pi^2 k^2 =: \sigma_k^{-1}$ and $v_k(x) := \sqrt{2}\sin(\sqrt{\lambda_k}x)$ there holds $K^*Kv_k = \sigma_k^2 v_k$ for all $k \in \mathbb{N}$ and the $(v_k)_{k\in\mathbb{N}}$ form an orthonormal basis of $\mathcal{N}(K)^\perp \subset L^2(0,1)$. Moreover, for*

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

and

$$v_{k,m}(\cdot) := \sum_{l=1}^{m} \sin\left(\sqrt{\lambda_k}\xi_l\right)\kappa(\xi_{l,m},\cdot)/\sigma_{k,m} \quad \text{and} \quad u_{k,m} := \sqrt{\frac{2}{m+1}}\left(\sin(k\pi\xi_{j,m})\right)_{j=1}^m$$

it holds that $K_m v_{k,m} = \sigma_{k,m}u_{k,m}$ and $K_m^* u_{k,m} = \sigma_{k,m}v_{k,m}$, with $(v_{k,m})_{k\le m}$ and $(u_{k,m})_{k\le m}$ orthonormal bases of $\mathcal{N}(K_m)^\perp \subset L^2(0,1)$ and $\mathbb{R}^m$ respectively.

The proof will be given below in Section A. We define an approximation to the unknown $f^\dagger$ via spectral cut-off and set

$$(2.3) \qquad f_{k,m}^\delta := \sum_{j=1}^{k} \frac{(g_m^\delta, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} v_{j,m}$$

and the ultimate goal will be to determine the stopping index $k \le m$ dependent only on m (and without knowledge of δ or assumptions on the smoothness of $f^\dagger$). For the determination of the truncation index k we choose generalized cross-validation due to Wahba. It is defined as follows:

$$(2.4) \qquad k_m = k_m(\delta, f^\dagger, g_m^\delta) = \arg\min_{k=0,\dots,\frac{m}{2}} \frac{\sum_{j=k+1}^m (g_m^\delta, u_{j,m})^2}{\left(1-\frac{k}{m}\right)^2} =: \arg\min_{k=0,\dots,\frac{m}{2}} V_m(k).$$

This choice was introduced by Vogel [20] and can be derived from the original method from Wahba [21], when Tikhonov regularization is replaced with spectral cut-off regularization. The only difference to [20] is that the minimizing set is restricted to $k \le m/2$ instead of $k \le m$. Other choices, say $k \le \frac{2}{3}m$ would be possible as well, as long as it is avoided that single random coefficients dominate the functional. In [20] such restriction was not needed, since there the expectation of the functional was considered. Note that the cross-validation functional V_m is kind of an approximation of the weak (predictive) norm

$$S_m(k) := \|K_m f_{k,m}^\delta - K_m f^\dagger\|^2 = \sum_{j=1}^{k} \delta^2 \varepsilon_j^2 + \sum_{j=k+1}^{m} \sigma_{j,m}^2 (f^\dagger, v_{j,m})^2.$$

In fact, it holds that

$$\mathbb{E}[S_m(k)] = k\delta^2 + \sum_{j=k+1}^{m} \sigma_{j,m}^2 (f^\dagger, v_{j,m})^2 \tag{2.5}$$

$$\mathbb{E}[V_m(k)] = \frac{(m-k)\delta^2 + \sum_{j=k}^{m} \sigma_{j,m}^2 (f^\dagger, v_{j,m})^2}{\left(1-\frac{k}{m}\right)^2}. \tag{2.6}$$

As already mentioned in the introduction, most results for cross-validation are of weak form, in the sense that they do not investigate k_m, but rather $k_m^* = \arg\min_k \mathbb{E}[V_m(k)]$. The results are usually that $k_m^* = (1+o(1)) \arg\min_k \mathbb{E}[S_m(k)]$ (as $m \to \infty$) under certain assumptions on the singular value decomposition of K, K_m and $f^\dagger$, and the constants hidden in $o(1)$ are not given or unknown. In this note we will investigate the data-driven choice k_m, and we will exactly calculate all involved constants. It is classic to calculate this error explicitly assuming that $f^\dagger$ belongs to some unknown subset of L^2 with a certain smoothness. For the given kernel we define the subsets as Hölder source conditions

$$\mathcal{X}_{s,\rho} := \left\{ f = (K^*K)^{\frac{s}{2}} h \ : \ h \in L^2, \ \|h\| \le \rho \right\}.$$

Below we will relate $\mathcal{X}_{s,\rho}$ to classical smoothness in Proposition 3.9. We will use the following function to quantify the uncertainty of our estimator. For $t \in \mathbb{N}$ and $\varepsilon \le \frac{1}{12}$, set

$$p_\varepsilon(t) := \frac{3}{\varepsilon} \mathbb{E}\left[\left| \frac{1}{t} \sum_{j=1}^{t} (\varepsilon_j^2 - 1) \right| \right].$$

Clearly, since the ε_j's are unbiased with unit variance, we have $p_\varepsilon(t) \to 0$ as $t \to \infty$. We are ready to formulate our main result:

Theorem 2.2. *Assume that $s > \frac{3}{4}$. Then, uniformly over $f^\dagger \in \mathcal{X}_{s,\rho}$, the probability that*

$$\begin{aligned} &\|f^\delta_{k_{\mathrm{gcv}},m} - f^\dagger\| \\ &\le L_s' \left(\frac{\delta}{\sqrt{m+1}} \right)^{\frac{4s}{5+4s}} \rho^{\frac{5}{5+4s}} + L_s'' \frac{\rho}{m^{2s}} + \frac{\|{f^\dagger}'\|}{\sqrt{2}(m+1)} \chi_{\{\frac{3}{4} < s \le \frac{5}{4}\}} + \frac{\|{f^\dagger}''\|}{2(m+1)^2} \chi_{\{s > \frac{5}{4}\}} \end{aligned}$$

is larger then $1 - p_\varepsilon \left(\frac{2}{3} \frac{\varepsilon}{\varepsilon+1} C_s \left(\frac{(m+1)\rho^2}{\delta^2} \right)^{\frac{1}{5+4s}} \right)$, where the constants L_s', L_s'' and C_s are given below in (3.11) *and* (3.7).

We comment on the result. The first term in the upper bound resembles the optimal convergence rate for the source condition $\mathcal{X}_{s,\rho}$ in the idealized functional white noise model with variance $\frac{\delta^2}{m+1}$, for m the number of point evaluations tending to infinity. In the latter model we again seek the solution $Kf = g$, but instead of having m noisy point evaluations, we can measure scalar products (g^δ, h) with $h \in L^2$. Hereby, the latter has the same distribution as $(g^\dagger, h) + \frac{\delta}{\sqrt{m+1}} \varepsilon_1$. The second term comes from the restriction $k_{\mathrm{gcv}}^\delta \le \frac{m}{2}$ and usually is dominated by the first term, unless the noise level δ is very small. The remaining two terms are upper bounds for the discretization error, under different smoothness s of the exact solution and expresses how good the exact solution $f^\dagger$ can be represented in the span of $\kappa(\xi_{1,m},\cdot),...,\kappa(\xi_{m,m},\cdot)$ (note that those span the space of piece-wise linear functions on the grid given by $\xi_{1,m},...,\xi_{m,m}$). Note that the assumption $s>\frac{3}{4}$ imposes a substantial differentiability condition onto the solution $f^\dagger$. If this assumption is violated a similar bound will still hold, however it is not possible to explicitly bound the aforementioned discretization error anymore.

A key advantage of GCV is that it does not require any knowledge of the noise level δ. Therefore it belongs to the class of heuristic parameter choice rules. The term heuristic stems from the fact that these rules provably do not assemble convergent regularization schemes under a classical deterministic worst-case noise model, due to the seminal work by Bakushinskii [2]. Still, for the white noise error model some heuristic parameter choice rules, i.e. the quasi-opimality criterion and the heuristic discrepancy principle yield convergent regularization methods, see Bauer & Reiß [3] and Jahn [10]. In order to prove mini-max optimality for those approaches, however additional to the classical source condition the true solution must fulfill a self-similarity condition, which is a substantial structural assumption as it demands a concrete relation between the high and low frequency parts of the unknown solution. Therefore it is remarkable that GCV yields mini-max optimality without assuming self-similarity. On the other hand, as will be explained below, the GCV is probably not consistent for general ill-posed problems, as it might lack stability for exponentially falling singular values. Such limitations regarding the robustness for exponentially ill-posed problems have recently been studied for several related methods based on unbiased risk estimation from Lucka & al [14].

We finally mention here modifications of GCV which are designed to improve the stability of the method when applied to inverse problems. Those methods were developed by Lukas and are called robust and strong robust cross-validation, see [16] and [17].

3. Proof of the main result.

We first prove the following lemma which holds for general kernel κ and evaluation points. Note however that in this case the singular system $(\sigma_{j,m},v_{j,m},u_{j,m})$ of K_m is not computable and has to be approximated numerically. Here no source condition is required, but we define the so called weak and strong oracles for each individual $f^\dagger$:

$$t_m^\delta := t_m^\delta(f^\dagger) := \max\left\{0\le k\le m \;:\; k\delta^2 \le \sum_{j=k+1}^m \sigma_{j,m}^2 (f^\dagger, v_{j,m})^2\right\}, \tag{3.1}$$

$$s_m^\delta := s_m^\delta(f^\dagger) := \max\left\{0\le k\le m \;:\; \frac{k\delta^2}{\sigma_{k,m}^2} \le \sum_{j=k+1}^m (f^\dagger, v_{j,m}^2)\right\}. \tag{3.2}$$

Lemma 3.1. *For L_s and C_a given below and uniformly for all $f^\dagger$ with $t_m^\delta(f^\dagger)\ge t\in\mathbb{N}$, it holds that*

$$\mathbb{P}\left(\left\|f^\delta_{k_{\mathrm{gcv}},m} - P_{\mathcal{N}(K_m)^\perp} f^\dagger\right\| \le \frac{L_s\sqrt{s_m^\delta}\delta + C_a\sqrt{\sum_{j=\frac{m}{2}+1}^m (g_m^\dagger, u_{j,m})^2_{\mathbb{R}^m}}}{\sigma_{\frac{s_m^\delta}{\varepsilon^2},m}}\right) \ge 1 - p_\varepsilon\left(\frac{2}{3}\frac{\varepsilon}{1+\varepsilon}t\right).$$

Remark 3.2. The above is kind of an oracle inequality for our estimator with respect to the projected exact solution. While the second summand of the denominator is due to the constraint $k^\delta_{\mathrm{gcv}} \le \frac{m}{2}$ and is usually negligible, the fact that we have $\sigma_{\frac{s^\delta_m}{\varepsilon^2},m}$ instead of $\sigma_{s^\delta_m,m}$ in the nominator is more sincere, since this term explodes for rapidly falling singular values.

Proof of Lemma 3.1. For the analysis we define the event

$$\Omega_t := \left\{ \left| \sum_{j=k+1}^{l} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} - (l-k)\delta^2 \right| \le \varepsilon(l-k)\delta^2,\ \forall l \ge t,\ k \le \frac{l}{2} \right\}. \tag{3.3}$$

On Ω_t we can control the random errors, and for its probability we claim that

$$\mathbb{P}(\Omega_t) \ge 1 - p_\varepsilon\left(\frac{2}{3}\frac{\varepsilon}{1+\varepsilon}t\right). \tag{3.4}$$

Remark 3.3. Note that if $l \ge t$, but $\frac{l}{2} < k \le l$, we will occasionally use the upper bound

$$\sum_{j=k+1}^{l} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} \chi_{\Omega_t} \le \sum_{j=1}^{l} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} \chi_{\Omega_t} \le (1+\varepsilon) l \delta^2.$$

We first prove the claim (3.4) and define, for $\varepsilon' := \frac{2}{3}\frac{\varepsilon}{1+\varepsilon}$,

$$\Omega_t' := \left\{ \left| \frac{1}{l} \sum_{j=1}^{l} \left((g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} - \delta^2 \right) \right| \le \frac{\varepsilon}{3},\ \forall l \ge \varepsilon' t \right\}.$$

Using the Kolmogorov-Doob inequality for backwards martingales one can prove that (see, e.g., Proposition 4.1 of [9])

$$\mathbb{P}(\Omega_t') \ge 1 - \frac{3}{\varepsilon}\mathbb{E}\left[\left| \frac{1}{\varepsilon' t} \sum_{j=1}^{\varepsilon' t} (\varepsilon_j^2 - 1) \right| \right] = 1 - p_\varepsilon(\varepsilon' t).$$

and it remains to show that $\Omega_t' \subset \Omega_t$. For this, we refine the argumentation in the proof of Proposition 3.1 of [10]. So let $l \ge t$ and first assume that $k \ge \varepsilon' l$. Then $k \ge \varepsilon' t$ and thus

$$\begin{aligned} \sum_{j=k+1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} &= \sum_{j=1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} - \sum_{j=1}^{k} \varepsilon_j^2 \chi_{\Omega_t'} \le \left(1 + \frac{\varepsilon}{3}\right) l - \left(1 - \frac{\varepsilon}{3}\right) k = (1+\varepsilon)(l-k) - \frac{2}{3}\varepsilon l + \frac{4}{3}\varepsilon k \\ &\le (1+\varepsilon)(l-k), \end{aligned}$$

since $k \le l/2$. Similar, $\sum_{j=k+1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} \ge (1-\varepsilon)(l-k)\chi_{\Omega_t'}$. For $k < \varepsilon' l$, we obtain

$$\begin{aligned} \sum_{j=k+1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} &\le \sum_{j=1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} \le \left(1 + \frac{\varepsilon}{3}\right) l = (1+\varepsilon)(l-k) - \frac{2}{3}\varepsilon l + (1+\varepsilon) k \\ &\le (1+\varepsilon)(l-k) - \frac{2}{3}\varepsilon l + (1+\varepsilon)\varepsilon' l = (1+\varepsilon)(l-k), \end{aligned}$$

by definition of ε'. Finally,

$$\sum_{j=k+1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} \geq \sum_{j=\varepsilon' l+1}^{l} \varepsilon_j^2 \chi_{\Omega_t'} \geq \left(1-\frac{\varepsilon}{3}\right) l\chi_{\Omega_t'} - \left(1+\frac{\varepsilon}{3}\right)\varepsilon' l \chi_{\Omega_t'}$$
$$= (1-\varepsilon)(l-k)\chi_{\Omega_t'} + \left(\frac{2}{3}\varepsilon - \left(1+\frac{\varepsilon}{3}\right)\varepsilon'\right) l\chi_{\Omega_t'} + (1-\varepsilon)k\chi_{\Omega_t'}$$
$$= (1-\varepsilon)(l-k)\chi_{\Omega_t'} + (1-\varepsilon)k\chi_{\Omega_t'} \geq (1-\varepsilon)(l-k)\chi_{\Omega_t'}.$$

This proves $\Omega_t' \subset \Omega_t$ and therefore the claim (3.4).

In the following we fix $\varepsilon \leq \frac{1}{12}$. We first show stability.

Proposition 3.4. *For $t \leq t_m^\delta$ it holds that $k_{\mathrm{gcv},m}^\delta \chi_{\Omega_t} \leq \frac{t_m^\delta}{\varepsilon^2}$.*

Proof of Proposition 3.4. It suffices to show that

$$\Psi_m(t_m^\delta)\chi_{\Omega_t} < \Psi_m(k) \tag{3.5}$$

for all $\frac{t_m^\delta}{\varepsilon^2} < k \leq \frac{m}{2}$. By definition of ε, in this case $t_m^\delta < \frac{k}{2}$. Now, on the one hand

$$\Psi_m(t_m^\delta)\chi_{\Omega_t}$$
$$= \frac{\sum_{j=t_m^\delta+1}^m (g_m^\delta, u_{j,m})_{\mathbb{R}^m}^2}{\left(1-\frac{t_m^\delta}{m}\right)^2}\chi_{\Omega_t} = \frac{\sum_{j=t_m^\delta+1}^k (g_m^\delta, u_{j,m})_{\mathbb{R}^m}^2}{\left(1-\frac{t_m^\delta}{m}\right)^2}\chi_{\Omega_t} + \frac{\sum_{j=k+1}^m (g_m^\delta, u_{j,m})_{\mathbb{R}^m}^2}{\left(1-\frac{t_m^\delta}{m}\right)^2}\chi_{\Omega_t}$$
$$\leq \frac{\left(\sqrt{\sum_{j=t_m^\delta+1}^k (g_m^\delta - g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2} + \sqrt{\sum_{j=t_m^\delta+1}^k (g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2}\right)^2}{\left(1-\frac{t_m^\delta}{m}\right)^2}\chi_{\Omega_t} + \left(\frac{1-\frac{k}{m}}{1-\frac{t_m^\delta}{m}}\right)^2 \Psi_m(k)$$
$$\leq \frac{\left((1+\varepsilon)\sqrt{k}\delta + \sqrt{t_m^\delta}\delta\right)^2}{\left(1-\frac{t_m^\delta}{m}\right)^2} + \left(\frac{m-k}{m-t_m^\delta}\right)^2 \Psi_m(k)$$
$$\leq \frac{\left((1+\varepsilon)\sqrt{k}\delta + \sqrt{\varepsilon^2 k}\delta\right)^2}{\left(1-\frac{t_m^\delta}{m}\right)^2} + \left(\frac{m-k}{m-t_m^\delta}\right)^2 \Psi_m(k) \leq \frac{(1+2\varepsilon)^2 k\delta^2}{\left(1-\frac{t_m^\delta}{m}\right)^2} + \left(\frac{m-k}{m-t_m^\delta}\right)^2 \Psi_m(k)$$

Note that $k \leq m-k$ and $t_m^\delta \leq \varepsilon^2 k$. Then, on the other hand,

$$\Psi_m(k)\chi_{\Omega_t} \geq \frac{\left(\sqrt{\sum_{j=k+1}^m (g_m^\delta - g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2} - \sqrt{\sum_{j=k+1}^m (g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2}\right)^2}{(1-\frac{k}{m})^2}\chi_{\Omega_t}$$
$$\geq \frac{\left((1-\varepsilon)\sqrt{m-k}\delta - \sqrt{t_m^\delta}\delta\right)^2}{(1-\frac{k}{m})^2}\chi_{\Omega_t} \geq \frac{\left((1-\varepsilon)\sqrt{m-k}\delta - \varepsilon\sqrt{k}\delta\right)^2}{(1-\frac{k}{m})^2}\chi_{\Omega_t}$$
$$\geq \frac{((1-\varepsilon)\sqrt{m-k}\delta - \varepsilon\sqrt{m-k}\delta)^2}{(1-\frac{k}{m})^2}\chi_{\Omega_t} \geq \frac{(1-2\varepsilon)^2(m-k)\delta^2}{(1-\frac{k}{m})^2}\chi_{\Omega_t}$$

We solve the second inequality for δ and plug into the first equation and obtain

$$\begin{aligned}
\Psi_m(t_m^\delta)\chi_{\Omega_t} &\le \frac{(1+2\varepsilon)^2 k\delta^2}{\left(1-\frac{t_m^\delta}{m}\right)^2}\chi_{\Omega_t} + \left(\frac{m-k}{m-t_m^\delta}\right)^2\Psi_m(k) \\
&\le \frac{(1+2\varepsilon)^2 k}{\left(1-\frac{t_m^\delta}{m}\right)^2}\frac{\left(1-\frac{k}{m}\right)^2}{(1-2\varepsilon)^2(m-k)}\Psi_m(k) + \left(\frac{m-k}{m-t_m^\delta}\right)^2\Psi_m(k) \\
&= \Psi_m(k)\frac{m-k}{(m-t_m^\delta)^2}\left(k\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2 + m-k\right) \\
&= \Psi_m(k)\frac{m^2-\left(2-\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2\right)mk-\left(\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2-1\right)k^2}{m^2-2mt_m^\delta+{t_m^\delta}^2} \\
&< \Psi_m(k)\frac{m^2-\left(2-\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2\right)mk}{m^2-2mt_m^\delta} < \Psi_m(k),
\end{aligned}$$

since

$$\frac{2-\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2}{2}\frac{k}{t_m^\delta} \ge \frac{2-\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^2}{2\varepsilon^2} > 1$$

for $\varepsilon \le 1/12$. This proves that

$$\min_{\frac{t_m^\delta}{\varepsilon^2}\le k\le\frac{m}{2}} \Psi_m(k) > \Psi_m(t_m^\delta)$$

and hence $k_{\mathrm{gcv}}^\delta\chi_{\Omega_t} = \chi_{\Omega_t}\arg\min_{0\le k\le\frac{m}{2}}\Psi_m(k) < t_m^\delta/\varepsilon^2$. ■

The upper bound for $k_{\mathrm{gcv},m}^\delta$ directly yields an (up to a multiplicative constant optimal) bound for the (weak) data propagation error. We now deduce a bound for the (weak) approximation error also.

Proposition 3.5. *Let $t \le t_m^\delta$. If $t_m^\delta \le \frac{m}{2}$ it holds that*

$$\sum_{j=k_{\mathrm{gcv},m}^\delta+1}^{m} (g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2\chi_{\Omega_t} \le C_a t_m^\delta\delta^2$$

with $C_a := 35+34\varepsilon$, and if $t_m^\delta > \frac{m}{2}$ it holds that

$$\sum_{j=k_{\mathrm{gcv},m}^\delta+1}^{m} (g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2\chi_{\Omega_t} \le C_a' \sum_{j=\frac{m}{2}+1}^{m} (g_m^\dagger, u_{j,m})_{\mathbb{R}^m}^2$$

with $C_a' = 12+8\varepsilon$.

Proof of Proposition 3.5. Since $C_a \geq 1$ the assertion clearly holds for $k^\delta_{\mathrm{gcv},m}\chi_{\Omega_t} > t^\delta_m$. Now assume $k^\delta_{\mathrm{gcv},m} < t^\delta_m$ and $t^\delta_m \leq m/2$. Then, by definition of t^δ_m,

$$\begin{aligned}
&\sum_{j=g^\delta_{\mathrm{gcv},m}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} \\
&= \sum_{j=g^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} + \sum_{j=t^\delta_m+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} \leq \sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} + t^\delta_m\delta^2 \\
&\leq 2\sum_{j=k_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\delta_m, u_{j,m})^2_{\mathbb{R}^m} + 2\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} + t^\delta_m\delta^2 \\
&\leq 2\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\delta_m, u_{j,m})^2_{\mathbb{R}^m} + (3+2\varepsilon)t^\delta_m\delta^2.
\end{aligned}$$

Because

$$\begin{aligned}
\Psi_m(k^\delta_{\mathrm{gcv},m}) &= \frac{\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{m}(g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2} = \frac{\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m}(g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2} + \frac{\sum_{j=t^\delta_m+1}^{m}(g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2} \\
&= \frac{\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m}(g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2} + \left(\frac{m-t^\delta_m}{m-k^\delta_{\mathrm{gcv},m}}\right)^2\Psi_m(t^\delta_m)
\end{aligned}$$

we conclude, since $k^\delta_{\mathrm{gcv},m}$ is the minimizer of Ψ_m on $0 \le k \le m/2$ and $t^\delta_m \le \frac{m}{2}$,

$$
\begin{aligned}
&\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{t^\delta_m} (g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t}\\
&= \left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2 \Psi_m(k^\delta_{\mathrm{gcv},m})\chi_{\Omega_t} - \left(1-\frac{t^\delta_m}{m}\right)^2 \Psi_m(t^\delta_m)\chi_{\Omega_t}\\
&\le \left(1-\frac{k^\delta_{\mathrm{gcv},m}}{m}\right)^2 \Psi_m(t^\delta_m)\chi_{\Omega_t} - \left(1-\frac{t^\delta_m}{m}\right)^2 \Psi_m(t^\delta_m)\chi_{\Omega_t}\\
&= \frac{\Psi_m(t^\delta_m)}{m}\left(2t^\delta_m - 2k^\delta_{\mathrm{gcv},m} + \frac{{k^\delta_{\mathrm{gcv},m}}^2 - {t^\delta_m}^2}{m}\right)\chi_{\Omega_t} \le \frac{2t^\delta_m\Psi_m(t^\delta_m)}{m}\chi_{\Omega_t}\\
&= \frac{2t^\delta_m}{m}\frac{\sum_{j=t^\delta_m+1}^m (g^\delta_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{t^\delta_m}{m}\right)^2}\\
&\le \frac{4t^\delta_m}{m}\frac{\sum_{j=t^\delta_m+1}^m (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} + \sum_{j=t^\delta_m+1}^m (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}}{\left(1-\frac{t^\delta_m}{m}\right)^2}\\
&\le \frac{4t^\delta_m}{m}\frac{(1+\varepsilon)(m-t^\delta_m)\delta^2 + t^\delta_m\delta^2}{\left(1-\frac{t^\delta_m}{m}\right)^2} \le 4(1+\varepsilon)\frac{t^\delta_m\delta^2}{\frac{1}{2^2}} = 16(1+\varepsilon)t^\delta_m\delta^2.
\end{aligned}
$$

Putting everything together we obtain

$$
\sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} \le 32(1+\varepsilon)t^\delta_m\delta^2 + (3+2\varepsilon)t^\delta_m\delta^2 = (35+34\varepsilon)t^\delta_m\delta^2 = C_a t^\delta_m\delta^2.
$$

Finally, assume that $k^\delta_{\mathrm{gcv}} < t^\delta_m$ and $t^\delta_m > m/2$. Then, using $\frac{m}{2}\delta^2 < \sum_{j=\frac{m}{2}+1}^m (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}$ in this case, we get

$$
\begin{aligned}
&\sum_{j=k^\delta_{\mathrm{gcv}}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}\chi_{\Omega_t} \le 2\sum_{j=k^\delta_{\mathrm{gcv}}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} + 2\sum_{j=k^\delta_{\mathrm{gcv}}+1}^{m} (g^\delta_m - g^\dagger_m, u_{j,m})^2\\
&\le 2\left(1-\frac{k^\delta_{\mathrm{gcv}}}{m}\right)^2\Psi_m(k^\delta_{\mathrm{gcv}}) + 2(1+\varepsilon)m\delta^2 \le 2\left(1-\frac{k^\delta_{\mathrm{gcv}}}{m}\right)^2\Psi_m\left(\frac{m}{2}\right) + 2(1+\varepsilon)m\delta^2\\
&\le 4\sum_{j=\frac{m}{2}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} + 4\sum_{j=\frac{m}{2}+1}^{m} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} + 2(1+\varepsilon)m\delta^2\\
&\le (12+8\varepsilon)\sum_{j=\frac{m}{2}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} = C'_a\sum_{j=\frac{m}{2}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}.
\end{aligned}
$$

■

We move on to the main proof. Note that

$$(g^\dagger_m, u_{j,m})_{\mathbb{R}^m} = (K_m f^\dagger, u_{j,m})_{\mathbb{R}^m} = (f^\dagger, K^*_m u_{j,m}) = \sigma_{j,m}(f^\dagger, v_{j,m}).$$

Splitting the error yields

$$\begin{aligned} f^\delta_{k^\delta_{\mathrm{gcv},m},m} - P_{\mathcal{N}^\perp(K_m)} f^\dagger &= \sum_{j=1}^{k^\delta_{\mathrm{gcv},m}} \frac{(g^\delta_m, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} v_{j,m} - \sum_{j=1}^{m} (f^\dagger, v_{j,m}) v_{j,m} \\ &= \sum_{j=1}^{k^\delta_{\mathrm{gcv},m}} \frac{(g^\delta_m - g^\dagger_m, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} v_{j,m} - \sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{m} (f^\dagger, v_{j,m}) v_{j,m}. \end{aligned}$$

For the first term we obtain

$$\begin{aligned} \sum_{j=1}^{k^\delta_{\mathrm{gcv},m}} \frac{\left(g^\delta_m - g^\dagger_m, u_{j,m}\right)^2_{\mathbb{R}^m}}{\sigma^2_{j,m}} \chi_{\Omega_t} &\le \frac{1}{\sigma^2_{k_{\mathrm{gcv},m}}} \sum_{j=1}^{k^\delta_{\mathrm{gcv},m}} (g^\delta_m - g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m} \chi_{\Omega_t} \le (1+\varepsilon) \frac{k^\delta_{\mathrm{gcv},m}\delta^2}{\sigma^2_{k_{\mathrm{gcv},m}}} \chi_{\Omega_t} \\ &\le \frac{1+\varepsilon}{\varepsilon^2} \frac{t^\delta_m \delta^2}{\sigma^2_{\frac{t^\delta_m}{\varepsilon^2}}}, \end{aligned}$$

and for the second,

$$\begin{aligned} \sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{m} (f^\dagger, v_{j,m}) \chi_{\Omega_t} &= \sum_{j=k^\delta_{\mathrm{gcv},m}+1}^{s^\delta_m} (f^\dagger, v_{j,m})^2 \chi_{\Omega_t} + \sum_{j=s^\delta_m+1}^{m} (f^\dagger, v_{j,m})^2 \\ &\le \frac{1}{\sigma^2_{s^\delta_m,m}} \sum_{k^\delta_{\mathrm{gcv},m}+1}^{s^\delta_m} (g^\dagger, u_{j,m})^2_{\mathbb{R}^m} \chi_{\Omega_t} + \frac{s^\delta_m \delta^2}{\sigma^2_{s^\delta_m,m}} \\ &\le \frac{1}{\sigma^2_{s^\delta_m,m}} \sum_{k^\delta_{\mathrm{gcv},m}+1}^{m} (g^\dagger, u_{j,m})^2_{\mathbb{R}^m} \chi_{\Omega_t} + \frac{s^\delta_m \delta^2}{\sigma^2_{s^\delta_m,m}}. \end{aligned}$$

Combining the preceding both estimates and using Proposition 3.5 together with the fact that $t^\delta_m(f^\dagger) \le s^\delta_m(f^\dagger)$, we conclude

$$\left\| f^\delta_{k^\delta_{\mathrm{gcv},m},m} - P_{\mathcal{N}^\perp(K_m)} f^\dagger \right\| \chi_{\Omega_t} \le \frac{L_s \sqrt{s^\delta_m}\,\delta + C'_a \sqrt{\sum_{j=\frac{m}{2}+1}^{m} (g^\dagger_m, u_{j,m})^2_{\mathbb{R}^m}}}{\sigma_{\frac{s^\delta_m}{\varepsilon^2},m}}$$

with $L_s := \frac{\sqrt{1+\varepsilon}}{\varepsilon} + \sqrt{C_a + 1}$ and the proof of Lemma 3.1 is finished. ■

As a corollary of the preceding two propositions we formulate an oracle inequality for the empirical predictive error of our estimator. Note that it holds for arbitrary continuous kernel κ. For simplicity we exclude the case $t^\delta_m(f^\dagger) > \frac{m}{2}$, that is when the balancing weak oracle is not in the range of the cross-validation.

Corollary 3.6. *It holds that*

$$\inf_{\substack{f^\dagger \\ t\leq t_m^\delta(f^\dagger)\leq \frac{m}{2}}} \mathbb{P}\left(\|K_m f_{k,m}^\delta - g_m^\dagger\|_{\mathbb{R}^m} \leq \sqrt{\frac{1}{\varepsilon^2}+C_a}\sqrt{t_m^\delta}\delta\right) \geq 1 - p_\varepsilon\left(\frac{2}{3}\frac{\varepsilon}{1+\varepsilon}t\right).$$

We now use the concrete form of the singular value decomposition of the semi-discrete and the continuous operator to calculate the error to the continuous solution $f^\dagger$ for the proof of Theorem 2.2. The following Lemma gives a first estimate for s_m^δ uniformly over the source condition $\mathcal{X}_{s,\rho}$.

Lemma 3.7. *It holds that*

$$\sup_{f\in\mathcal{X}_{\nu,\rho}} s_m^\delta(f) \leq C_s\left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{1}{5+8s}},$$

$$\sup_{f\in\mathcal{X}_{\nu,\rho}} t_m^\delta(f) \leq C_s\left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{1}{5+8s}}.$$

with C_s given below in the proof.

Proof of Lemma 3.7. The following auxiliary proposition is needed and will be proved in Appendix B.

Proposition 3.8. *For $j = t(m+1)+s$ with $m\in\mathbb{N}, t\in\mathbb{N}_0$ and $s\in\{0,...,m\}, k\in\{1,...,m\}$, it holds that*

$$(v_j, v_{k,m}) = \sqrt{m+1}\frac{\sigma_j}{\sigma_{k,m}}\begin{cases}1 & \text{for } s=k \text{ and } t \text{ even}\\ -1 & \text{for } s+k=m+1 \text{ and } t \text{ odd}\\ 0 & \text{else}\end{cases}$$

By Proposition 3.8, it holds that

$$\begin{aligned} v_{j,m} &= \sum_{l=1}^\infty (v_{j,m}, v_l)v_l \\ &= \sqrt{m+1}\frac{\sigma_j}{\sigma_{j,m}}v_j - \sqrt{m+1}\sum_{t=1}^\infty \frac{\sigma_{2t(m+1)-j}v_{2t(m+1)-j} - \sigma_{2t(m+1)+j}v_{2t(m+1)+j}}{\sigma_{j,m}}. \end{aligned}$$

Therefore, with $f^\dagger = \sum_{j=1}^\infty \varphi(\sigma_j^2)(h, v_j)v_j =: \sum_{j=1}^\infty f_j v_j$, we obtain

$$\begin{aligned}
&(f, v_{j,m})\\
= &\sum_{l=1}^\infty f_l (v_l, v_{j,m}) = \frac{\sqrt{m+1}}{\sigma_{j,m}} \left(\sigma_j f_j - \sum_{t=1}^\infty \left(\sigma_{2t(m+1)-j} f_{2t(m+1)-j} - \sigma_{2t(m+1)+j} f_{2t(m+1)+j} \right) \right)\\
= &\varphi_s(\sigma_{j,m}^2)\sqrt{m+1} \frac{\sigma_j \varphi_s(\sigma_j^2)}{\sigma_{j,m}\varphi_s(\sigma_{j,m}^2)} * \Bigg((h, v_j)\\
&- \sum_{t=1}^\infty \frac{\sigma_{2t(m+1)-j}\varphi_s(\sigma_{2t(m+1)-j})(h, v_{2t(m+1)-j}) - \sigma_{2t(m+1)+j}\varphi_s(\sigma_{2t(m+1)+j})(h, v_{2t(m+1)+j})}{\sigma_j \varphi_s(\sigma_j^2)} \Bigg).
\end{aligned}$$

Using the Cauchy-Schwartz-inequality gives

$$\begin{aligned}
&\Bigg((h, v_j)\\
&- \sum_{t=1}^\infty \frac{\sigma_{2t(m+1)-j}\varphi_s(\sigma_{2t(m+1)-j})(h, v_{2t(m+1)-j}) - \sigma_{2t(m+1)+j}\varphi_s(\sigma_{2t(m+1)+j})(h, v_{2t(m+1)+j})}{\sigma_j \varphi_s(\sigma_j^2)} \Bigg)^2\\
\leq &\ 2(h, v_j)^2\\
&+ 2\left(\sum_{t=1}^\infty \left(\frac{\sigma_{2t(m+1)-j}^2}{\sigma_j^2} \right)^{\frac{s+1}{2}} |(h, v_{2t(m+1)-j})| + \left(\frac{\sigma_{2t(m+1)+j}^2}{\sigma_j^2} \right)^{\frac{s+1}{2}} |(h, v_{2t(m+1)+j})| \right)^2.
\end{aligned}$$

For the second term, we further obtain

$$\begin{aligned}
&\left(\sum_{t=1}^\infty \left(\frac{\sigma_{2t(m+1)-j}^2}{\sigma_j^2} \right)^{\frac{s+1}{2}} |(h, v_{2t(m+1)-j})| + \left(\frac{\sigma_{2t(m+1)+j}^2}{\sigma_j^2} \right)^{\frac{s+1}{2}} |(h, v_{2t(m+1)+j})| \right)^2\\
= &\left(\sum_{t=1}^\infty \left(\frac{1}{2t\frac{m+1}{j} - 1} \right)^{2s+2} |(h, v_{2t(m+1)-j})| + \left(\frac{1}{2t\frac{m+1}{j} + 1} \right)^{2s+2} |(h, v_{2t(m+1)+j})| \right)^2\\
\leq &\left(\sum_{t=1}^\infty \left(\frac{1}{2t\frac{m+1}{j} - 1} \right)^{4s+4} + \left(\frac{1}{2t\frac{m+1}{j} + 1} \right)^{4s+4} \right) \left(\sum_{t=1}^\infty (h, v_{2t(m+1)-j})^2 + (h, v_{2t(m+1)+j})^2 \right)\\
\leq &2^{-3-4s} \left(\sum_{t=1}^\infty t^{-4s-4} \right) \left(\sum_{t=1}^\infty (h, v_{2t(m+1)-j})^2 + (h, v_{2t(m+1)+j})^2 \right)\\
\leq &\frac{1}{2^{4s}(4s+3)} \left(\sum_{t=1}^\infty (h, v_{2t(m+1)-j})^2 + (h, v_{2t(m+1)+j})^2 \right)
\end{aligned}$$

and finally

$$\left((h,v_j) - \sum_{t=1}^{\infty} \frac{\sigma_{2t(m+1)-j}\varphi_s(\sigma_{2t(m+1)-j})(h,v_{2t(m+1)-j}) - \sigma_{2t(m+1)+j}\varphi_s(\sigma_{2t(m+1)+j})(h,v_{2t(m+1)+j})}{\sigma_j\varphi_s(\sigma_j^2)}\right)^2$$

$$\leq 2\left((h,v_j)^2 + \sum_{t=1}^{\infty}(h,v_{2t(m+1)-j})^2 + (h,v_{2t(m+1)+j})^2\right).$$

Moreover, we use $\sin^2(x) \in [0,1]$ and $\sin(x) \leq x$ and obtain

$$(m+1)\frac{\sigma_j^2\varphi_s^2(\sigma_j^2)}{\sigma_{j,m}^2\varphi_s^2(\sigma_{j,m}^2)} = (m+1)\left(\frac{\sigma_j^2}{\sigma_{j,m}^2}\right)^{s+1} = (m+1)\left(\frac{16(m+1)^3\sin^4\left(\frac{j\pi}{2(m+1)}\right)}{\pi^4 j^4\left(1-\frac{2}{3}\sin^2\left(\frac{j\pi}{2(m+1)}\right)\right)}\right)^{s+1}$$

$$\leq (m+1)\left(\frac{3}{(m+1)}\right)^{s+1} = \frac{3^{s+1}}{(m+1)^s}.$$

Putting both estimates together yields

$$\sum_{j=k+1}^{m}(f,v_{j,m})^2$$

$$\leq\; 2*3^{s+1}\frac{\varphi_s^2\left(\sigma_{k+1,m}^2\right)}{(m+1)^{2s}}\sum_{j=k+1}^{m}\left((h,v_j)^2 + \sum_{t=1}^{\infty}(h,v_{2t(m+1)-j})^2 + (h,v_{2t(m+1)+j})^2\right)$$

$$\leq \frac{2*3^{s+1}}{(m+1)^s}\varphi_s^2\left(\frac{1-\frac{2}{3}\sin^2\left(\frac{(k+1)\pi}{2(m+1)}\right)}{16(m+1)^3\sin^4\left(\frac{(k+1)\pi}{2(m+1)}\right)}\right)\sum_{l=k+1}^{\infty}(h,v_l)^2$$

$$\leq \frac{2*3^{s+1}}{(m+1)^s}\varphi_s^2\left(\frac{m+1}{2^4(k+1)^4}\right)\rho^2 = \frac{3^{s+1}}{2^{4s-1}}k^{-4s}\rho^2, \tag{3.6}$$

where we used that $\sin(x) \geq \frac{2}{\pi}x$ for $0 \leq x \leq \frac{\pi}{2}$ in the third step and the fact that for every $l \geq m+1$ there is at most one pair (j,t) such that $l = 2t(m+1)-j$ or $l = 2t(m+1)+j$ in the second step. Therefore, on the one hand,

$$\sup_{f^\dagger\in\mathcal{X}_{s,\rho}}\sum_{j=k+1}^{\infty}(f^\dagger,v_{j,m})^2 \leq \frac{3^{s+1}}{2^{4s-1}}k^{-4s}\rho^2,$$

while on the other hand

$$\frac{k\delta^2}{\sigma_{k,m}^2} = \frac{16(m+1)^3\sin^4\left(\frac{k\pi}{2(m+1)}\right)}{1-\frac{2}{3}\sin^2\left(\frac{k\pi}{2(m+1)}\right)}k\delta^2 \leq \frac{16\pi^4k^4}{\frac{1}{3}2^4(m+1)}k\delta^2 = 3\pi^4\frac{k^5\delta^2}{m+1}.$$

Consequently,

$$3\pi^4 \frac{k\delta^2}{m+1} \overset{!}{\leq} \frac{3^{s+1}}{2^{4s-1}} k^{-4s}\rho^2$$
$$\implies \quad k \leq C_s \left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{1}{5+4s}}$$

with

$$C_s := \left(\frac{3^s}{2^{4s-1}\pi^4}\right)^{\frac{1}{5+4s}}. \tag{3.7}$$

We conclude

$$\sup_{f^\dagger \in \mathcal{X}_{s,\rho}} s_m^\delta(f^\dagger) \leq C_s \left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{1}{5+4s}}.$$

With similar arguments we also get

$$\sup_{f^\dagger \in \mathcal{X}_{s,\rho}} t_m^\delta(f^\dagger) \leq C_s \left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{1}{5+4s}}.$$

For $t_m^\delta \geq t$ we therefore obtain, with (3.6),

$$\|f^\delta_{k^\delta_{\mathrm{gcv},m},m} - P_{\mathcal{N}(K_m)^\perp} f^\dagger\| \chi\Omega_t$$
$$\leq \frac{L_s\sqrt{s_m^\delta}\delta + C_a\sqrt{\sum_{j=\frac{m}{2}+1}^m (g_m^\dagger, u_{j,m})^2_{\mathbb{R}^m}}}{\sigma_{\frac{s_m^\delta}{\varepsilon^2},m}} \tag{3.8}$$
$$\leq \frac{\sqrt{3}L_s\pi^2}{\varepsilon^4} s_m^{\delta\,\frac{5}{2}} \frac{\delta}{\sqrt{m+1}} + C_a \frac{\sigma_{\frac{m}{2}+1,m}}{\sigma_{\frac{s_m^\delta}{\varepsilon^2},m}} \sqrt{\sum_{j=\frac{m}{2}+1}^m (f^\dagger, v_{j,m})^2} \tag{3.9}$$
$$\leq \frac{\sqrt{3}C_s^{\frac{5}{2}} L_s\pi^2}{\varepsilon^4} \left(\frac{(m+1)\rho^2}{\delta^2}\right)^{\frac{5}{2(5+4s)}} \frac{\delta}{\sqrt{m+1}} + 3\sqrt{2}C_a \sqrt{\sum_{j=\frac{m}{2}+1}^m (f^\dagger, v_{j,m})^2}$$
$$= L_s' \left(\frac{\delta}{\sqrt{m+1}}\right)^{\frac{4s}{5+8s}} \rho^{\frac{5}{5+4s}} + L_s'' \frac{\rho}{m^{2s}}, \tag{3.10}$$

with

$$L_s := \frac{\sqrt{3}C_s^{\frac{5}{2}} L_s\pi^2}{\varepsilon^4} \quad \text{and} \quad L_s'' := \frac{3^{s+2}}{2^{4s-\frac{3}{2}}} C_a. \tag{3.11}$$

Finally, we treat the discretization error $\|P_{\mathcal{N}^\perp(K_m)} f^\dagger - f^\dagger\|$. First, by definition of κ we see that the span $< v_{1,m}, ..., v_{m,m} >$ is equal to the space of piece-wise linear functions on the grid $\xi_{1,m},...,\xi_{m,m}$, and $f_m^\dagger = P_{\mathcal{N}(K_m)^\perp} f^\dagger$ is the L^2-projection of $f^\dagger$ onto that space. The error depends on classical smoothness of $f^\dagger$ and we now relate the Hölder source condition to classical smoothness.

Proposition 3.9. *Assume that $f^\dagger \in \mathcal{X}_{s,\rho}$. If $s > \frac{3}{4}$, then $f^\dagger$ is differentiable. if $s > \frac{5}{4}$, then $f^\dagger$ is twice differentiable.*

Proof of Proposition 3.9. First $f^\dagger \in \mathcal{X}_{s,\rho}$ implies that there exists $h \in L^2$ with $\|h\| \leq \rho$, such that $f^\dagger = \sum_{j=1}^\infty \varphi_s(\sigma_j^2)(h,v_j)v_j$. Differentiating the sum formally term-by-term, we obtain

$$\sqrt{2}\sum_{j=1}^\infty \pi j \varphi_s(\sigma_j^2)(h,v_j)\cos\left(\pi j \cdot\right).$$

We now show that this series converges uniformly in x. Indeed, using Cauchy-Schwartz,

$$\sum_{j=1}^\infty \pi j \varphi_s(\sigma_j^2)|(h,v_j)||\cos(j\pi x)| \leq \pi \sqrt{\sum_{j=1}^\infty (h,v_j)^2}\sqrt{\sum_{j=1}^\infty j^2\varphi_s^2(\sigma_j^2)} \leq \pi^{1+2s}\rho\sqrt{\sum_{j=1}^\infty j^{2-4s}},$$

and the right hand side converges whenever $s > \frac{3}{4}$, uniformly in x. Consequently, it holds that

$$(f^\dagger)' = \sqrt{2}\sum_{j=1}^\infty j\pi\varphi_s(\sigma_j^2)(h,v_j)\cos(\pi j\cdot).$$

Similar, differentiating $f^\dagger$ twice formally term-by-term, we get

$$-\sqrt{2}\sum_{j=1}^\infty j^2\pi^2\varphi_s(\sigma_j^2)(h,v_j)v_j(\cdot),$$

and

$$\sum_{j=1}^\infty \pi^2 j^2\varphi_s(\sigma_j^2)|(h,v_j)||v_j(x)| \leq \pi^2\sqrt{\sum_{j=1}^\infty (h,v_j)^2}\sqrt{\sum_{j=1}^\infty j^4\varphi_s^2(\sigma_j^2)} \leq \pi^{2+2s}\rho\sqrt{\sum_{j=1}^\infty j^{4-4s}},$$

where the right hand side converges uniformly in x whenever $s > \frac{5}{4}$. ■

Proposition 3.9 and classical estimates for the linear interpolating spline then yield the following bound for the discretization error,

$$\|P_{\mathcal{N}(K_m)^\perp}f^\dagger - f^\dagger\|_{L^2} \leq \begin{cases} \frac{\|(f^\dagger)'\|_{L^2}}{\sqrt{2}(m+1)}, & \text{for } s \geq \frac{3}{4} \\ \frac{\|(f^\dagger)''\|_{L^2}}{2(m+1)^2}, & \text{for } s \geq \frac{5}{4} \end{cases}. \tag{3.12}$$

Finally, plugging the estimates (3.10) and (3.12) into the decomposition

$$\|f^\delta_{k^\delta_{\mathrm{gcv}},m} - f^\dagger\|_{\chi\Omega_t} \leq \|f^\delta_{k^\delta_{\mathrm{gcv}},m} - P_{\mathcal{N}(K_m)^\perp}f^\dagger\|_{\chi\Omega_t} + \|P_{\mathcal{N}(K_m)^\perp}f_m^\dagger - f^\dagger\|_{\chi\Omega_t}$$

and applying Lemma 3.1 and Lemma 3.7 finishes the proof of Theorem 2.2.

4. Numerical experiments. We now implement GCV and apply it to the integral equation (2.1). First, we set $D = 2^{14} = 16384$ and fix, for all simulations, X_j i.i.d. standard Gaussian random variables, $j = 1, ..., D$. Based on this we define three exact solutions

$$f^{i,\dagger} := \sum_{j=1}^{D} \sigma_j^{s_i} X_j v_j$$

with $s_i \in \left\{\frac{1}{4}, \frac{3}{4}, \frac{5}{4}\right\}$ varying the smoothness of the solution. We define the corresponding exact data as

$$g_m^{i,\dagger} := \left(Kf^{i,\dagger}(\xi_{l,m})\right)_{l=1}^{m} = \sqrt{2}\left(\sum_{j=1}^{D} (j\pi)^{2(s_i+1)} X_j \sin(j\pi\xi_{l,m})\right)_{l=1}^{m} \in \mathbb{R}^m.$$

We generate the perturbed data

$$g_m^{i,\delta} := g_m^{i,\dagger} + \delta \begin{pmatrix} Z_1 \\ ... \\ Z_m \end{pmatrix}, \tag{4.1}$$

with $Z_1, ..., Z_m$ i.i.d. standard Gaussian, sampled anew in every simulation loop. We first give formulas to calculate the error of our estimator. Using Proposition 3.8, the projection $(f^{i,\dagger}, v_{k,m}) = \sum_{j=1}^{D} \sigma_j^{s_i+1} X_j (v_j, v_{k,m})$ can be calculated exactly for $k = 1, ..., m$, and we define $f_m^{i,\dagger} := \sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m}) v_{j,m}$. We have

$$\|f_{k,m}^{\delta} - f_m^{i,\dagger}\|^2 = \sum_{j=1}^{k} \left(\frac{(g_m^{i,\delta}, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} - (f^{i,\dagger}, v_{j,m})\right)^2 + \sum_{j=k+1}^{m} (f^{i,\dagger}, v_{j,m})^2$$

and

$$\begin{aligned} f_m^{i,\dagger} - f^{i,\dagger} &= \sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m}) v_{j,m} - \sum_{l=1}^{D} (f^{i,\dagger}, v_l) v_l \\ &= \sum_{l=1}^{D} \left(\sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m})(v_{j,m}, v_l) - (f^{i,\dagger}, v_l)\right) v_l + \sum_{l=D+1}^{\infty} \sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m})(v_{j,m}, v_l) v_l. \end{aligned}$$

Thus, by orthogonality ($\|f_k^{i,\delta} - f^{i,\dagger}\|^2 = \|f_k^{i,\delta} - f_m^{i,\dagger}\|^2 + \|f_m^{i,\dagger} - f^{i,\dagger}\|^2$),

$$\begin{aligned} &\|f_{k,m}^{\delta} - f^{i,\dagger}\|^2 \\ &= \sum_{j=1}^{k} \left(\frac{(g_m^{i,\delta}, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} - (f^{i,\dagger}, v_{j,m})\right)^2 + \sum_{j=k+1}^{m} (f^{i,\dagger}, v_{j,m})^2 \\ &\quad + \sum_{j=1}^{D} \left(\sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m})(v_{j,m}, v_l) - (f^{i,\dagger}, v_l)\right)^2 + \sum_{l=D+1}^{\infty} \left(\sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m})(v_{j,m}, v_l)\right)^2 \end{aligned}$$

and we define, suppressing the dependence on δ and m, i, the approximative error of the estimator:

$$e_k := \left(\sum_{j=1}^{k} \left(\frac{(g_m^{i,\delta}, u_{j,m})_{\mathbb{R}^m}}{\sigma_{j,m}} - (f^{i,\dagger}, v_{j,m}) \right)^2 + \sum_{j=k+1}^{m} (f^{i,\dagger}, v_{j,m})^2 \right. \tag{4.2}$$

$$\left. + \sum_{j=1}^{D} \left(\sum_{j=1}^{m} (f^{i,\dagger}, v_{j,m})(v_{j,m}, v_l) - (f^{i,\dagger}, v_l) \right)^2 \right)^{\frac{1}{2}}. \tag{4.3}$$

In the simulations we calculate the computable GCV estimator

$$k_{\mathrm{gcv}} := \arg \min_{0 \le k \le \frac{m}{2}} \frac{\sum_{j=k+1}^{m} (g_m^{i,\delta}, u_{j,m})_{\mathbb{R}^m}}{(1 - \frac{k}{m})^2}, \tag{4.4}$$

and the in practice unfeasible optimal estimator

$$k_{\mathrm{opt}} := \arg \min_{0 \le k \le m} e_k, \tag{4.5}$$

for reference. The error we make in approximating $\|f_{k,m}^{\delta} - f^{\dagger}\|$ by (4.2) can be bounded from above as follows (where expectation is with respect to the $X_j' s$):

$$\begin{aligned}
&\mathbb{E}\left[\left| e_k^2 - \|f_k^{i,\delta} - f^{i,\dagger}\|^2 \right| \right] \\
&= \sum_{l=D+1}^{\infty} \mathbb{E}\left[\left(\sum_{j=1}^{m} \sigma_j^{s_i} X_j (v_{j,m}, v_l) \right)^2 \right] = \sum_{l=D+1}^{\infty} \sum_{j=1}^{m} \sigma_j^{2s_i} (v_{j,m}, v_l)^2 \\
&\le \sum_{l=D+1}^{\infty} \max_{j=1,\dots,m} \sigma_j^{2s_i} (m+1) \frac{\sigma_l^2}{\sigma_{j,m}^2} \le 3 \max_{j=1,\dots,m} \sigma_j^{2s_i - 2} \sum_{l=D+1}^{\infty} \sigma_l^2 \le \frac{3}{\pi^4} \frac{1}{D^3} \max_{j=1,\dots,m} \sigma_j^{2s_i - 2}
\end{aligned}$$

and so

$$\delta_i^2 := \frac{3}{\pi^4} \begin{cases} \frac{(m\pi)^3}{D^3} & , \quad \text{for } s = \frac{1}{4} \\ \frac{m\pi}{D^3} & , \quad \text{for } s_i = \frac{3}{4} \\ \frac{1}{\pi D^3} & , \quad \text{for } s_i = \frac{5}{4} \end{cases}.$$

is an upper bound for $\mathbb{E}\left[\left| e_k^2 - \|f_k^{i,\delta} - f^{i,\dagger}\|^2 \right| \right]$. For our choices of m and D we thus obtain

$$\delta_i \asymp \begin{cases} 2^{-9} & , \quad \text{for } s_i = \frac{1}{4} \\ 2^{-17} & , \quad \text{for } s_i = \frac{3}{4} \\ 2^{-21} & , \quad \text{for } s_i = \frac{5}{4} \end{cases}.$$

We will see below in the error plots that δ_i is of smaller order than e_k in all cases. We consider different noise levels δ, which we determine implicitly via the signal-to-noise ratio (SNR). The

SNR is defined as

$$\mathrm{SNR} := \frac{\|\mathrm{signal}\|}{\|\mathrm{noise}\|} = \frac{\|g^{i,\dagger}\|_m}{\sqrt{m}\delta}.$$

For each exact solution $f^{i,\dagger}$ and each SNR, we generate 200 independent noisy measurements g_m^δ (in (4.1)), and calculate $k.$ along with the corresponding errors $e_{k.}$, where $\cdot \in \{\mathrm{gcv}, \mathrm{opt}\}$, see (4.2) − (4.5). We fix the number of measurements as $m = 2^9$ and let SNR vary over $\{1, 10, ..., 10^8\}$ (that is we effectively vary the noise level δ). The results are presented in Figure 1. In the left column we visualize the statistics as box plots and in the right column we give the corresponding sample means and sample standard deviations in tabular form. In each box plot, the upper and lower edge give the 75- respective 25% quantile of the statistic $e_{k.}$ for $\cdot = \mathrm{gcv}$ (red) and $\cdot = \mathrm{opt}$ (blue). The median of the statistic is given as a red bar inside the boxes. The whiskers extend to the samples whose distance to the upper respectively lower edge is less than six times the height of the box. All samples which fall outside of the whiskers are plotted individually as red crosses (outliers). Outliers above the upper limit 1 are plotted just above, retaining their relative order, but not given the exact value.

We clearly observe the convergence of the error, as the noise level decreases (that is as the SNR increases). Hereby, the convergence rate of the generalized cross-validation is comparable to the one of the optimal rate at least for small noise levels. For larger noise levels (smaller SNR) the statistic for the generalized cross-validation is rather spread out. Moreover we observe saturation of the error for rougher solutions with smoothness parameter $s_i \in \{1/4, 3/4\}$, due to a dominating discretization error. The difference between $e_{k^\delta_{\mathrm{gcv}}}$ and $e_{k^\delta_{\mathrm{opt}}}$ in the saturation regime is due to the constraint $k^\delta_{\mathrm{gcv}} \leq \frac{m}{2}$. Note that in all cases the error for the largest SNR is still of higher order than the errors δ_i we make in the approximation.

5. Concluding remarks.

In this article we deduced rigorously a non-asymptotic error bound (in probability) for GCV as a parameter choice rule for the solution of a specific ill-posed integral equation. In particular we verified the optimality of the rule in the mini-max sense, remarkably without imposing a self-similarity condition onto the unknown solution, which up to our knowledge so far was required for any rigorous and consistent optimality result for heuristic parameter choice rules in the context of ill-posed problems. We conclude with listing three possible further research directions. First, the findings could be extended to integral equations with a general kernel κ. As mentioned above, see e.g. Corollary 3.6, the probabilistic analysis of the rule remains largely unchanged. However, it remains to analyze the discretization error given by the relation between the decomposition of the continuous operator K and the semi-discrete one K_m. In particular, the design matrix T_m cannot be calculated exactly in this case and has to be approximated by, e.g., a quadrature rule, and the estimator should be based on the decomposition of the quadrature approximation. Second, instead of spectral cut-off other regularization methods, like Tikhonov regularization or some iterative scheme should be considered. This will require non-trivial changes of the probabilistic analysis of GCV. Finally, it would be interesting to extend the analysis to more contemporary settings, for example non-parametric regression based on kernelized spectral-filter algorithms.

Appendix A. Proof of Lemma 2.1.

Proof of Lemma 2.1. It is well-known that in our setting the kernel is the Green's function of the Laplace equation, i.e., $(Kf)'' = -f$. It is then straight forward to check that the solutions of the differential equation are eigenfunctions of K, which yield σ_j and v_j. While the discretization of the differential equation has been analyzed in detail, see, e.g., [4], we have not found results for the corresponding discretization of the integral equation in the literature. We first show that the singular value decomposition of the semi-discrete K_m is strongly related to the eigenvalue decomposition of the symmetric $m \times m$ matrix $(T_m)_{ij} := \int \kappa(\xi_{i,m}, y)\kappa(\xi_{j,m}, y)\mathrm{d}y = \frac{\xi_i(1-\xi_l)}{6}(-\xi_i^2 - \xi_j^2 + 2\xi_j)$. Indeed, since $K_m^*\alpha = \sum_{j=1}^m \alpha_j \kappa(\xi_{j,m}, \cdot)$ for $\alpha \in \mathbb{R}^m$, we obtain for $f_\alpha := \sum_{j=1}^m \alpha_j\kappa(\xi_{j,m}, \cdot) \in L^2$ and $\lambda \in \mathbb{R}$ the relation

$$K_m^* K_m f_\alpha = \lambda f_\alpha \qquad \Longleftrightarrow \qquad T_m \alpha = \lambda\alpha.$$

and consequently we need to find the eigenvalue decomposition of T_m. As auxiliary tools, we need the following $m \times m$-dimensional symmetric matrices:

(A.1)
$$\Delta_m := \begin{pmatrix} 2 & -1 & \dots & & \\ -1 & 2 & -1 & \dots & \\ \vdots & & & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}, \quad R_m := \begin{pmatrix} 4 & 1 & \dots & & \\ 1 & 4 & 1 & \dots & \\ \vdots & & & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{pmatrix}, \quad S_m := \big(\kappa(\xi_{s,m}, \xi_{t,m})\big)_{st}$$

Note that $(m+1)^2\Delta_m$ is the discretization of the second derivative via centered second order finite differences on the homogeneous grid $\xi_{1,m}, ..., \xi_{m,m}$ and $(R_m)_{ij} = \frac{6}{m+1}(\Lambda_i^m, \Lambda_j^m)_{L^2(0,1)}$, with the hat functions $\Lambda_i(x) := (x - \xi_{i-1,m})(m+1)\chi_{(\xi_{i-1,m}, \xi_{i,m}]}(x) + (\xi_{i+1,m} - x)(m+1)\chi_{(\xi_{i,m}, \xi_{i+1,m}]}(x)$. First we show that T_m and the matrices in (A.1) have mutual eigenvectors

$$z_{k,m} := \sqrt{\frac{2}{m+1}}\left(\sin\left(\sqrt{\lambda_k}\xi_{1m}\right) \quad \dots \quad \sin\left(\sqrt{\lambda_k}\xi_{mm}\right)\right)^T \in \mathbb{R}^m, \tag{A.2}$$

with $k = 1, ..., m$. Using the polar identity $2i\sin(x) = e^{ix} - e^{-ix}$ and the closed-form expression for the partial geometric series with $q = e^{\frac{ik\pi}{m+1}}$, one sees that $\|z_{k,m}\|_{\mathbb{R}^m}^2 = 1$. By exploiting the polar identity again one easily verifies that $z_{k,m}$ are the eigenvectors of the circulant matrices Δ_m and R_m, and moreover that $\rho_{k,m} := 4 + 2\cos\left(\frac{k\pi}{m+1}\right)$ are the corresponding eigenvalues for R_m. Moreover, slightly lengthy but straightforward computations yield $\Delta_m S_m - S_m\Delta_m = 0 = \Delta_m T_m - T_m\Delta_m$, which implies that the $z_{k,m}$ are also the eigenvectors of S_m and T_m. Next we show that the eigenvalues $\mu_{k,m}$ of S_m are given by $\mu_{k,m} = (-1)^{k+1}\cos\left(\frac{\sqrt{\lambda_k}}{2(1+m)}\right)\sin\left(\frac{\sqrt{\lambda_k}}{2(1+m)}\right)^{-1}\sin\left(\frac{\sqrt{\lambda_k}}{1+m}\right)^{-1}$. Using the polar identity for $q = e^{\frac{ik\pi}{2(m+1)}}$ and

$$\sum_{j=1}^m q^j j = \frac{q + q^{1+m}(-1 - m + mq)}{(1-q)^2}$$

yields

$$\sum_{l=1}^{m} \sin\left(\frac{\sqrt{\lambda_k} l}{m+1}\right) l = \frac{m+1}{2} (-1)^{k+1} \frac{\cos\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)}{\sin\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)},$$

and because $\sin(k\pi m/(m+1)) = \sin(k\pi/(m+1))$, the $\mu_{k,m}$ can be computed with the defining relation of the eigenvalues:

$$\mu_{k,m} \sin\left(\frac{\sqrt{\lambda_k} m}{m+1}\right) = \sqrt{\frac{2}{m+1}} \mu_{k,m} (z_{k,m})_m = \sqrt{\frac{2}{m+1}} (S_m z_{k,m})_m \tag{A.3}$$

$$= \sum_{l=1}^{m} \xi_{l,m} (1 - \xi_{m,m}) \sin\left(\frac{\sqrt{\lambda_k} l}{m+1}\right) = \frac{(-1)^{k+1}}{2(m+1)} \frac{\cos\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)}{\sin\left(\frac{\sqrt{\lambda_k}}{2(m+1)}\right)}. \tag{A.4}$$

To finally determine the eigenvalues of $\sigma_{k,m}^2$ of T_m we set $w_{k,m} := \sum_{l=1}^{m} (z_{k,m})_l \kappa(\xi_{l,m}, \cdot)$ and normalize in two ways. First,

$$\|w_{k,m}\|^2 = \sum_{l,l'=1}^{m} (z_{k,m})_l (z_{k,m})_{l'} (\kappa(\xi_{l,m}, \cdot), \kappa(\xi_{l',m}, \cdot)) = z_{k,m}^T T_m z_{k,m} = \sigma_{k,m}^2.$$

Second, expanding $\kappa(\xi_{j,m}, \cdot) = \sum_{i=1}^{m} \kappa(\xi_{l,m}, \xi_{i,m}) \Lambda_i(\cdot)$ in terms of the hat functions,

$$\|w_{k,m}\|^2$$

$$= \|\sum_{l=1}^{m} (z_{k,m})_l \sum_{i=1}^{m} \kappa(\xi_{l,m}, \xi_{i,m}) \Lambda_i\|^2 = \|\sum_{i=1}^{m} (S_m z_{k,m})_i \Lambda_i^m\|^2 = \mu_{k,m}^2 \|\sum_{i=1}^{m} (z_{k,m})_i \Lambda_i^m\|^2 \tag{A.5}$$

$$= \mu_{k,m}^2 \sum_{i,i'=1}^{m} (z_{k,m})_i (z_{k,m})_{i'} (\Lambda_i^m, \Lambda_{i'}^m) = \mu_{k,m}^2 \frac{1}{6(m+1)} \sum_{i=1}^{m} (z_{k,m})_i (R_m z_{k,m})_i$$

$$= \mu_{k,m}^2 \frac{4 + 2\cos\left(\frac{\sqrt{\lambda_k}}{m+1}\right)}{6(m+1)}. \tag{A.6}$$

Putting (A.3) and (A.6) together, using $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = 1 - 2\sin^2(x)$, then yields the explicit formulas for the eigenvalues $\sigma_{k,m}$ and the left singular functions $v_{k,m}$. Finally, we calculate the right singular vectors $u_{k,m}$:

$$(u_{k,m})_j = \frac{1}{\sigma_{k,m}} (K_m v_{k,m})(\xi_{j,m}) = \frac{1}{\sigma_{k,m}} \sum_{l=1}^{m} (z_{k,m})_l (K_m \kappa(\xi_{l,m}, \cdot))(\xi_{j,m})$$

$$= \frac{1}{\sigma_{k,m}} \sum_{l=1}^{m} (T_m)_{j,l} (z_{k,m})_l = (z_{k,m})_j = \sqrt{\frac{2}{m+1}} \sin(k\pi \xi_{j,m}). \qquad \blacksquare$$

Appendix B. Proof of Proposition 3.8 .

Proof of Proposition 3.8. We need the following auxiliary identity: For $m \in \mathbb{N}, t \in \mathbb{N}_0$ and $k \in \{1, ..., m\}, s \in \{0, ..., m\}$ and $j = t(m+1) + s$ there holds

$$\text{(B.1)} \qquad \sum_{l=1}^{m} \sin\left(\frac{j\pi l}{m+1}\right) \sin\left(\frac{k\pi l}{m+1}\right) = \begin{cases} \frac{m+1}{2} & \text{for } s = k \text{ and } t \text{ even} \\ -\frac{m+1}{2} & \text{for } s + k = m+1 \text{ and } t \text{ odd} \\ 0 & \text{else} \end{cases}.$$

We first prove the claim. With $q_1 = \exp(i(j+k)\pi/(m+1))$ and $q_2 = \exp(i(j-k)\pi/(m+1))$ and the polar identity we obtain

$$\sum_{l=1}^{m} \sin\left(\frac{j\pi l}{m+1}\right) \sin\left(\frac{k\pi l}{m+1}\right) = \frac{1}{4} \sum_{l=1}^{m} \left(q_2^l + q_2^{-l} - (q_1^l + q_1^{-l})\right).$$

For $q \in \{q_1, q_2\}$, if $q \neq 0, 1$, if holds that

$$\sum_{i=1}^{m} (q^i + q^{-i}) = -1 + \frac{q^{m+\frac{1}{2}} - q^{-(m+\frac{1}{2})}}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}} = -1 + (-1)^{k+j}(-1) = -(1 + (-1)^{k+j})$$

since $q^{m+\frac{1}{2}} = (-1)^{k+j} q^{-\frac{1}{2}}$. If t is even and $s = k$, then $j - k = t(m+1)$ which implies that $q_2 = 1$, while, since $0 < 2k < 2(m+1)$, the sum $j + k = t(m+1) + 2k$ cannot be a multiple of $2(m+1)$, therefore $q_1 \neq 0, 1$ and thus, since $j+k$ is even, $\sum_{l=1}^m \sin\left(\frac{j\pi l}{m+1}\right) = \frac{m+1}{2}$. Similar, if t is odd and $s + k = m+1$, then $j + k = (t+1)(m+1)$ implies $q_1 = 1$, and now $j - k = t(m+1) + s - k = (t+1)(m+1) - 2k$ is not a multiple of $2(m+1)$, which yields $q_2 \neq 0, 1$. Since $j+k$ is again even we deduce $\sum_{l=1}^m \sin\left(\frac{j\pi l}{m+1}\right) \sin\left(\frac{k\pi l}{m+1}\right) = -\frac{m+1}{2}$. In any other case it hold that $q_1, q_2 \neq 0, 1$ and therefore $\sum_{l=1}^m \sin\left(\frac{j\pi l}{m+1}\right) \sin\left(\frac{k\pi l}{m+1}\right) = 0$, which finishes the proof of the claim (B.1). We come to the proof of the proposition. As above we can write $j = t(m+1) + s$ with $t \in \mathbb{N}_0$ and $s \in \{0, ..., m\}$. Using the claim (B.1) together with

$$\sigma_{j,m} \frac{m+1}{2} (v_k, v_{j,m}) = \left(\sin(\sqrt{\lambda_k}\cdot), \sum_{l=1}^{m} \sin\left(\sqrt{\lambda_j}\xi_l\right) \kappa(\xi_{l,m}, \cdot)\right)$$

$$= \sum_{l=1}^{m} \sin\left(\sqrt{\lambda_j}\xi_l\right) \left(\sin(\sqrt{\lambda_k}\cdot), \kappa(\xi_{l,m}, \cdot)\right) = \sigma_k \sum_{l=1}^{m} \sin\left(\sqrt{\lambda_j}\xi_l\right) \sin\left(\sqrt{\lambda_k}\xi_l\right)$$

concludes the proof. ∎

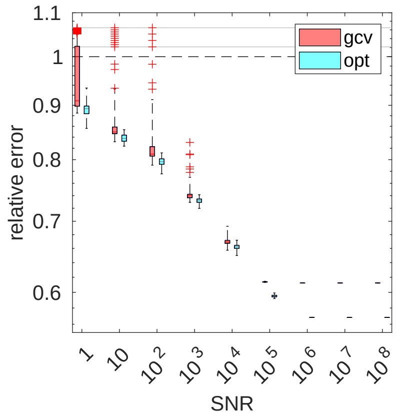
REFERENCES

[1] S. Arlot and A. Celisse, *A survey of cross-validation procedures for model selection*, Statist. Surv., 4 (2010), pp. 40–79.

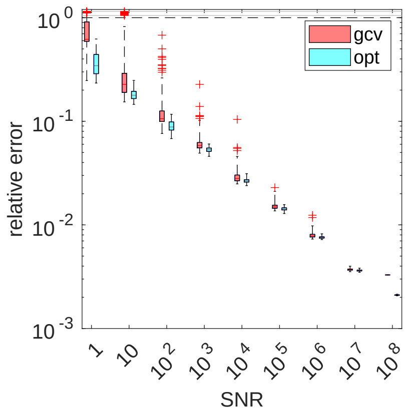
[2] A. Bakushinskii, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zh. Vychisl. Mat. i Mat. Fiz., 24 (1984), pp. 1258–1259.

[3] F. Bauer and M. Reiss, *Regularization independent of the noise level: an analysis of quasi-optimality*, Inverse Problems, 24 (2008), p. 055009.

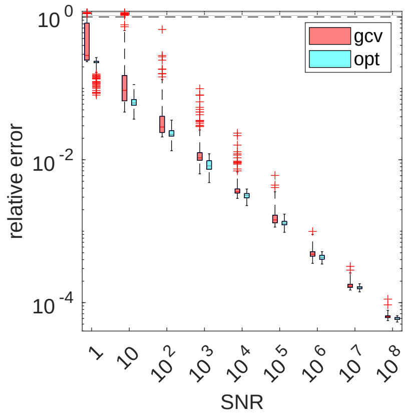
[4] D. Boffi, *Finite element approximation of eigenvalue problems*, Acta Numerica, 19 (2010), pp. 1–120.
[5] A. Caponnetto and Y. Yao, *Cross-validation based adaptation for regularization operators in learning theory*, Analysis and Applications, 8 (2010), pp. 161–183.
[6] P. Craven and G. Wahba, *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation*, Numerische mathematik, 31 (1978), pp. 377–403.
[7] C. Giraud, *Introduction to high-dimensional statistics*, CRC Press, 2021.
[8] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer, 2009.
[9] 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 (2022), pp. 35–56.
[10] T. Jahn, *Noise level free regularization of general linear inverse problems under unconstrained white noise*, SIAM/ASA Journal on Uncertainty Quantification, 11 (2023), pp. 591–615.
[11] M. Kuhn, K. Johnson, et al., *Applied predictive modeling*, vol. 26, Springer, 2013.
[12] K.-C. Li, *Asymptotic optimality of cl and generalized cross-validation in ridge regression with application to spline smoothing*, The Annals of Statistics, (1986), pp. 1101–1112.
[13] K.-C. Li, *Asymptotic optimality for cp, cl, cross-validation and generalized cross-validation: discrete index set*, The Annals of Statistics, (1987), pp. 958–975.
[14] F. Lucka, K. Proksch, C. Brune, N. Bissantz, M. Burger, H. Dette, and F. Wübbeling, *Risk estimators for choosing regularization parameters in ill-posed problems-properties and limitations*, Inverse Problems & Imaging, 12 (2018), pp. 1121–1155.
[15] M. A. Lukas, *Asymptotic optimality of generalized cross-validation for choosing the regularization parameter*, Numerische Mathematik, 66 (1993), pp. 41–66.
[16] M. A. Lukas, *Robust generalized cross-validation for choosing the regularization parameter*, Inverse problems, 22 (2006), p. 1883.
[17] M. A. Lukas, *Strong robust generalized cross-validation for choosing the regularization parameter*, Inverse Problems, 24 (2008), p. 034006.
[18] P. Speckman, *Spline smoothing and optimal rates of convergence in nonparametric regression models*, The Annals of Statistics, (1985), pp. 970–983.
[19] M. Stone, *Cross-validatory choice and assessment of statistical predictions*, Journal of the royal statistical society: Series B (Methodological), 36 (1974), pp. 111–133.
[20] C. Vogel, *Optimal choice of a truncation level for the truncated svd solution of linear first kind integral equations when data are noisy*, SIAM journal on numerical analysis, 23 (1986), pp. 109–117.
[21] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM journal on numerical analysis, 14 (1977), pp. 651–667.
[22] I. H. Witten and E. Frank, *Data mining: practical machine learning tools and techniques with java implementations*, Acm Sigmod Record, 31 (2002), pp. 76–77.



SNR	e_{gcv}	e_{opt}
1	1.2e0 ± 8.3e-1	8.9e-1 ± 1.1e-2
10	9.8e-1 ± 1.2e0	8.4e-1 ± 6.9e-3
10^2	8.3e-1 ± 7.1e-2	8.0e-1 ± 6.9e-3
10^3	7.4e-1 ± 1.2e-2	7.3e-1 ± 4.1e-3
10^4	6.7e-1 ± 4.6e-3	6.6e-1 ± 3.6e-3
10^5	6.1e-1 ± 2.4e-4	6.0e-1 ± 1.4e-4
10^6	6.1e-1 ± 1.9e-6	5.7e-1 ± 3.0e-5
10^7	6.1e-1 ± 2.1e-8	5.7e-1 ± 3.6e-7
10^8	6.1e-1 ± 8.3e-10	5.7e-1 ± 1.5e-8



SNR	e_{gcv}	e_{opt}
1	2.5e0 ± 1.0e1	3.8e-1 ± 1.1e-1
10	3.7e-1 ± 5.9e-1	1.8e-1 ± 2.3e-2
10^2	1.3e-1 ± 7.7e-2	9.0e-2 ± 1.0e-2
10^3	6.4e-2 ± 1.8e-2	5.3e-2 ± 2.8e-3
10^4	3.0e-2 ± 7.3e-3	2.7e-2 ± 1.2e-3
10^5	1.5e-2 ± 1.3-3	1.4e-2 ± 5.1e-4
10^6	7.9e-3 ± 5.5e-4	7.6e-3 ± 1.7e-4
10^7	3.7e-3 ± 7.3e-5	3.6e-3 ± 5.8e-5
10^8	3.3e-3 ± 6.0e-7	2.1e-3 ± 1.5e-5



SNR	e_{gcv}	e_{opt}
1	1.5e0 ± 4.2e0	2.2e-1 ± 4.2e-2
10	2.2e-1 ± 4.8e-1	6.6e-2 ± 1.4e-2
10^2	4.7e-2 ± 6.1e-2	2.4e-2 ± 3.4e-3
10^3	1.5e-2 ± 1.2e-2	8.4e-3 ± 1.5e-3
10^4	4.4e-3 ± 2.6e-3	3.1e-3 ± 3.0e-4
10^5	1.6e-3 ± 5.6e-4	1.3e-3 ± 1.2e-4
10^6	4.9e-4 ± 8.0e-5	4.3e-4 ± 3.5e-5
10^7	1.8e-4 ± 2.1e-5	1.6e-4 ± 7.9e-6
10^8	6.4e-5 ± 5.3e-6	6.0e-5 ± 2.6e-6

Figure 1. *Left column: Boxplots of the errors for 200 independent runs, with different signal-to-noise ratios (SNR). Right column: The corresponding sample mean and sample standard deviation of the errors. First row: rough solution. Second row: differentiable solution. Third row: twice differentiable solution.*