

REGULARISING LINEAR INVERSE PROBLEMS UNDER UNKNOWN NON-GAUSSIAN WHITE NOISE

BASTIAN HARRACH*, TIM JAHN†, AND ROLAND POTTHAST‡

Abstract. We deal with the solution of a generic linear inverse problem in the Hilbert space setting. The exact right hand side is unknown and only accessible through discretised measurements corrupted by white noise with unknown arbitrary distribution. The measuring process can be repeated, which allows to estimate the measurement error through averaging. We show convergence against the true solution of the infinite-dimensional problem for a priori and a posteriori regularisation schemes, as the number of measurements and the dimension of the discretisation tend to infinity, under natural and easily verifiable conditions for the discretisation.

Key words. statistical inverse problems, discretisation, white noise, discrepancy principle

AMS subject classifications. 65J22

1. Introduction and Preliminaries. We consider a compact linear operator $K : \mathcal{X} \rightarrow \mathcal{Y}$ between Hilbert spaces. The goal is to solve the ill-posed equation $K\hat{x} = \hat{y}$ for a given $\hat{y} \in \mathcal{D}(K^+)$, where K^+ is the generalised inverse and the right hand side \hat{y} is ad hoc unknown and has to be reconstructed from measurements. Solving the problem then typically requires specific a priori information about the noise. Here, our key assumption will be that we are able to perform multiple measurements and we do not require any other specific assumption for the error distribution of one measurement. Measuring the same quantity repeatedly is a standard engineering practice to decrease the measurement error, known as 'signal averaging' and was extensively studied in [13] in the context of infinite-dimensional inverse problems with (strongly L^2 -bounded) unknown noise. In this article we take discretisation into account and generalise the error distribution further to arbitrary unknown white noise.

As an arbitrary element of an infinite-dimensional space \hat{y} cannot be measured directly, but we may measure $l(\hat{y})$ for various linear functionals $l \in \mathcal{L}(\mathcal{Y}, \mathbb{R})$. If the unknown \hat{y} is for example a continuous function, one may think of performing point evaluations or measuring the integrals of that function over small parts of the domain. We will refer to these linear functionals as measurement channels in the following. We assume that we have multiple and unbiased samples on each measurement channel, corrupted randomly by additive noise. So,

$$(1) \quad Y_{ij} := l_j(\hat{y}) + \delta_{ij}$$

is the i -th sample on the j -th measurement channel, with $\|l_1\| = \|l_2\| = \dots$ and unbiased and independent measurement errors δ_{ij} , $i, j \in \mathbb{N}$ with arbitrary unknown distribution. Thus $(Y_{i1} - l_1(\hat{y}) \quad \dots \quad Y_{im} - l_m(\hat{y}))^T \in \mathbb{R}^m$, $i \in \mathbb{N}$ are i.i.d white noise vectors with unknown distribution. We assume, that $(l_j)_{j \in \mathbb{N}}$ is complete and square-summable, i.e. for all $y \in \mathcal{Y} \setminus \{0\}$ there is a l_j with $l_j(y) \neq 0$ and $\sum_{j=1}^{\infty} l_j(y)^2 < \infty$. For a fixed number m of measurement channels and a large number n of repetitions we obtain an approximation

*Institut für Mathematik, Goethe-Universität Frankfurt am Main (harrach@math.uni-frankfurt.de).

†Institut für Mathematik, Goethe-Universität Frankfurt am Main (jahn@math.uni-frankfurt.de).

‡Deutscher Wetterdienst, Offenbach am Main (Roland.Potthast@dwd.de).

$$\bar{Y}_n^{(m)} := \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n Y_{i1} \\ \dots \\ \frac{1}{n} \sum_{i=1}^n Y_{im} \end{pmatrix} \approx \begin{pmatrix} l_1(\hat{y}) \\ \dots \\ l_m(\hat{y}) \end{pmatrix}.$$

As a first approach we are using the ideas of Tikhonov and minimise the following functional with finite-dimensional residuum (fdr)

$$(2) \quad \arg \min_{x \in \mathcal{X}} \left\| \begin{pmatrix} l_1(Kx) \\ \dots \\ l_m(Kx) \end{pmatrix} - \bar{Y}_n^{(m)} \right\|_{\mathbb{R}^m}^2 + \alpha \|x\|_{\mathcal{X}}^2.$$

The main question of this work is whether the unique minimiser of (2), denoted by $R_\alpha^{(m)} \bar{Y}_n^{(m)}$, converges to \hat{x} for $m, n \rightarrow \infty$, for adequately chosen $\alpha = \alpha(m, n)$. Hereby, an important quantity is the measurement error $\|\bar{Y}_n^{(m)} - (l_1(\hat{y}) \dots l_m(\hat{y}))^T\|$, which by randomness is unknown and has to be guessed. The i.i.d assumption yields a natural estimator

$$(3) \quad \delta_{m,n}^{est} := \sqrt{\frac{m}{n} s_{m,n}^2},$$

where $s_{m,n}^2 := \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n (Y_{ij} - \frac{1}{n} \sum_{l=1}^n Y_{lj})^2$ is the mean of the sample variances. Clearly, $\delta_{m,n}^{est}$ converges to 0 in probability (and a.s. and in root mean square), iff $m/n \rightarrow 0$ (see Proposition 4.3). One of the most natural and popular strategies to determine the regularisation parameter α in (2) is the discrepancy principle [21], i.e. to solve

$$(4) \quad \left\| \begin{pmatrix} l_1(K R_\alpha^{(m)} \bar{Y}_n^{(m)}) \\ \dots \\ l_m(K R_\alpha^{(m)} \bar{Y}_n^{(m)}) \end{pmatrix} - \bar{Y}_n^{(m)} \right\| \approx \delta_{m,n}^{est}$$

(see Algorithm 1 with $C_0 = 1$ for the numerical implementation). We obtain the following convergence result for the discrepancy principle.

COROLLARY 1.1. *Assume that K is injective with dense range and that $(\delta_{ij})_{i,j \in \mathbb{N}}$ are independent and identically distributed with zero mean and bounded variance. Moreover assume, that $(l_j)_{j \in \mathbb{N}}$ is complete and square-summable. Then, with $\alpha_{m,n}$ determined by the discrepancy principle (4), we have that*

$$\lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty \\ m/n \rightarrow 0}} \mathbb{P} \left(\left\| R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y} \right\| \geq \varepsilon \right) = 0$$

for all $\varepsilon > 0$.

All the details to this result can be found in Section 2, where we also more generally treat filter based regularisations, as well as a priori parameter choice rules and discretisations $l_j^{(m)}$, $j = 1, \dots, m$, $m \in \mathbb{N}$. Let us stress, that Corollary 1.1 guarantees convergence without any quantitative knowledge of the quality of the discretisation

(error), for an arbitrary unknown error distribution. This might be surprising in view of the Bakushinskii veto [2], which states, that quantitative a priori knowledge about the noise is a crucial requirement for solving an inverse problem. Corollary 1.1 does not give a convergence rate, however, the numerical experiments in Section 5 indicate, that there might hold order optimality in various settings.

In order to obtain convergence rates we consider a second approach, which is about to first construct from the measured data in \mathbb{R}^m continuous measurements in the Hilbert space \mathcal{Y} , see e.g. the recent preprint [9]. For that we first solve the following optimisation problem

$$(5) \quad Z_n^{(m)} := \arg \min_{y \in \mathcal{Y}} \left\| \begin{pmatrix} l_1(y) \\ \dots \\ l_m(y) \end{pmatrix} - \bar{Y}_n^{(m)} \right\|.$$

We restrict to discretisations, for which (5) is well-conditioned, see Assumption 3.1. For general discretisations one would need to add an additional regularisation term. Then, instead of (2), we solve the following optimisation problem with infinite-dimensional residuum (idr)

$$(6) \quad \arg \min_{x \in \mathcal{X}} \left\| Kx - Z_n^{(m)} \right\|_{\mathcal{Y}}^2 + \alpha \|x\|_{\mathcal{X}}^2$$

and the regularisation parameter α has to be chosen accordingly to $\|Z_n^{(m)} - \hat{y}\|$. With

$$y^{(m)} := \arg \min_{y \in \mathcal{Y}} \left\| \begin{pmatrix} l_1(y) \\ \dots \\ l_m(y) \end{pmatrix} - \begin{pmatrix} l_1(\hat{y}) \\ \dots \\ l_m(\hat{y}) \end{pmatrix} \right\|$$

we may decompose this term into a measurement error and a discretisation error

$$\left\| Z_n^{(m)} - \hat{y} \right\| \leq \left\| Z_n^{(m)} - y^{(m)} \right\| + \left\| y^{(m)} - \hat{y} \right\|.$$

Assume that we know an asymptotic bound δ_m^{disc} for the discretisation error $\|\hat{y} - y^{(m)}\|$ (which is natural in various settings, see Section 3). One may estimate $\|Z_n^{(m)} - y^{(m)}\|$ (see Algorithm 2), and should use that many repetitions $n(m, \delta_m^{disc})$, such that this estimator approximately equals δ_m^{disc} . The regularisation parameter α is then again determined via the discrepancy principle

$$(7) \quad \left\| KR_\alpha Z_{n(m, \delta_m^{disc})}^{(m)} - Z_{n(m, \delta_m^{disc})}^{(m)} \right\| \approx 2\delta_m^{disc},$$

with $R_\alpha Z_{n(m, \delta_m^{disc})}^{(m)}$ the unique solution of (6) (see Algorithm 2 with $C_0 = 1$ for the numerical implementation). We obtain the following result on the convergence and the order optimality.

COROLLARY 1.2. *Assume that K is injective with dense range and that $(\delta_{ij})_{i,j \in \mathbb{N}}$ are independent with zero mean and finite variance. Moreover, the discretisation is complete and well-conditioned (see Proposition 3.3). Let $(\delta_m^{disc})_{m \in \mathbb{N}}$ be an known*

upper bound for the discretisation error converging to 0 and determine α_m with the discrepancy principle (7). Then

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\left\| R_{\alpha_m} Z_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \geq \varepsilon \right) = 0$$

for all $\varepsilon > 0$. If moreover there is a $0 < \nu \leq 1$ and a $\rho > 0$ such that $K^+ \hat{y} = (K^* K)^{\nu/2} w$ for some $w \in \mathcal{X}$ with $\|w\| \leq \rho$, then

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\left\| R_{\alpha_m} Z_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \leq L' \rho^{\frac{1}{\nu+1}} \delta_m^{\frac{\nu}{\nu+1}} \right) = 1$$

for some constant L' .

Solution strategies for inverse problems typically require a priori knowledge about the noise. For example, in the classical deterministic case an upper bound for the error is given, or in the stochastic case, one restricts to certain classes of distribution (often Gaussian). In [13] there was for the first time presented a rigorous convergence theory without any knowledge of the error distribution, if one has multiple measurements (strongly bounded in L^2) of the right hand side \hat{y} . Here we consider more realistic semi-discretised measurements, under arbitrary unknown white noise. It is widely known that discretisation has a regularising effect, see for example [18],[12] for the discretisation in the deterministic setting, [18], [19], [17] for the statistical frequentist setting and [16], [14] for the Bayesian approach. In general, one can either first regularise the infinite-dimensional problem and then discretise, or, as it is done in this article, one first discretises and then regularises the finite-dimensional problem. So far, inverse problems under white noise [7],[6] are treated the first way, and the white noise is modelled as a Hilbert space process operating on \mathcal{Y} , see [3], [5]. The major challenge of this modelling is, that then the measurements are not elements of \mathcal{Y} . This implies some drawbacks, e.g. one has to restrict to sufficiently smoothing operators and to include correction terms in the convergence rates (compared to the classical deterministic rates). Most importantly, the discrepancy principle, one of the most widely used parameter choice rules in practice, cannot be applied due to the unboundedness of the noise. Thus one rather relies on other parameter choice rules, f.e. cross validation [23] or Lepski's balancing principle [20], even though a modified discrepancy principle could be applied [4]. These technical difficulties are not present in the semi-discretised setting considered here. It is notable, that the main convergence result in this work guarantees convergence for arbitrary unknown distribution, as long as one is able to measure repeatedly, under quite general assumptions on the discretisation, which are only of qualitative nature and most importantly are independent of the unknown exact right hand side. In this paper we restrict to the discrepancy principle as an a posteriori rule, which is known to be challenging in stochastic regularisation even for strongly L^2 -bounded noise, see [13],[15]. Still, we expect that the results can be extended to other a posteriori parameter choice rules as well, since the central tools to handle the stochastic noise, namely Lemma 4.4 and Lemma 4.6, do not depend on the chosen regularisation or parameter choice rule.

The rest of the article is organised as follows. In Section 2 and Section 3 we will show the L^2 convergence (a.k.a. convergence of the mean squared error) of a priori parameter choice rules and the convergence in probability of the discrepancy principle for the both approaches respectively. The proofs are deferred to Section 4 and we conclude with a numerical study in Section 5 and some final remarks in Section 6.

2. Approach with finite-dimensional residuum. We start with a precise and more general definition of our discretisation scheme. Therefore we introduce as follows the discretisation (operators)

$$(8) \quad P_m : \mathcal{Y} \rightarrow \mathbb{R}^m, \quad y \mapsto \begin{pmatrix} l_1^{(m)}(y) \\ \dots \\ l_m^{(m)}(y) \end{pmatrix},$$

with the corresponding measurements

$$Y_{ij}^{(m)} := l_j^{(m)}(\hat{y}) + \delta_{ij}^{(m)}.$$

and $\|l_1^{(m)}\| = \dots = \|l_m^{(m)}\|$. That is, the measurement channels and also the error distribution may depend on the number m of measurement channels now. We will often use, that by the Riesz representation theorem there are unique $(\eta_j^{(m)})_{j \leq m, m \in \mathbb{N}}$ such that $l_j^{(m)}(y) = (\eta_j^{(m)}, y)$ for all $y \in \mathcal{Y}$. For convenience we will assume that P_m is injective.

From now on we consider more generally filter-based regularisations $R_\alpha^{(m)} := F_\alpha((P_m K)^* P_m K)(P_m K)^*$, where $(F_\alpha)_\alpha$ fulfills Assumption 2.1 below.

Assumption 2.1 (Filter). $(F_\alpha)_{\alpha > 0}$ is a family of piecewise continuous real valued functions on $[0, \|K\|^2]$, with

$$(9) \quad \lim_{\alpha \rightarrow 0} \sup_{\varepsilon \leq \lambda \leq \|K\|^2} |F_\alpha(\lambda) - 1/\lambda| = 0$$

for all $\varepsilon > 0$ and $\lambda |F_\alpha(\lambda)| \leq C_R \in \mathbb{R}$ for all $\lambda \in (0, \|K\|^2]$ and $\alpha > 0$. Moreover it has qualification $\nu_0 \geq 0$, i.e. ν_0 is maximal such that for all $0 \leq \nu \leq \nu_0$ there is a constant $C_\nu \in \mathbb{R}$ such that

$$\sup_{\lambda \in (0, \|K\|^2]} \lambda^{\frac{\nu}{2}} |1 - F_\alpha(\lambda)\lambda| \leq C_\nu \alpha^{\frac{\nu}{2}}.$$

Hereby, for $\nu = 0$ the constant C_0 is assumed to be known. Finally, there is a constant $C_F \in \mathbb{R}$ with $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all $\alpha > 0$ and $\lambda \in (0, \|K\|^2]$.

Remark 2.2. Assumption 2.1 coincides with the classical ones in [8] up to (9), which is usually replaced by the weaker condition $\lim_{\alpha \rightarrow 0} F_\alpha(\lambda) = 1/\lambda$, for all $\lambda \in (0, \|K\|^2]$. However, it is easy to verify that the generating filter of all popular methods, e.g. truncated singular value, (iterated) Tikhonov or Landweber regularisation, fulfill Assumption 2.1. In all this cases it holds that $C_0 = 1$.

We impose the following more abstract condition on the discretisation, which generalises the ones from the introduction.

Assumption 2.3 (Discretisation for finite-dimensional residuum). There exists an injective operator $A \in \mathcal{L}(\mathcal{Y})$ such that $\lim_{m \rightarrow \infty} P_m^* P_m y = Ay$ for all $y \in \mathcal{Y}$.

We list some popular discretisation schemes which fulfill Assumption 2.3, starting with the one from the introduction.

PROPOSITION 2.4. Assume that $l_j^{(m)} = l_j$ for all $j = 1, \dots, m$ and $m \in \mathbb{N}$ with $(l_j)_{j \in \mathbb{N}} \subset \mathcal{L}(\mathcal{Y}, \mathbb{R})$, where $(l_j)_{j \in \mathbb{N}}$ is complete and square-summable, i.e. for all $y \in \mathcal{Y} \setminus \{0\}$ there is a l_j such that $l_j(y) \neq 0$ and there holds $\sum_{j=1}^{\infty} l_j(y)^2 < \infty$. Then Assumption 2.3 is fulfilled.

Often the limit operator A will be the identity $Id = Id_{\mathcal{Y}}$, e.g. in the case when we discretise by box or hat functions.

PROPOSITION 2.5. Assume that $\mathcal{Y} = L^2([0, 1])$ and we discretise by box functions, i.e. $l_j^{(m)} = (\eta_j^{(m)}, \cdot)$, with $\eta_j^{(m)} = \sqrt{m} \chi_{[\frac{j-1}{m}, \frac{j}{m}]}$ for $j = 1, \dots, m$ and $m \geq 2$. Then Assumption 2.3 is fulfilled with $A = Id$.

PROPOSITION 2.6. Assume that $\mathcal{Y} = L^2([0, 1])$ and we discretise by hat functions, i.e. $l_j^{(m)} = (\eta_j^{(m)}, \cdot)$, with

1. $\frac{\eta_j^{(m)}}{\sqrt{m-1}} = (1-j+(m-1)x) \chi_{[\frac{j-1}{m-1}, \frac{j}{m-1}]} + (j+1-(m-1)x) \chi_{[\frac{j}{m-1}, \frac{j+1}{m-1}]}$ for $j = 2, \dots, m-1$,
2. $\eta_1^{(m)} = \sqrt{2(m-1)}(1+j-(m-1)x) \chi_{[\frac{j}{m-1}, \frac{j+1}{m-1}]}$,
3. $\eta_m^{(m)} = \sqrt{2(m-1)}((m-1)x-j+1) \chi_{[\frac{j-1}{m-1}, \frac{j}{m-1}]}$.

Then Assumption 2.3 is fulfilled with $A = Id$.

2.1. A priori regularisation with finite-dimensional residuum. We start with a priori regularisations and impose the following assumption on the error, which is weaker than the one in the introduction. Basically, solely independence on each measurement channel and a uniform boundedness of the variances are required.

Assumption 2.7 (Error for a priori regularisation). For all $m, j \in \mathbb{N}$, the random variables $(\delta_{ij}^{(m)})_{i \in \mathbb{N}}$ are independent with zero mean and there exists $C_d \in \mathbb{R}$ with

$$\sup_{\substack{m, i, j \in \mathbb{N} \\ j \leq m}} \mathbb{E}[\delta_{ij}^{(m)2}] \leq C_d.$$

Since the sample variance depends on the data, we set $s_{m,n}^2 = 1$ here, such that $\delta_{m,n}^{est} = \sqrt{m/n}$. This has the advantage, that the regularisation parameter α is independent of the measurements $Y_{ij}^{(m)}$. We obtain convergence in L^2 for a priori regularisation.

THEOREM 2.8. Assume that K is injective, the discretisation fulfills Assumption 2.3, the error is accordingly to Assumption 2.7 and $(F_\alpha)_{\alpha > 0}$ fulfills Assumption 2.1. Take an a priori parameter choice rule with $\alpha(\delta) \xrightarrow{\delta \rightarrow 0} 0$ and $\delta/\sqrt{\alpha(\delta)} \xrightarrow{\delta \rightarrow 0} 0$. Then there holds

$$\lim_{\substack{m, n \rightarrow \infty \\ m/n \rightarrow 0}} \mathbb{E} \left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y} \right\|^2 = 0.$$

2.2. A posteriori regularisation with finite dimensional residuum. We turn our attention to the discrepancy principle. The regularisation parameter is determined through

$$(10) \quad \left\| P_m K R_{\alpha}^{(m)} \bar{Y}_n^{(m)} - \bar{Y}_n^{(m)} \right\| \approx \delta_{m,n}^{est}$$

and in the definition of $\delta_{m,n}^{est} = \sqrt{s_{m,n}^2 m/n}$ we choose the mean of the sample variances

$$s_{m,n}^2 := \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2,$$

since we will need a sharp estimation of the right hand side. We implement the discrepancy principle with Algorithm 1.

Algorithm 1 Discrepancy principle with fdr approach

- 1: Choose $\tau > C_0$ (from Assumption 2.1) and $q \in (0, 1)$;
 - 2: Input: Measurements $Y_{ij}^{(m)} = l_j^{(m)}(\hat{y}) + \delta_{ij}^{(m)}$ with $i \leq n$ and $j \leq m$;
 - 3: Set $\bar{Y}_n^{(m)} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} Y_{i1}^{(m)} \\ \dots \\ Y_{im}^{(m)} \end{pmatrix}$;
 - 4: Set $\delta_{m,n}^{est} = \sqrt{\frac{m}{n} \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2}$;
 - 5: $k = 0$;
 - 6: **while** $\left\| \begin{pmatrix} l_1^{(m)}(KR_{q^k} \bar{Y}_n^{(m)}) \\ \dots \\ l_n^{(m)}(KR_{q^k} \bar{Y}_n^{(m)}) \end{pmatrix} - \bar{Y}_n^{(m)} \right\| > \tau \delta_{m,n}^{est}$ **do**
 - 7: $k = k + 1$;
 - 8: **end while**
 - 9: $\alpha_{m,n} = q^k$;
-

Algorithm 1 terminates (with probability tending to 1 for $m \rightarrow \infty$), if K has dense range and (for m large enough) $\mathbb{E}(Y_{11}^{(m)} - \mathbb{E}Y_{11}^{(m)})^2 > 0$, for details see [13]. We extend the assumptions of the error in the introduction.

Assumption 2.9 (Error for a posteriori regularisation). It holds that either

1. the random variables $(\delta_{ij}^{(m)})_{i,j,m \in \mathbb{N}}$ are i.i.d. with zero mean and bounded variance, or,
2. there are $C_d \in \mathbb{R}$ and $p > 1$ such that for all $m \in \mathbb{N}$ the random variables $(\delta_{ij}^{(m)})_{i,j \in \mathbb{N}}$ are i.i.d with zero mean and $\frac{\mathbb{E}|\delta_{ij}^{(m)}|^{2p}}{(\mathbb{E}\delta_{ij}^{(m)2})^p} \leq C_d$.

The main difference between Assumption 2.9.1 and 2.9.2 is, that for the latter the error distribution may vary with m , to the cost of a uniform moment condition.

Remark 2.10. Assumption 2.9.2 guarantees, that the error distribution does not degenerate too much. It is trivially fulfilled, if f.e. $\delta_{ij}^{(m)} \stackrel{d}{=} c_m X$, with $\mathbb{E}|X|^{2p} < \infty, (c_m)_{m \in \mathbb{N}} \subset \mathbb{R} \setminus \{0\}$.

Now we are ready to prove convergence of the discrepancy principle. In contrast to the previous section, where we showed convergence in L^2 for a priori regularisation methods, the result will now be on convergence in probability (compare this to the counter example in 3.1 in [13]).

THEOREM 2.11. *Assume that K is injective with dense range and that the discretisation fulfills Assumption 2.3 and that the error is accordingly to Assumption 2.9*

and $(F_\alpha)_{\alpha>0}$ fulfills Assumption 2.1 with a qualification $\nu_0 > 1$. Then, with $\alpha_{m,n}$ the output of Algorithm 1,

$$\lim_{\substack{m,n \rightarrow \infty \\ m/n \rightarrow 0}} \mathbb{P} \left(\left\| R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y} \right\| \geq \varepsilon \right) = 0$$

for all $\varepsilon > 0$.

Corollary 1.1 is an easy consequence of Theorem 2.11 and Proposition 2.4. We conclude the section with some remarks regarding Assumptions 2.3 and 2.9.

Remark 2.12. A natural generalisation of the assumption in the introduction, that $(l_j)_{j \in \mathbb{N}}$ is complete, would be the following: For all $y \in \mathcal{Y} \setminus \{0\}$ there is a $\varepsilon > 0$ such that $\|P_m y\| \geq \varepsilon$ for m large enough. However, the following counter example shows that this condition is not sufficient to guarantee that the discretisation error tends to 0: Let $(e_j)_{j \in \mathbb{N}}$ be an orthonormal basis of \mathcal{Y} . Set $l_j^{(m)}(y) = (y, e_j)$ for $j = 2, \dots, m$ and $l_1^{(m)}(y) = (y, e_1/\sqrt{2} + e_{m+1}/\sqrt{2})$. For $y \neq 0$ we set $\varepsilon = |(y, e_j)|/2$ with $j = \min\{j' : (y, e_{j'}) \neq 0\}$. Then clearly, $\|P_m y\| \geq \varepsilon$ for m large enough. But, it is $\mathcal{N}(P_m) = \langle e_1/\sqrt{2} - e_{m+1}/\sqrt{2}, e_{m+2}, e_{m+3}, \dots \rangle$, thus

$$e_1 - P_m^+ P_m e_1 = P_{\mathcal{N}(P_m)} e_1 = e_1/\sqrt{2} \not\rightarrow 0$$

for $m \rightarrow \infty$.

Remark 2.13. As already mentioned, Assumption 2.9 excludes distributions which are too degenerated and guarantees, that $\mathbb{E} \delta_{11}^{(m)2}$ is in some sense uniformly estimatable. We quickly sketch what can go wrong, if the distributions degenerate too much. Assume that $(\delta_{ij}^{(m)})_{ij}$ are i.i.d. for all $m \in \mathbb{N}$, with

$$\mathbb{P}(\delta_{ij}^{(m)} = x) = \begin{cases} \frac{1}{m^4} & \text{for } x = -\sqrt{m^4 - 1} \\ \frac{m^4 - 1}{m^4} & \text{for } x = 1/\sqrt{m^4 - 1} \end{cases}.$$

Thus $\mathbb{E} \delta_{11}^{(m)} = 0$ and $\mathbb{E} \delta_{11}^{(m)2} = 1$, but, for any $p > 1$,

$$\frac{\mathbb{E} |\delta_{11}^{(m)}|^{2p}}{(\mathbb{E} \delta_{11}^{(m)2})^p} \geq \frac{1}{m^4} |\sqrt{m^4 - 1}|^{2p} = \left(1 - \frac{1}{m^4}\right) |m^4 - 1|^p \rightarrow \infty$$

as $m \rightarrow \infty$. Thus Assumption 2.9 is violated and with the choice $n(m) = m^2$ it holds that $\lim_{m \rightarrow \infty} \frac{m}{n(m)} = 0$, but we have that

$$\begin{aligned} \mathbb{P}(\delta_{m,n(m)}^{est} = 0) &= \mathbb{P}(s_{m,n(m)}^2 = 0) \\ &= \mathbb{P}(\delta_{ij}^{(m)} = 1/\sqrt{m^4 - 1}, i = 1, \dots, m^2, j = 1, \dots, m) \\ &= \left(1 - \frac{1}{m^4}\right)^{m^3} = \left(\left(1 - \frac{1}{m^4}\right)^{m^4}\right)^{\frac{1}{m}} \rightarrow 1 \end{aligned}$$

as $m \rightarrow \infty$. Thus with asymptotic probability 1 the discrepancy principle cannot even be applied for this choice of n . The number of repetitions $n(m) = m^2$ is simply too small to estimate the variance of $\delta_{11}^{(m)}$ adequately.

3. Approach with infinite-dimensional residuum. We turn our attention to the second approach (6). The strategy is to use the measured data to construct virtual measurements in the infinite-dimensional Hilbert space \mathcal{Y} and then to regularise the infinite-dimensional problem using classical methods. For the regularisation we will need in the following an upper bound for the discretisation error, which we denote by $\delta_m^{disc} \geq \|\hat{y} - P_m^+ P_m \hat{y}\|$. Decomposing the true data error yields

$$\left\| \hat{y} - P_m^+ \bar{Y}_n^{(m)} \right\| \leq \left\| \hat{y} - P_m^+ P_m \hat{y} \right\| + \left\| P_m^+ P_m \hat{y} - P_m^+ \bar{Y}_n^{(m)} \right\|.$$

As in the approach with a finite-dimensional residuum, there is a generic way (given below) to estimate the (projected) measurement error $\|P_m^+ \bar{Y}_n^{(m)} - P_m^+ P_m \hat{y}\|$. So that it is natural to choose the number of repetitions $n = n(m, \delta_m^{disc})$ in such a way, that this estimator approximately equals the discretisation error δ_m^{disc} . After that one may use any deterministic regularisation together with total estimated noise level

$$(11) \quad 2\delta_m^{disc} \approx \left\| \hat{y} - P_m^+ P_m \hat{y} \right\| + \left\| P_m^+ P_m \hat{y} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\| \geq \left\| \hat{y} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\|.$$

We again consider regularisations $R_\alpha := F_\alpha(K^*K)K^*$ induced by a regularising filter (see Assumption 2.1) and make the following assumptions for the discretisation and our a priori knowledge of it.

Assumption 3.1 (Discretisation for infinite-dimensional residuum). We assume that we know an asymptotic upper bound $(\delta_m^{disc})_{m \in \mathbb{N}}$ for the discretisation error and asymptotic upper and lower bounds $(c_m)_{m \in \mathbb{N}}, (C_m)_{m \in \mathbb{N}}$ for the singular values $(\sigma_j^{(m)})_{j \leq m, m \in \mathbb{N}}$ of $(P_m)_{m \in \mathbb{N}}$. More precisely, these bounds have to fulfill $\|\hat{y} - P_m^+ P_m \hat{y}\| \leq \delta_m^{disc}, 0 < c_m \leq \sigma_j^{(m)} \leq C_m$ for all $j = 1, \dots, m$ and m large enough, and $\delta_m^{disc} \rightarrow 0$ as $m \rightarrow \infty$ and

$$(12) \quad \limsup_{m \rightarrow \infty} \kappa(P_m) := \limsup_{m \rightarrow \infty} \|P_m\| \|P_m^+\| = \limsup_{m \rightarrow \infty} \frac{\max_{j=1, \dots, m} \sigma_j^{(m)}}{\min_{j=1, \dots, m} \sigma_j^{(m)}} \leq \limsup_{m \rightarrow \infty} \frac{C_m}{c_m} < \infty.$$

Often the stability assumption (12) can be guaranteed by an angle condition for the unique $\eta_j^{(m)} \in \mathcal{Y}$, which fulfill $l_j^{(m)}(y) = (\eta_j, y)$ for all $y \in \mathcal{Y}$.

PROPOSITION 3.2. *Assume that*

$$\sup_{m \in \mathbb{N}} \sup_{j \leq m} \sum_{i \neq j} \frac{|(\eta_i^{(m)}, \eta_j^{(m)})|}{\|\eta_1^{(m)}\|^2} \leq c < 1.$$

Then $c_m := \|\eta_1^{(m)}\|^2(1 - c) \leq \sigma_j^{(m)} \leq \|\eta_1^{(m)}\|^2(1 + c) =: C_m$ for $j = 1, \dots, m$ and m large enough and thus $\kappa(P_m) \leq \frac{1+c}{1-c}$.

Clearly, the angle condition is always satisfied for orthogonal discretisations. We now show that Assumption 3.1 is fulfilled for various popular discretisation schemes. We start with the example from the introduction.

PROPOSITION 3.3. *Assume that $l_j^{(m)} = l_j = (\eta_j, \cdot)$ for all $j = 1, \dots, m$ and $m \in \mathbb{N}$, with $(l_j)_{j \in \mathbb{N}} \subset \mathcal{L}(\mathcal{Y}, \mathbb{R})$ and $(\eta_j)_{j \in \mathbb{N}} \subset \mathcal{Y}$, and that we know c and δ_m^{disc} such that $\delta_m^{disc} \geq \|\hat{y} - P_m^+ P_m \hat{y}\|$ and $(l_j)_{j \in \mathbb{N}}$ is complete, i.e. for all $y \in \mathcal{Y} \setminus \{0\}$ there is a l_j*

such that $l_j(y) \neq 0$, and well-conditioned, that is

$$\sup_{j \in \mathbb{N}} \sum_{\substack{i=1 \\ i \neq j}}^{\infty} |(\eta_i, \eta_j)| / \|\eta_1\|^2 \leq c < 1.$$

Then Assumption 3.1 is fulfilled for δ_m^{disc} and $c_m = 1 - c$, $C_m = 1 + c$.

Next we consider discretisation along the singular directions of K , see the beginning of Section 4 for the definition of the singular value decomposition.

PROPOSITION 3.4. *Assume that the singular value decomposition $(\sigma_l, v_l, u_l)_{l \in \mathbb{N}}$ of K is known. Then for the discretisation $l_j^{(m)} = (u_j, \cdot)$ Assumption 3.1 is (asymptotically) fulfilled, with the bounds $\delta_m^{disc} = f_m \sigma_{m+1}$ (where f_m is any sequence with $f_m \rightarrow \infty$ as $m \rightarrow \infty$) and $c_m = C_m = 1$.*

In many important cases, for example if K is a Fredholm integral equation with sufficient smoothing kernel, Assumption 3.1 is also fulfilled for discretisation with box or hat functions.

PROPOSITION 3.5. *Consider $\mathcal{X} = \mathcal{Y} = L^2(0, 1)$ and $\eta_j^{(m)}$ the box functions from Proposition 2.5. If \hat{y} is continuously differentiable, then Assumption 3.1 is fulfilled with bounds $\delta_m = f_m/m$ and $c_m = C_m = 1$, where $(f_m)_m$ is arbitrary with $\lim_m f_m = \infty$.*

PROPOSITION 3.6. *Consider $\mathcal{X} = \mathcal{Y} = L^2(0, 1)$ and $\eta_j^{(m)}$ the hat functions from Proposition 2.6. If \hat{y} is continuously differentiable, then Assumption 3.1 is fulfilled with bounds $\delta_m = f_m/m$ and $c_m = 1/6$ and $C_m = 7/6$, where $\lim_m f_m = \infty$. If \hat{y} is twice continuously differentiable, then Assumption 3.1 is fulfilled with bounds $\delta_m = f_m/m^2$ and $c_m = 1/6$ and $C_m = 7/6$, with $\lim_m f_m = \infty$.*

It remains to determine the number of repetitions $n(m, \delta_m^{disc})$, such that the (back projected) measurement error fulfills $\|P_m^+ P_m \hat{y} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)}\| \approx \delta_m^{disc}$. This number depends on the singular value decomposition of P_m and the variance $\mathbb{E} \delta_{11}^{(m)2}$. More precisely, with $(\sigma_j^{(m)}, v_j^{(m)}, u_j^{(m)})_{j \leq m}$ the singular value decomposition of P_m and $e_1^{(m)}, \dots, e_m^{(m)}$ the standard basis of \mathbb{R}^m , it is

$$\begin{aligned} \|P_m^+ \bar{Y}_n^{(m)} - P_m^+ P_m \hat{y}\|^2 &= \sum_{j=1}^m \frac{1}{\sigma_j^{(m)2}} \left(\sum_{l=1}^m \sum_{i=1}^n \frac{\delta_{ij}^{(m)}}{n} (u_j^{(m)}, e_l^{(m)}) \right)^2 \\ \implies \mathbb{E} \|P_m^+ \bar{Y}_n^{(m)} - P_m^+ P_m \hat{y}\|^2 &= \frac{\mathbb{E} \delta_{11}^{(m)2}}{n} \sum_{j=1}^m \frac{1}{\sigma_j^{(m)2}}. \end{aligned}$$

Thus with our lower bound $c_m \leq \sigma_j^{(m)}$ we determine

$$n(m, \delta_m^{disc}) := \min \left\{ n \geq 2 : \frac{m s_{m,n}^2}{n c_m^2} \leq \delta_m^{disc2} \right\},$$

with $s_{m,n}^2 = 1$ or $s_{m,n}^2 = \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2$.

3.1. A priori regularisation with infinite-dimensional residuum. For a priori regularisations we set $s_{m,n}^2 = 1$, so that $n(m, \delta)$ and the measurements $Y_{ij}^{(m)}$ are independent. The convergence result holds true with the same assumption for the error as in Section 2.1.

THEOREM 3.7. *Assume that K is injective, the discretisation fulfills Assumption 3.1, the error is accordingly to Assumption 2.7 and $(F_\alpha)_{\alpha>0}$ fulfills Assumption 2.1. Take an a priori parameter choice rule with $\alpha(\delta) \xrightarrow{\delta \rightarrow 0} 0$ and $\delta/\sqrt{\alpha(\delta)} \xrightarrow{\delta \rightarrow 0} 0$. Then there holds*

$$\lim_{m \rightarrow \infty} \mathbb{E} \left\| R_{\alpha(\delta_m^{disc})} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\|^2 = 0$$

$$\text{for } n(m, \delta_m^{disc}) = \left\lceil \frac{m}{c_m^2 \delta_m^{disc^2}} \right\rceil.$$

Remark 3.8. Note that for a priori regularisation one can relax the condition on δ_m^{disc} in Assumption 3.1 to $\lim_{m \rightarrow \infty} \delta_m^{disc} = 0$ and $\limsup_{m \rightarrow \infty} \frac{\delta_m^{disc}}{\|\hat{y} - P_m^+ P_m \hat{y}\|} > 0$.

3.2. A posteriori regularisation with infinite-dimensional residuum. We determine the stopping index $n(m, \delta_m^{disc})$ more accurately with the sample variance and set $s_{m,n}^2 := \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2$. We implement the discrepancy principle in Algorithm 2.

Algorithm 2 Discrepancy principle with idr approach

- 1: Choose $\tau > C_0$ (from Assumption 2.1) and $q \in (0, 1)$;
 - 2: Input: Number of measurement channels m , measurements $Y_{ij}^{(m)}$, $j \leq m, i \in \mathbb{N}$, upper bound δ_m^{disc} for discretisation error, lower bound c_m for singular values of P_m ;
 - 3: Determine $n(m, \delta_m^{disc}) := \min \left\{ n' \geq 1 : \frac{ms_{m,n'}^2}{n'c_m^2} \leq \delta_m^{disc^2} \right\}$ from measurements $Y_{ij}^{(m)}$.
 - 4: Set $\bar{Y}_{n(m, \delta_m^{disc})}^{(m)} = \frac{1}{n(m, \delta_m^{disc})} \sum_{i=1}^{n(m, \delta_m^{disc})} \begin{pmatrix} Y_{i1}^{(m)} \\ \dots \\ Y_{in}^{(m)} \end{pmatrix}$;
 - 5: $k = 0$;
 - 6: **while** $\left\| (KR_{q^k} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)}) \right\| > 2\tau \delta_m^{disc}$ **do**
 - 7: $k = k + 1$;
 - 8: **end while**
 - 9: $\alpha_m = q^k$;
-

Algorithm 2 terminates under the same conditions as Algorithm 1. The back propagating of the measurements induces correlations, which forces us to impose slightly stricter conditions on the error distribution than in the setting before. On the other hand, the regularisation is now done in \mathcal{Y} (no matter which m), which allows to use classical results to obtain a convergence rate.

THEOREM 3.9. *Assume that K is injective with dense range and that the discretisation fulfills Assumption 3.1 and that the error is accordingly to Assumption 2.9, with $p \geq 2$ in the case of 2.9.2, and $(F_\alpha)_{\alpha>0}$ fulfills Assumption 2.1 with a qualification*

$\nu_0 > 1$. For $\tau > C_0$, let α_m and $\bar{Y}_{n(m, \delta_m^{disc})}^{(m)}$ be the output of the discrepancy principle as implemented in Algorithm 2. Then

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \geq \varepsilon \right) = 0.$$

If moreover, there is a $0 < \nu \leq \nu_0 - 1$ and a $\rho > 0$ such that $K^+ \hat{y} = (K^* K)^\nu / 2 w$ for some $w \in \mathcal{X}$ with $\|w\| \leq \rho$, then

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \leq L' \rho^{\frac{1}{\nu+1}} (\delta_m^{disc})^{\frac{\nu}{\nu+1}} \right) = 1$$

for some constant L' .

Now Corollary 1.2 in the introduction is an easy consequence of Theorem 3.9 and Proposition 3.3.

4. Proofs. In this section we collect the proofs. We will need the singular value decomposition of an injective compact operator A (see [5]): there exists a monotone sequence $\|A\| = \sigma_1 \geq \sigma_2 \geq \dots > 0$. Moreover there are families of orthonormal vectors $(u_l)_{l \leq \dim(\mathcal{R}(A))}$ and $(v_l)_{l \leq \dim(\mathcal{R}(A))}$ with $\overline{\text{span}(u_l : l \leq \dim(\mathcal{R}(A)))} = \mathcal{R}(A)$, $\overline{\text{span}(v_l : l \leq \dim(\mathcal{R}(A)))} = \mathcal{N}(A)^\perp$ such that $Av_l = \sigma_l v_l$ and $A^* u_l = \sigma_l v_l$.

4.1. Proofs for finite-dimensional residuum. The assumptions for the discretisation when using the first approach (with finite-dimensional residuum) are such, that the discretised operators $K^* P_m^* P_m K$ converge uniformly to a compact and injective operator $K^* A K$. The uniform convergence guarantees, that the eigenvalues and spaces of the former converge pointwise to the ones of the latter, and the injectivity of the limit operator assures, that the unknown \hat{x} is determined arbitrarily precisely by finitely many eigenvectors of the latter. We make this precise with the following lemma.

LEMMA 4.1. *Assume that K is injective and that Assumption 2.3 holds true. Then*

$$\|K^* P_m^* P_m K - K^* A K\| \rightarrow 0$$

for $m \rightarrow \infty$ and $K^* A K$ is injective, compact, self-adjoint and positive semidefnite. Denote by $(\lambda_j^{(m)})_{j \leq m}$ and $(\lambda_j^{(\infty)})_{j \in \mathbb{N}}$ the nonzero eigenvalues with corresponding orthonormal eigenvectors $(v_j^{(m)})_{j \leq m}$ of $K^* P_m^* P_m K$ and $K^* A K$ respectively, ordered decreasingly. Then

1. $\lim_{m \rightarrow \infty} \lambda_j^{(m)} = \lambda_j^{(\infty)}$ for all $j \in \mathbb{N}$, and
2. for all $x \in \mathcal{X}$ and $\varepsilon > 0$, there is a $M = M(x, \varepsilon) \in \mathbb{N}$, such that

$$\limsup_{m \rightarrow \infty} \sum_{j=M+1}^m (x, v_j^{(m)})^2 \leq \varepsilon.$$

Proof. Denote by (σ_j, u_j, v_j) the singular value decomposition of K and set $A_m = P_m^* P_m$ and $C := \max \{\|A\|, \sup_m \|A_m\|\} < \infty$ (uniform boundedness principle). For $\varepsilon > 0$ arbitrary define $M \in \mathbb{N}$ implicitly through $2C\sigma_{M+1} \leq \varepsilon/2$. Then

$$\begin{aligned}
 & \|A_m K - AK\| \\
 &= \sup_{\substack{x \in \mathcal{X} \\ \|x\|=1}} \|A_m Kx - AKx\| = \sup_{\substack{\sum \alpha_j^2=1 \\ x=\sum \alpha_j u_j}} \left\| \sum_{j=1}^{\infty} \alpha_j (A_m K - AK) u_j \right\| \\
 &\leq \sup_{\substack{\sum \alpha_j^2=1 \\ x=\sum \alpha_j u_j}} \sum_{j=1}^M \sigma_j |\alpha_j| \|(A_m - A)v_j\| + \sup_{\substack{\sum \alpha_j^2=1 \\ x=\sum \alpha_j u_j}} \left\| (A_m - A) \sum_{j=M+1}^{\infty} \sigma_j \alpha_j v_j \right\| \\
 &\leq \sigma_1 \sum_{j=1}^M \|(A_m - A)v_j\| + \|A_m - A\| \sup_{\substack{\sum \alpha_j^2=1 \\ x=\sum \alpha_j u_j}} \left\| \sum_{j=M+1}^{\infty} \sigma_j \alpha_j v_j \right\| \\
 &\leq \sigma_1 \sum_{j=1}^M \|(A_m - A)v_j\| + 2C\sigma_{M+1} \leq \sigma_1 \sum_{j=1}^M \|(A_m - A)v_j\| + \varepsilon/2
 \end{aligned}$$

Because $A_m \rightarrow A$ pointwise there is an $m_0 \in \mathbb{N}$, such that $\sigma_1 \sum_{j=1}^M \|(A_m - A)v_j\| \leq \varepsilon/2$ for all $m \geq m_0$, thus $A_m K \rightarrow AK$ and therefore $K^* A_m K \rightarrow K^* AK$ for $m \rightarrow \infty$ uniformly. Since $K^* P_m^* P_m K$ is compact, self-adjoint and positive semidefinite, so is $K^* AK$ as its uniform limit. Then (1.) holds by Section 6 of [1]. We define iteratively $I_1 := \{j : \lambda_j^{(\infty)} = \lambda_1^{(\infty)}\}$, $I_i := \{j : \lambda_j^{(\infty)} = \lambda_{\max(I_{i-1})+1}^{(\infty)}\}$. So the cardinality of I_i is the algebraic multiplicity of the i -th largest eigenvalue of $K^* AK$. We define the corresponding eigenspaces $E_i := \text{span}(v_j^{(\infty)}, j \in I_i)$, $E_i^m := \text{span}(v_j^{(m)}, j \in I_i)$. With $P_{E_i}, P_{E_i^m}$ the orthogonal projections onto E_i and E_i^m , by Theorem 7.1 of [1] there is a constant C_i such that $\|P_{E_i^m} - P_{E_i}\| \leq C_i \|K^* P_m^* P_m K - K^* AK\|$ (for m sufficiently large). Thus there is a $M \in \mathbb{N}$ with $M = \sum_{i=1}^{i^*} |I_i|$ for some $i^* \in \mathbb{N}$ such that

$$\begin{aligned}
 \left| \sum_{j=1}^M (\hat{x}, v_j^{(m)})^2 - \sum_{j=1}^M (\hat{x}, v_j^{(\infty)})^2 \right| &\leq \sum_{i=1}^{i^*} \left| \|P_{E_i^m} \hat{x}\|^2 - \|P_{E_i} \hat{x}\|^2 \right| \\
 &\leq \sum_{i=1}^{i^*} (\|P_{E_i^m} \hat{x}\| + \|P_{E_i} \hat{x}\|) \|P_{E_i^m} \hat{x}\| - \|P_{E_i} \hat{x}\| \\
 &\leq 2\|\hat{x}\| \sum_{i=1}^{i^*} \|P_{E_i^m} \hat{x} - P_{E_i} \hat{x}\| \\
 &\leq 2\|\hat{x}\|^2 \|K^* P_m^* P_m K - K^* AK\| \sum_{i=1}^{i^*} C_i \leq \varepsilon/2
 \end{aligned}$$

for m sufficiently large and

$$\left| \|\hat{x}\|^2 - \sum_{j=1}^M (\hat{x}, v_j^{(\infty)})^2 \right| = \sum_{j=M+1}^{\infty} (\hat{x}, v_j^{(\infty)})^2 \leq \varepsilon/2,$$

where the second assertion followed from the injectivity of $K^* AK$. Thus

$$\begin{aligned}
\sum_{j=M+1}^m \left(\hat{x}, v_j^{(m)} \right)^2 &\leq \left\| P_{(P_m K)^\perp} \hat{x} \right\|^2 - \sum_{j=1}^M \left(\hat{x}, v_j^{(m)} \right)^2 \\
&\leq \|\hat{x}\|^2 - \sum_{j=1}^M \left(\hat{x}, v_j^{(\infty)} \right)^2 + \sum_{j=1}^M \left(\hat{x}, v_j^{(m)} \right)^2 - \sum_{j=1}^M \left(\hat{x}, v_j^{(\infty)} \right)^2 \leq \varepsilon
\end{aligned}$$

for m sufficiently large. \square

4.1.1. Proof of Proposition 2.4. It is $\sup_{m \in \mathbb{N}} \|P_m y\| = \sup_{m \in \mathbb{N}} \sum_{j=1}^m l_j(y)^2 < \infty$, thus $\sup_m \|P_m\| < \infty$ and with the embedding $\mathbb{R}^m \subset l^2(\mathbb{N})$ it follows that $\lim P_m y = P_\infty y$, with $P_\infty y = (l_1(y) \ l_2(y) \ \dots)$. Thus $P_m^* P_m y \rightarrow A y$ with $A = P_\infty^* P_\infty$ and A is injective because of the completeness condition.

4.1.2. Proof of Proposition 2.5. Since smooth functions are dense in L^2 , it suffices to consider the case where y is smooth. We have that $P_m^* P_m = P_m^+ P_m = P_{\mathcal{N}(P_m)^\perp}$ and $\mathcal{N}(P_m)^\perp := \{\sum_{j=1}^m \alpha_j \Lambda_j^{(m)}\}$ is the set of all functions constant on a homogeneous grid with m elements. Since the set of all functions constant on a homogeneous grid is dense in the set of smooth functions, the claim follows.

4.1.3. Proof of Proposition 2.6. As above w.l.o.g. y is assumed to be smooth. We denote by $A_m \in \mathbb{R}^{m \times m}$ the matrix representing $P_m : \mathcal{N}(P_m)^\perp \rightarrow \mathbb{R}^m$ with respect to the bases $(\eta_j^{(m)})_{j=1, \dots, m} \subset \mathcal{N}(P_m)^\perp$ and $(e_j)_{j=1, \dots, m} \subset \mathbb{R}^m$, where the latter is the canonical basis of \mathbb{R}^m . So

$$P_m^* P_m \eta_j^{(m)} = \sum_{i=1}^m (A_m^* A_m)_{ij} \eta_i^{(m)},$$

and

$$(A_m)_{ij} = \left(P_m \eta_i^{(m)}, e_j \right)_{\mathbb{R}^m} = l_j^{(m)}(\eta_i^{(m)}) = (\eta_j^{(m)}, \eta_i^{(m)})_y$$

with

$$(\eta_j^{(m)}, \eta_i^{(m)}) = \begin{cases} 2/3 & , i = j \\ 1/3 & , |i - j| = 1, \min(i, j) = 1 \text{ or } \max(i, j) = m \\ 1/6 & , |i - j| = 1, \min(i, j) > 1 \text{ and } \max(i, j) < m \\ 0 & , \text{else.} \end{cases}$$

So it is

$$\|P_m\| \leq \sqrt{\|P_m\|_1 \|P_m\|_\infty} = \max_{j=1, \dots, m} \sum_{i=1}^m |(A_m)_{ij}| = \frac{7}{6},$$

where $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are the spectral, the maximum absolute column and row norm respectively, and

$$P_m^* P_m \eta_j^{(m)} = \frac{\eta_{j-2}^{(m)}}{36} + \frac{2\eta_{j-1}^{(m)}}{9} + \frac{\eta_j^{(m)}}{2} + \frac{2\eta_{j+1}^{(m)}}{9} + \frac{\eta_{j+2}^{(m)}}{36},$$

for $j = 4, \dots, m-3$. Denote by $y_m = \sum_{j=1}^m y\left(\frac{j-1}{m-1}\right) \eta_j^{(m)} \sqrt{\frac{3}{2(m-1)}}$ the interpolating spline of y , then

$$\begin{aligned} & \|y - P_m^* P_m y\| \\ & \leq \|y_m - P_m^* P_m y_m\| + \|(I - P_m^* P_m)(y - y_m)\| \\ & \leq \left\| \sum_{j=1}^m y\left(\frac{j-1}{m-1}\right) \sqrt{\frac{3}{2(m-1)}} (I_m - P_m^* P_m) \eta_j^{(m)} \right\| + 2\|y - y_m\| \\ & \leq 2\|y - y_m\| + 6 \sup_t |y(t)| \sqrt{\frac{3}{2(m-1)}} \left(1 + \frac{7^2}{6^2}\right) + \\ & \left\| \sum_{j=4}^{m-3} \left(\frac{y\left(\frac{j}{m-1}\right)}{2} - \frac{2y\left(\frac{j+1}{m-1}\right)}{9} - \frac{2y\left(\frac{j-1}{m-1}\right)}{9} - \frac{y\left(\frac{j+2}{m-1}\right)}{36} - \frac{y\left(\frac{j-2}{m-1}\right)}{36} \right) \frac{\sqrt{3}\eta_j^{(m)}}{\sqrt{2(m-1)}} \right\| \\ & \leq 2\|y - y_m\| + 30 \sup_t |y(t)| \frac{1}{\sqrt{m-1}} \\ & + \sup_{j \leq m} \left| \frac{y\left(\frac{j}{m-1}\right)}{2} - \frac{2y\left(\frac{j+1}{m-1}\right)}{9} - \frac{2y\left(\frac{j-1}{m-1}\right)}{9} - \frac{y\left(\frac{j+2}{m-1}\right)}{36} - \frac{y\left(\frac{j-2}{m-1}\right)}{36} \right| \\ & * \left\| \sum_{j=4}^{m-3} \frac{\sqrt{3}\eta_j^{(m)}}{\sqrt{2(m-1)}} \right\| \\ & \leq 2\|y - y_m\| + 30 \sup_{t \in (0,1)} |y(t)| \frac{1}{\sqrt{m-1}} + \sup_{t \in (0,1)} |y'(t)| \frac{3}{m} \rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$.

4.1.4. Proof of Theorem 2.8. We will need the following proposition for the convergence proofs.

PROPOSITION 4.2. *Assume that Assumption 2.3 is fulfilled. Then, $P_{\mathcal{N}(P_m K)} x \rightarrow 0$ as $m \rightarrow \infty$, for all $x \in \mathcal{X}$.*

Proof. We assume w.l.o.g. that $x_m := P_{\mathcal{N}(P_m K)} x \rightarrow z \in \mathcal{X}$ for $m \rightarrow \infty$ (weakly). Then $\lim_{m \rightarrow \infty} K x_m = K z$. Thus

$$\begin{aligned} \|AKz\| &= \limsup_{m \rightarrow \infty} \|P_m^* P_m Kz\| = \limsup_{m \rightarrow \infty} \|P_m^* P_m (Kz - Kx_m)\| \\ &\leq \limsup_{m \rightarrow \infty} \|P_m\|^2 \|Kz - Kx_m\| = 0, \end{aligned}$$

so $AKz = 0$ hence by injectivity $z = 0$. In particular, $(P_{\mathcal{N}(P_m K)} v_i, v_i) \rightarrow (0, v_i) = 0$ for $m \rightarrow \infty$ and $i \in \mathbb{N}$ (set $x = v_i$ the i -th singular vector of K), so

$$1 \geq \|P_{\mathcal{N}(P_m K)} v_i - v_i\|^2 = \|P_{\mathcal{N}(P_m K)} v_i\|^2 - 2(P_{\mathcal{N}(P_m K)} v_i, v_i) + 1$$

and therefore

$$\limsup_{m \rightarrow \infty} \|P_{\mathcal{N}(P_m K)} v_i\| = 0.$$

Finally, by injectivity of K , for $\varepsilon > 0$ there is a $M \in \mathbb{N}$ with $\sum_{j=M+1}^{\infty} (x, v_j)^2 \leq \varepsilon$, so

$$\limsup_{m \rightarrow \infty} \|P_{\mathcal{N}(P_m K)} x\|^2 \leq \sum_{j=1}^M (x, v_j)^2 \limsup_{m \rightarrow \infty} \|P_{\mathcal{N}(P_m K)} v_j\|^2 + \varepsilon = \varepsilon$$

and the claim follows with $\varepsilon \rightarrow 0$. \square

We come to the main proof and split

$$\begin{aligned} & \mathbb{E} \left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y} \right\|^2 \\ & \leq \left\| K^+ \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\|^2 + \mathbb{E} \left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} \bar{Y}_n^{(m)} \right\|^2 \\ & \leq \left\| K^+ \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\|^2 + \left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \right\|^2 \mathbb{E} \left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\|^2 \end{aligned}$$

and because of independence,

$$\mathbb{E} \left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\|^2 = \mathbb{E} \sum_{j=1}^m \left(\frac{1}{n} \sum_{i=1}^n \delta_{ij}^{(m)} \right)^2 = \frac{1}{n} \sum_{j=1}^m \mathbb{E} \delta_{1j}^{(m)2} \leq \frac{m}{n} C_d = (\delta_{m,n}^{est})^2 C_d.$$

Assumption 2.1 implies, that

$$(13) \quad \|R_\alpha\| \leq \sqrt{C_R C_F / \alpha},$$

see f.e. [8] or Proposition 1 of [13]. Therefore it follows that

$$(14) \quad \left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \right\|^2 \mathbb{E} \delta_m^{meas2} \leq \left(\left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \right\| \delta_{m,n}^{est} \right)^2 C_d \leq C_d C_R C_F \frac{\delta_{m,n}^{est2}}{\alpha(\delta_{m,n}^{est})} \rightarrow 0$$

for $m, n \rightarrow \infty, m/n \rightarrow 0$. Now

$$\begin{aligned} & \left\| K^+ \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\| \\ & \leq \left\| K^+ \hat{y} - (P_m K)^+ P_m \hat{y} \right\| + \left\| (P_m K)^+ P_m \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\| \\ & = \left\| K^+ K \hat{x} - (P_m K)^+ P_m K \hat{x} \right\| + \left\| (P_m K)^+ P_m \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\| \\ & = \left\| \hat{x} - P_{\mathcal{N}(P_m K)^\perp} \hat{x} \right\| + \left\| (P_m K)^+ P_m \hat{y} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m \hat{y} \right\| \end{aligned}$$

and

$$(15) \quad \lim_{m \rightarrow \infty} \left\| \hat{x} - P_{\mathcal{N}(P_m K)^\perp} \hat{x} \right\| = \lim_{m \rightarrow \infty} \left\| P_{\mathcal{N}(P_m K)} \hat{x} \right\| = 0$$

by Proposition 4.2. Finally, for $\varepsilon > 0$, by Lemma 4.1.2 there is a $M \in \mathbb{N}$ such that $\sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2 \leq \varepsilon$ for m large enough and therefore

$$\begin{aligned} & \left\| (P_m K)^+ P_m K \hat{x} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m K \hat{x} \right\|^2 \\ &= \sum_{j=1}^m \left| 1 - F_{\alpha(\delta_{m,n}^{est})}(\sigma_j^{(m)}) \sigma_j^{(m)} \right|^2 (\hat{x}, v_j^{(m)})^2 \\ &\leq \sum_{j=1}^M \left| 1 - F_{\alpha(\delta_{m,n}^{est})}(\sigma_j^{(m)}) \sigma_j^{(m)} \right|^2 (\hat{x}, v_j^{(m)})^2 + \sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2 \\ &\leq \|\hat{x}\|^2 \sup_{j=1, \dots, M} \left| 1 - F_{\alpha(\delta_{m,n}^{est})}(\sigma_j^{(m)}) \sigma_j^{(m)} \right|^2 + \varepsilon. \end{aligned}$$

By Lemma 4.1.1, (9) and since $\alpha(\delta_{m,n}^{est}) \rightarrow 0$ for $m, n \rightarrow \infty, m/n \rightarrow 0$,

$$\sup_{j=1, \dots, M} \left| 1 - F_{\alpha(\delta_{m,n}^{est})}(\sigma_j^{(m)}) \sigma_j^{(m)} \right| \leq \sup_{\frac{\sigma_{M/2}^{(\infty)2}}{2} \leq \lambda \leq \|K\|^2} \left| 1 - F_{\alpha(\delta_{m,n}^{est})}(\lambda) \lambda \right| \leq \frac{\sqrt{\varepsilon}}{\|\hat{x}\|}$$

for all m, n sufficiently large and m/n sufficiently small. Thus with $\varepsilon \rightarrow 0$ it follows that

$$\lim_{\substack{m, n \rightarrow \infty \\ m/n \rightarrow 0}} \left\| (P_m K)^+ P_m K \hat{x} - R_{\alpha(\delta_{m,n}^{est})}^{(m)} P_m K \hat{x} \right\| = 0,$$

which concludes the proof together with (14) and (15).

4.1.5. Proof of Theorem 2.11. By the nature of white noise we cannot expect the error to concentrate along a certain direction, in contrast to [13]. However, the independence between the measurement channels implies, that its amplitude is highly concentrated. First, the following Proposition affirms that we are estimating the variance correctly.

PROPOSITION 4.3. *Assume that the error fulfills Assumption 2.9. Then for the sample variance*

$$s_{m,n}^2 = \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2$$

there holds

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\sup_{n \geq 2} \left| s_{m,n}^2 - \mathbb{E} \delta_{11}^{(m)2} \right| \geq \varepsilon \mathbb{E} \delta_{11}^{(m)2} \right) = 0$$

for all $\varepsilon > 0$.

Proof. As a sum of m reversed martingales, $\left(s_{m,-n}^2 - \mathbb{E} \delta_{11}^{(m)2} \right)_{n \leq -2}$ is a reversed martingale adapted to the filtration

$$\mathcal{F}_{-n} = \sigma \left(\sum_{i=1}^n (\delta_{i1}^{(m)} - \overline{\delta_{i1}^{(m)}})^2, \dots, \sum_{i=1}^n (\delta_{im}^{(m)} - \overline{\delta_{im}^{(m)}})^2 \right), n \geq 2.$$

Under Assumption 2.9.2, by the Kolmogorov-Doob-inequalities there holds

$$\mathbb{P}\left(\sup_{n \geq 2} \left| s_{m,n}^2 - \mathbb{E}\delta_{11}^{(m)2} \right| \geq \varepsilon \mathbb{E}\delta_{11}^{(m)2}\right) \leq \frac{\mathbb{E}\left| s_{m,2}^2 - \mathbb{E}\delta_{11}^{(m)2} \right|^p}{\left(\varepsilon \mathbb{E}\delta_{11}^{(m)2}\right)^p}.$$

By Marcinkiewicz-Zygmund inequality [10] there exists C_p such that

$$\begin{aligned} \mathbb{E}\left| s_{m,2}^2 - \mathbb{E}\delta_{11}^{(m)2} \right|^p &= \mathbb{E}\left| \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^2 (\delta_{ij}^{(m)} - \overline{\delta_{ij}^{(m)}})^2 - \mathbb{E}\delta_{11}^{(m)2} \right|^p \\ &\leq \frac{C_p}{m^{p-1}} \mathbb{E}\left| \sum_{i=1}^2 (\delta_{i1}^{(m)} - \overline{\delta_{i1}^{(m)}})^2 - \mathbb{E}\delta_{11}^{(m)2} \right|^p \\ &\leq \frac{2^{p-1}(4^p + 1)C_p}{m^{p-1}} \mathbb{E}|\delta_{11}^{(m)}|^{2p}, \end{aligned}$$

so

$$\mathbb{P}\left(\sup_{n \geq 2} \left| s_{m,n}^2 - \mathbb{E}\delta_{11}^{(m)2} \right| \geq \varepsilon \mathbb{E}\delta_{11}^{(m)2}\right) \leq \frac{\mathbb{E}\left| s_{m,2}^2 - \mathbb{E}\delta_{11}^{(m)2} \right|^p}{\left(\mathbb{E}\delta_{11}^{(m)2}\right)^p} \leq \frac{2^{p-1}(4^p + 1)C_p C_d}{\varepsilon^p m^{p-1}} \rightarrow 0$$

as $m \rightarrow \infty$. Under Assumption 2.9.1, by the Kolmogorov-Doob-inequality,

$$\mathbb{P}\left(\sup_{n \geq 2} \left| s_{m,n}^2 - \mathbb{E}\delta_{11}^{(m)2} \right| \geq \varepsilon \mathbb{E}\delta_{11}^{(m)2}\right) \leq \frac{\mathbb{E}\left| s_{m,2}^2 - \mathbb{E}\delta_{11}^{(1)2} \right|^2}{\varepsilon \mathbb{E}\delta_{11}^{(1)2}}.$$

It is

$$s_{m,2}^2 - \mathbb{E}\delta_{11}^{(1)2} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^2 \left(\delta_{ij}^{(m)} - \overline{\delta_{ij}^{(m)}} \right)^2 - \mathbb{E}\delta_{11}^{(1)2} =: \frac{1}{m} \sum_{j=1}^m X_j^{(m)}$$

with $X_j^{(m)}, j = 1, \dots, m, m \in \mathbb{N}$ are i.i.d and $\mathbb{E}X_j^{(m)} = 0, \mathbb{E}|X_j^{(m)}| < \infty$. To finish the proof we need to show that $\mathbb{E}|\sum_j X_m/m| \rightarrow 0$ as $m \rightarrow \infty$. Let $\varepsilon' > 0$. By dominated convergence and integrability of $X_j^{(m)}$, there is $M > 0$ large enough such that for $Y_j^{(m)} := X_j^{(m)} \chi_{\{|X_j^{(m)}| \leq M\}}$ and $Z_j^{(m)} := X_j^{(m)} \chi_{\{|X_j^{(m)}| > M\}}$ it holds that

$\mathbb{E}|Z_1^{(1)}| \leq \varepsilon$. So, since $X_j^{(m)}$ are i.i.d,

$$\begin{aligned} \mathbb{E} \left| \sum_{j=1}^m X_j^{(m)} \right| &\leq \mathbb{E} \left| \sum_{j=1}^m Y_j^{(m)} - \mathbb{E}Y_j^{(m)} \right| + \mathbb{E} \left| \sum_{j=1}^m Z_j^{(m)} - \mathbb{E}Z_j^{(m)} \right| \\ &\leq \sqrt{\mathbb{E} \left| \sum_{j=1}^m Y_j^{(m)} - \mathbb{E}Y_j^{(m)} \right|^2} + \sum_{j=1}^m \mathbb{E} |Z_j^{(m)} - \mathbb{E}Z_j^{(m)}| \\ &\leq \sqrt{m\mathbb{E} |Y_1^{(1)} - \mathbb{E}Y_1^{(1)}|^2} + 2m\mathbb{E}|Z_1^{(1)}| \leq \sqrt{m2M\mathbb{E}|X_1^{(1)}|} + 2m\varepsilon, \end{aligned}$$

thus $\mathbb{E} \left| \sum_{j=1}^m X_j^{(m)} / m \right| \leq 3\varepsilon$ for m large enough. \square

Now we need the following Lemma.

LEMMA 4.4. *Assume that the error model is accordingly to Assumption 2.9. Then there holds*

$$\lim_{m,n \rightarrow \infty} \mathbb{P} \left(\left| \frac{\left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\| - \delta_{m,n}^{est}}{\delta_{m,n}^{est}} \right| \geq \varepsilon \right) = 0.$$

Proof. It is

$$\begin{aligned} &\frac{\left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\| - \delta_{m,n}^{est}}{\delta_{m,n}^{est}} \\ &= \sqrt{\frac{\mathbb{E}\delta_{11}^{(m)2}}{s_{m,n}^2}} \left(\frac{\left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\| - \sqrt{m\mathbb{E}\delta_{11}^{(m)2}/n}}{\sqrt{m\mathbb{E}\delta_{11}^{(m)2}/n}} + 1 - \sqrt{\frac{s_{m,n}^2}{\mathbb{E}\delta_{11}^{(m)2}}} \right). \end{aligned}$$

Thus by Proposition 4.3 it suffices to show that

$$\lim_{m,n \rightarrow \infty} \mathbb{P} \left(\left| \frac{\left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\|^2 - \frac{m}{n} \mathbb{E}\delta_{11}^{(m)2}}{\frac{m}{n} \mathbb{E}\delta_{11}^{(m)2}} \right| \geq \varepsilon \right) = 0.$$

Let us first assume that Assumption 2.9.1 holds true. Then, by Markov's inequality

$$\begin{aligned} \mathbb{P} \left(\left| \frac{\left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\|^2 - \frac{m}{n} \mathbb{E}\delta_{11}^{(m)2}}{\frac{m}{n} \mathbb{E}\delta_{11}^{(m)2}} \right| \geq \varepsilon \right) &\leq \frac{\mathbb{E} \left| \left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\|^2 - \frac{m}{n} \mathbb{E}\delta_{11}^{(m)2} \right|}{\varepsilon \frac{m}{n} \mathbb{E}\delta_{11}^{(m)2}} \\ &= \frac{1}{m\varepsilon} \mathbb{E} \left| \sum_{j=1}^m \left(\left(\frac{\sum_{i=1}^n \delta_{ij}^{(m)}}{\sqrt{n\mathbb{E}\delta_{11}^{(m)2}}} \right)^2 - 1 \right) \right|. \end{aligned}$$

Now with

$$X_j^{(m)} := \left(\frac{\sum_{i=1}^n \delta_{ij}^{(m)}}{\sqrt{n\mathbb{E}\delta_{11}^{(m)^2}}} \right)^2 - 1,$$

it holds that $(X_j^{(m)})_{j=1, j=1, \dots, m, m \in \mathbb{N}}$ are i.i.d and $\mathbb{E}X_j^{(m)} = 0, \mathbb{E}|X_j^{(m)}| = 2 < \infty$. The proof can then be finished following the arguments at the end of the proof of Proposition 4.3.

Now assume that Assumption 2.9.2 holds true. Then, by Markov's inequality,

$$\mathbb{P} \left(\left| \frac{\|\bar{Y}_n^{(m)} - P_m \hat{y}\|^2 - \frac{m}{n} \mathbb{E}\delta_{11}^{(m)^2}}{\frac{m}{n} \mathbb{E}\delta_{11}^{(m)^2}} \right| \geq \varepsilon \right) \leq \frac{\mathbb{E} \left| \frac{n}{m\mathbb{E}\delta_{11}^{(m)^2}} \|\bar{Y}_n^{(m)} - P_m \hat{y}\|^2 - 1 \right|^p}{\varepsilon^p}$$

and using further twice the Marcinkiewicz-Zygmund inequality, one obtains

$$\begin{aligned} & \mathbb{E} \left| \frac{n}{m\mathbb{E}\delta_{11}^{(m)^2}} \|\bar{Y}_n^{(m)} - P_m \hat{y}\|^2 - 1 \right|^p \\ &= \frac{1}{m^p} \mathbb{E} \left| \sum_{j=1}^m \left(\left(\frac{\sum_{i=1}^n \delta_{ij}^{(m)}}{\sqrt{n\mathbb{E}\delta_{11}^{(m)^2}}} \right)^2 - 1 \right) \right|^p \\ &\leq \frac{B_p m^{\max(1, p/2)}}{m^p} \mathbb{E} \left| \left(\sum_{i=1}^n \delta_{i1}^{(m)} / \sqrt{n\mathbb{E}\delta_{11}^{(m)^2}} \right)^2 - 1 \right|^p \\ &\leq \frac{2^{p-1} B_p}{m^{\min(p-1, p/2)}} \left(\mathbb{E} \left| \sum_{i=1}^n \delta_{i1}^{(m)} / \sqrt{n\mathbb{E}\delta_{11}^{(m)^2}} \right|^{2p} + 1^p \right) \\ &\leq \frac{2^{p-1} B_p}{m^{\min(p-1, p/2)}} \left(B_{2p} \mathbb{E} \left| \delta_{11}^{(m)} \right|^{2p} / \left(\mathbb{E}\delta_{11}^{(m)^2} \right)^p + 1 \right) \leq \frac{C}{m^{\min(p-1, p/2)}} \rightarrow 0 \quad \square \end{aligned}$$

as $m \rightarrow \infty$, where we have used independence and $\mathbb{E} \left(\sum_{i=1}^n \delta_{ij}^{(m)} / \sqrt{n\mathbb{E}\delta_{11}^{(m)^2}} \right)^2 = 1$ in the second step.

Before we will start with the main proof, we need one last proposition.

PROPOSITION 4.5. *For all $\varepsilon > 0$, there are $m_0 \in \mathbb{N}$ and $\alpha_0 > 0$ such that*

$$\lim_{m \rightarrow \infty} \left\| P_m K R_\alpha^{(m)} P_m K \hat{x} - P_m K \hat{x} \right\| / \sqrt{\alpha} \leq \varepsilon$$

for all $m \geq m_0$ and $\alpha \leq \alpha_0$.

Proof. Lemma 4.1.2 guarantees the existence of $M \in \mathbb{N}$, such that

$$C_1^2 \sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2 \leq \varepsilon/2$$

for m sufficiently large. Then

$$\begin{aligned}
 & \left\| (P_m K R_\alpha^{(m)} - Id) P_m K \hat{x} \right\|^2 / \alpha = \sum_{j=1}^m \left(F_\alpha(\sigma_j^{(m)}) \sigma_j^{(m)^2} - 1 \right)^2 \frac{\sigma_j^{(m)^2}}{\alpha} (\hat{x}, v_j^{(m)})^2 \\
 & \leq \left(\sup_{\lambda > 0} \lambda^{\frac{\nu_0}{2}} |F_\alpha(\lambda) \lambda - 1| \right)^2 \|\hat{x}\|^2 \sum_{j=1}^M \frac{\sigma_j^{(m)^{2(1-\nu_0)}}}{\alpha} \\
 & \quad + \left(\sup_{\lambda > 0} \lambda^{\frac{1}{2}} |F_\alpha(\lambda) \lambda - 1| \right)^2 \frac{\sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2}{\alpha} \\
 & \leq C_{\nu_0}^2 M \sigma_M^{(m)^{2(1-\nu_0)}} \|\hat{x}\|^2 \alpha^{\nu_0-1} + C_1^2 \sum_{l=M+1}^m (\hat{x}, v_j^{(m)})^2 \\
 & \leq 2C_{\nu_0}^2 M \sigma_M^{(\infty)^{2(1-\nu_0)}} \|\hat{x}\|^2 \alpha^{\nu_0-1} + \varepsilon/2 \leq \varepsilon
 \end{aligned}$$

for m sufficiently large and α sufficiently small, where we have used that the qualification of $(F_\alpha)_{\alpha>0}$ is bigger than one in the third and Lemma 4.1.1 in the fourth step. \square

We start with the main proof. We define

$$\Omega_{m,n} := \left\{ \left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\| \leq \frac{\tau + C_0}{2C_0} \delta_{m,n}^{est}, \delta_{m,n}^{est} \leq c\varepsilon \right\},$$

with $c \leq \frac{1}{2} \max \left\{ \frac{C_0 + 3\tau}{\sigma_M^{(\infty)^2}}, \frac{(\tau + C_0) \sqrt{C_R C_F}}{\sqrt{\varepsilon'}} \right\}^{-1}$, where ε' is given below.

By Proposition 4.2,

$$\left\| (P_m K)^+ P_m \hat{y} - K^+ \hat{y} \right\| = \left\| P_{\mathcal{N}(P_m K)} \hat{x} \right\| \leq \varepsilon$$

for m large enough, and by Lemma 4.1.2,

$$\begin{aligned}
 & \left\| R_{\alpha_{m,n}}^{(m)} P_m \hat{y} - K^+ \hat{y} \right\|^2 \\
 & \leq \sum_{j=1}^M \left| F_{\alpha_{m,n}}(\sigma_j^{(m)}) \sigma_j^{(m)^2} - 1 \right|^2 (\hat{x}, v_j^{(m)})^2 + \sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2 \\
 & \leq \frac{1}{\sigma_M^{(m)^2}} \sum_{j=1}^M \left| F_{\alpha_{m,n}}(\sigma_j^{(m)}) \sigma_j^{(m)^2} - 1 \right|^2 \sigma_j^{(m)^2} (\hat{x}, v_j^{(m)})^2 + \varepsilon/2 \\
 & = \frac{1}{\sigma_M^{(m)^2}} \left\| (P_m K R_{\alpha_{m,n}}^{(m)} - Id) P_m \hat{y} \right\|^2 + \varepsilon/2 \\
 & \leq \frac{1}{\sigma_M^{(m)^2}} \left(\left\| (P_m K R_{\alpha_{m,n}}^{(m)} - Id) \bar{Y}_n^{(m)} \right\| + \left\| (P_m K R_{\alpha_{m,n}}^{(m)} - Id) (P_m \hat{y} - \bar{Y}_n^{(m)}) \right\| \right) + \varepsilon/2
 \end{aligned}$$

for m sufficiently large. So Lemma 4.1.1 and the defining relation of the discrepancy principle and of $\Omega_{m,n}$ ensure that

$$\left\| R_{\alpha_{m,n}}^{(m)} P_m \hat{y} - K^+ \hat{y} \right\| \chi_{\Omega_{m,n}} \leq \frac{2}{\sigma_M^{(\infty)^2}} \left(\tau \delta_{m,n}^{est} + C_0 \frac{\tau + C_0}{2C_0} \delta_{m,n}^{est} \right) \chi_{\Omega_{m,n}} + \varepsilon/2 \leq \varepsilon$$

for m sufficiently large. Moreover,

$$\begin{aligned}
& \tau \delta_{m,n}^{est} \chi_{\Omega_{m,n}} \\
& \leq \left\| (P_m K R_{\alpha_{m,n}/q}^{(m)} - Id) \bar{Y}_n^{(m)} \right\| \chi_{\Omega_{m,n}} \\
& \leq \left\| (P_m K R_{\alpha_{m,n}/q} - Id) P_m \hat{y} \right\| + \left\| (P_m K R_{\alpha_{m,n}/q}^{(m)} - Id) (\bar{Y}_n^{(m)} - P_m \hat{y}) \right\| \chi_{\Omega_{m,n}} \\
& \leq \left\| (P_m K R_{\alpha_{m,n}/q} - Id) P_m \hat{y} \right\| + C_0 \frac{\tau + C_0}{2C_0} \delta_{m,n}^{est} \chi_{\Omega_{m,n}}, \\
& \implies \delta_{m,n}^{est} \chi_{\Omega_{m,n}} \leq \frac{2}{\tau - C_0} \left\| (P_m K R_{\alpha_{m,n}/q} - Id) P_m \hat{y} \right\|
\end{aligned}$$

Proposition 4.5 guarantees the existence of ε' such that for m large enough

$$\left\| P_m K R_{\alpha}^{(m)} P_m \hat{y} - P_m \hat{y} \right\| / \sqrt{\alpha} \leq \frac{(\tau - C_0) q C_0}{(\tau + C_0) \sqrt{C_R C_F}} \frac{\varepsilon}{2}$$

for all $\alpha \leq \varepsilon'/q$. So with (13),

$$\begin{aligned}
& \left\| R_{\alpha_{m,n}}^{(m)} (\bar{Y}_n^{(m)} - P_m \hat{y}) \right\| \chi_{\Omega_{m,n}} \\
& \leq \left\| R_{\alpha_{m,n}} \right\| \left\| \bar{Y}_n^{(m)} - P_m \hat{y} \right\| \chi_{\Omega_{m,n}} \leq \sqrt{\frac{C_R C_R}{\alpha_{m,n}}} \frac{\tau + C_0}{2C_0} \delta_{m,n}^{est} \chi_{\Omega_{m,n}} \\
& \leq \frac{(\tau + C_0) \sqrt{C_R C_F}}{2C_0} \left(\frac{\delta_{m,n}^{est}}{\sqrt{\alpha_{m,n}}} \chi_{\Omega_{m,n} \cap \{\alpha_{m,n} \leq \varepsilon'\}} + \frac{\delta_{m,n}^{est}}{\sqrt{\alpha_{m,n}}} \chi_{\Omega_{m,n} \cap \{\alpha_{m,n} \geq \varepsilon'\}} \right) \\
& \leq \frac{(\tau + C_0) \sqrt{C_R C_F}}{2C_0} \left(\frac{2}{(\tau - C_0) q} \frac{\left\| (P_m K R_{\alpha_{m,n}/q} - Id) P_m \hat{y} \right\|}{\sqrt{\alpha_{m,n}/q}} \chi_{\{\alpha_{m,n} \leq \varepsilon'\}} + \frac{\delta_{m,n}^{est}}{\sqrt{\varepsilon'}} \chi_{\Omega_{m,n}} \right) \\
& \leq \frac{(\tau + C_0) \sqrt{C_R C_F}}{2C_0} \left(\frac{2}{(\tau - C_0) q} \frac{(\tau - C_0) q C_0}{(\tau + C_0) \sqrt{C_R C_F}} \frac{\varepsilon}{2} + \frac{c\varepsilon}{\sqrt{\varepsilon'}} \right) \leq \varepsilon/2 + \varepsilon/2
\end{aligned}$$

for m large enough. Putting it all together,

$$\begin{aligned}
& \left\| R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y} \right\| \chi_{\Omega_{m,n}} \\
& \leq \left\| R_{\alpha_{m,n}}^{(m)} (\bar{Y}_n^{(m)} - P_m \hat{y}) \right\| \chi_{\Omega_{m,n}} + \left\| R_{\alpha_{m,n}}^{(m)} P_m \hat{y} - (P_m K)^+ P_m \hat{y} \right\| \chi_{\Omega_{m,n}} \\
& \quad + \left\| (P_m K)^+ P_m \hat{y} - K^+ \hat{y} \right\| \chi_{\Omega_{m,n}} \\
& \leq 3\varepsilon
\end{aligned}$$

for m sufficiently large, which together with $\lim_{m/n \rightarrow \infty} \mathbb{P}(\Omega_{m,n}) = 1$ finishes the proof.

4.2. Proofs for infinite-dimensional residuum. For the second approach (with infinite-dimensional residuum), we need to guarantee stable inversion of the discretisation operator P_m . Afterwards we will show strong concentration of the back projected measurements in \mathcal{Y} in order to use classical results from deterministic regularisation theory.

4.2.1. Proof of Proposition 3.2. It is $\kappa(P_m) = \kappa(P_m|_{\mathcal{N}(P_m)^\perp})$. We again denote by $A_m \in \mathbb{R}^{m \times m}$ the matrix representing $P_m : \mathcal{N}(P_m)^\perp \rightarrow \mathbb{R}^m$ with respect to the bases $(\eta_j^{(m)})_{j=1,\dots,m} \subset \mathcal{N}(P_m)^\perp$ and $(e_j)_{j=1,\dots,m} \subset \mathbb{R}^m$, where the latter is the canonical basis of \mathbb{R}^m . Thus

$$(A_m)_{ij} = \left(P_m \eta_i^{(m)}, e_j \right)_{\mathbb{R}^m} = l_j^{(m)}(\eta_i^{(m)}) = (\eta_j^{(m)}, \eta_i^{(m)})_{\mathcal{Y}}.$$

By assumption, we have that

$$\begin{aligned} \left\| \frac{A_m}{\|\eta_1^{(m)}\|^2} - I_m \right\| &\leq \sqrt{\left\| \frac{A_m}{\|\eta_1^{(m)}\|^2} - I_m \right\|_1 \left\| \frac{A_m}{\|\eta_1^{(m)}\|^2} - I_m \right\|_\infty} \\ &= \max_{j=1,\dots,m} \sum_{i \neq j} \frac{|(\eta_j^{(m)}, \eta_i^{(m)})|}{\|\eta_1^{(m)}\|^2} =: c < 1, \end{aligned}$$

where $I_m \in \mathbb{R}^{m \times m}$ is the identity and $\|\cdot\|, \|\cdot\|_1, \|\cdot\|_\infty$ are the spectral and the maximum absolute column or row norm. So by (2.3) in [22], it is

$$(16) \quad 1 - c \leq \sigma_j \left(\frac{A_m}{\|\eta_1^{(m)}\|^2} \right) \leq 1 + c,$$

for $j = 1, \dots, m$, where $\sigma_1(A), \dots, \sigma_m(A)$ denote the singular values of $A \in \mathbb{R}^{m \times m}$. This proves the proposition.

4.2.2. Proof of Proposition 3.3. The bounds c_m, C_m follow directly from Proposition 3.2. It remains to show that $\|\hat{y} - P_m^+ P_m \hat{y}\| \rightarrow 0$ as $m \rightarrow \infty$. It holds that $\mathcal{N}(P_1) \supseteq \mathcal{N}(P_2) \supseteq \dots$. In particular, there is an orthonormal basis $(w_i)_{i \in \mathbb{N}}$ such that $\mathcal{N}(P_m) = \text{span}(w_{m+1}, w_{m+2}, \dots)$. Thus, $\delta_m^{disc} = \|P_{\mathcal{N}(P_m)} y\| = \sqrt{\sum_{j=m+1}^{\infty} (y, w_j)^2} \rightarrow 0$ as $m \rightarrow \infty$.

4.2.3. Proof of Proposition 3.4. The bound for the discretisation error follows from

$$\|\hat{y} - P_m^+ P_m \hat{y}\|^2 = \sum_{j>m} (\hat{y}, u_j)^2 = \sum_{j>m} \sigma_j^{2+2\nu} (w, v_j)^2 \leq \sigma_{m+1}^{2(1+\nu)} \|w\|^2.$$

Since $(\eta_j^{(m)}, \eta_i^{(m)}) = (v_j, v_i)$ and $(v_j)_{j \in \mathbb{N}}$ is an orthonormal basis, the claim follows.

4.2.4. Proof of Proposition 3.5. The choice $c_m = C_m = 1$ follows from Proposition 3.2, since $(\eta_j^{(m)})_{j=1,\dots,m}$ are orthonormal for all $m \in \mathbb{N}$. Denote by $y_m = \sum_{j=1}^m \hat{y}((j-1)/m) \chi_{(\frac{j-1}{m}, \frac{j}{m})} \in \mathcal{R}(P_m^*) = \mathcal{N}(P_m)^\perp$ the piecewise constant interpolating spline of the continuously differentiable function \hat{y} . Then there holds

$$\begin{aligned}
\|\hat{y} - P_m^+ P_m \hat{y}\| &= \|\hat{y} - P_{\mathcal{N}(P_m)^\perp} \hat{y}\| \leq \|\hat{y} - y_m\| \leq \sqrt{\int_0^1 (\hat{y}(t) - y_m(t))^2 dt} \\
&= \sqrt{\sum_{j=1}^m \int_{\frac{j-1}{m}}^{\frac{j}{m}} \left(\hat{y}(t) - \hat{y}\left(\frac{j-1}{m}\right) \right)^2 dt} \\
&= \sqrt{\sum_{j=1}^m \int_{\frac{j-1}{m}}^{\frac{j}{m}} y'(\xi_t) \left(t - \frac{j-1}{m} \right)^2 dt} \leq \frac{\sup_{t' \in (0,1)} |\hat{y}'(t')|}{m},
\end{aligned}$$

with $\xi_t \in [\frac{j-1}{m}, \frac{j}{m}]$.

4.2.5. Proof of Proposition 3.6. It is

$$(\eta_j^{(m)}, \eta_i^{(m)}) = \begin{cases} 2/3 & , i = j \\ 1/3 & , |i - j| = 1, \min(i, j) = 1 \text{ or } \max(i, j) = m \\ 1/6 & , |i - j| = 1, \min(i, j) > 1 \text{ and } \max(i, j) < m \\ 0 & , \text{else} \end{cases}$$

Therefore

$$\sup_{m \in \mathbb{N}} \max_{j \leq m} \frac{\sum_{j \neq i} |(\eta_j^{(m)}, \eta_i^{(m)})|}{\|\eta_1^{(m)}\|^2} = \frac{1/2}{2/3} = \frac{3}{4},$$

so that the bounds c_m, C_m follow with Proposition 3.2. Let $y_m \in \mathcal{N}(P_m)^\perp$ be the interpolating spline of continuously differentiable \hat{y} . By the mean value theorem there are $\xi_t, \zeta_t \in [\frac{j-1}{m-1}, \frac{j}{m-1}]$ such that

$$\begin{aligned}
&\hat{y}(t) - y_m(t) \\
&= \hat{y}\left(\frac{j-1}{m-1}\right) + \hat{y}'(\xi_t) \left(t - \frac{j-1}{m-1} \right) \\
&\quad - \left(\hat{y}\left(\frac{j-1}{m-1}\right) + \left(\hat{y}\left(\frac{j}{m-1}\right) - \hat{y}\left(\frac{j-1}{m-1}\right) \right) ((m-1)t - (j-1)) \right) \\
&= (y'(\xi_t) - y'(\zeta_t)) \left(t - \frac{j-1}{m-1} \right)
\end{aligned}$$

for $t \in [\frac{j-1}{m-1}, \frac{j}{m-1}]$. Thus

$$\begin{aligned}
\|\hat{y} - P_m^+ P_m \hat{y}\| &\leq \|\hat{y} - y_m\| \leq \sqrt{\sum_{j=1}^m \int_{\frac{j-1}{m-1}}^{\frac{j}{m-1}} (\hat{y}'(\xi_t) - \hat{y}'(\zeta_t))^2 \left(t - \frac{j-1}{m-1} \right)^2 dt} \\
&\leq \frac{2\sqrt{m} \sup_{t \in (0,1)} |\hat{y}'(t)|}{(m-1)^{3/2}} \leq \frac{2^{5/2} \sup_{t' \in (0,1)} |\hat{y}'(t')|}{m}
\end{aligned}$$

If \hat{y} is twice continuously differentiable, then there are $\xi'_t, \zeta'_t \in (\frac{j-1}{m-1}, \frac{j}{m-1}]$ such that

$$\begin{aligned} |\hat{y}'(\xi_t) - \hat{y}'(\zeta_t)| &= \left| \hat{y}''(\xi'_t) \left(\xi_t - \frac{j-1}{m-1} \right) - \hat{y}''(\zeta'_t) \left(\zeta_t - \frac{j-1}{m-1} \right) \right| \\ &\leq \frac{2 \sup_{t' \in (0,1)} |\hat{y}''(t')|}{m-1} \end{aligned}$$

for $t \in [\frac{j-1}{m-1}, \frac{j}{m-1})$, so that

$$\begin{aligned} \|\hat{y} - P_m^+ P_m \hat{y}\| &\leq \|\hat{y} - y_m\| \leq \sqrt{\sum_{j=1}^m \int_{\frac{j-1}{m-1}}^{\frac{j}{m-1}} \left(\frac{2 \sup_{t' \in (0,1)} |\hat{y}''(t')|}{m-1} \right)^2 \left(t - \frac{j-1}{m-1} \right)^2 dt} \\ &\leq \frac{2\sqrt{m} \sup_{t' \in (0,1)} |\hat{y}''(t')|}{(m-1)^{5/2}} \leq \frac{2^{7/2} \sup_{t' \in (0,1)} |\hat{y}''(t')|}{m^2}. \end{aligned}$$

4.2.6. Proof of Theorem 3.7. We use the bias-variance decomposition

$$\begin{aligned} &\mathbb{E} \left\| R_{\alpha(\delta_m^{disc})} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\|^2 \\ &= \mathbb{E} \left\| R_{\alpha(\delta_m^{disc})} P_m^+ (\bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m \hat{y}) \right\|^2 + \left\| R_{\alpha(\delta_m^{disc})} P_m^+ P_m \hat{y} - K^+ \hat{y} \right\|^2 \\ &\leq \mathbb{E} \left\| R_{\alpha(\delta_m^{disc})} P_m^+ (\bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m \hat{y}) \right\|^2 + 2 \left\| R_{\alpha(\delta_m^{disc})} P_m^+ P_m \hat{y} - R_{\alpha(\delta_m^{disc})} \hat{y} \right\|^2 \\ &\quad + 2 \left\| R_{\alpha(\delta_m^{disc})} \hat{y} - K^+ \hat{y} \right\|^2 \\ &\leq \left\| R_{\alpha(\delta_m^{disc})} \right\|^2 \left(\mathbb{E} \left\| P_m^+ \right\|^2 \mathbb{E} \left\| P_m \hat{y} - \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\|^2 + 2 \left\| P_m^+ P_m \hat{y} - \hat{y} \right\|^2 \right) \\ &\quad + 2 \left\| R_{\alpha(\delta_m^{disc})} \hat{y} - K^+ \hat{y} \right\|^2 \\ &\leq \frac{C_R C_F}{\alpha(\delta_m^{disc})} \left(\frac{\mathbb{E} \delta_{11}^{(m)2} m}{c_m^2 n(m, \delta_m^{disc})} + 2 \delta_m^{disc2} \right) + 2 \left\| R_{\alpha(\delta_m^{disc})} \hat{y} - K^+ \hat{y} \right\|^2 \\ &\leq (C_R C_F (C_d + 2)) \frac{\delta_m^{disc2}}{\alpha(\delta_m^{disc})} + 2 \left\| R_{\alpha(\delta_m^{disc})} \hat{y} - K^+ \hat{y} \right\|^2 \rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$.

4.2.7. Proof of Theorem 3.9. The proof of Theorem 3.9 is more technical than the one of Theorem 2.11, due to correlations coming from the back projecting of the measurements and the data-dependent determination of the stopping index $n(m, \delta_m^{disc})$. However, under slightly stronger conditions we obtain a similar concentration property of the measurement error.

LEMMA 4.6. *Assume that the discretisation fulfills Assumption 2.3 and the error is accordingly to Assumption 2.9, with $p \geq 2$ in the case of Assumption 2.9.2. For $m \in \mathbb{N}$, $\delta_0, \delta > 0$ and the sample variance*

$$s_{m,n}^2 := \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{i=1}^n \left(Y_{ij}^{(m)} - \frac{1}{n} \sum_{l=1}^n Y_{lj}^{(m)} \right)^2,$$

consider the (random) choice

$$n(m, \delta) = \min \left\{ n' \geq 1 : \frac{ms_m^2(n')}{c_m^2 n'} \leq \delta^2 \right\}$$

with $\sigma_1^{(m)}, \dots, \sigma_m^{(m)}$ the singular values of P_m . Then for any $\varepsilon > 0$ there holds

$$\lim_{m \rightarrow \infty} \sup_{0 < \delta \leq \delta_0} \mathbb{P} \left(\left| \frac{\|P_m^+ \bar{Y}_{n(m, \delta)}^{(m)} - P_m^+ P_m \hat{y}\| - \delta_m}{\delta_m} \right| \geq \varepsilon \right) = 0$$

with $\bar{Y}_{n(m, \delta)}^{(m)} = \frac{1}{n(m, \delta)} \sum_{i=1}^{n(m, \delta)} \left(Y_{i1}^{(m)} \quad \dots \quad Y_{im}^{(m)} \right)^T$ and $\delta_m := \delta \sqrt{\sum_{j=1}^m \frac{c_m^2}{m \sigma_j^{m/2}}}$.

Proof. The auxiliary parameter δ_m has to be introduced due to the fact, that with the choice of $n(m, \delta)$ we are actually overestimating $\mathbb{E} \left\| P_m^+ \bar{Y}_{n(m, \delta)}^{(m)} - P_m^+ P_m \hat{y} \right\|^2$, since $c_m \leq \sigma_j^{(m)}$. We define

$$\mu_m^\delta := \frac{m \mathbb{E}[\delta_{11}^{(m)^2}]}{c_m^2 \delta^2}$$

$$I_\varepsilon(m, \delta) := [(1 - \varepsilon)\mu_m^\delta, (1 + \varepsilon)\mu_m^\delta].$$

$$\delta_{m,n}^{meas} := \|P_m^+ \bar{Y}_n^{(m)} - P_m^+ P_m \hat{y}\| = \sqrt{\sum_{j=1}^m \lambda_j^{(m)} \left(\sum_{l=1}^m \sum_{i=1}^n \frac{\delta_{ij}^{(m)}}{n} (u_j^{(m)}, e_l^{(m)}) \right)^2}$$

where $\lambda_j^{(m)} = \sigma_j^{(m)^{-2}}$ and $(u_j^{(m)})_{j \leq m}, (e_j^{(m)})_{j \leq m} \subset \mathbb{R}^m$ are the singular basis of P_m (fulfilling $P_m P_m^* u_j^{(m)} = \sigma_j^{(m)^2} u_j^{(m)}$) and the canonical basis of \mathbb{R}^m respectively. So

$$\mathbb{E} \delta_{m,n}^{meas^2} = \sum_{j=1}^m \lambda_j \mathbb{E} \left(\sum_{l=1}^m \sum_{i=1}^n \frac{\delta_{il}^{(m)}}{n} (u_j^{(m)}, e_l^{(m)}) \right)^2 = \frac{\mathbb{E} \delta_{11}^{(m)^2}}{n} \sum_{j=1}^m \lambda_j$$

and

$$\begin{aligned} \mathbb{P} \left(\left| \frac{\delta_{m,n(m, \delta)}^{meas^2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon \right) &\geq \mathbb{P} \left(\left| \frac{\delta_{m,n(m, \delta)}^{meas^2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon, n(m, \delta) \in I_{\varepsilon'} \right) \\ &\geq \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas^2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon, n(m, \delta) \in I_{\varepsilon'} \right) \\ &\geq 1 - \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas^2} - \delta_m^2}{\delta_m^2} \right| > \varepsilon \right) - \mathbb{P}(n(m, \delta) \notin I_{\varepsilon'}). \end{aligned}$$

Since

$$\left| \frac{\delta_{m,n}^{meas^2} - \delta_m^2}{\delta_m^2} \right| \leq \left(\left| \frac{\delta_{m,n}^{meas^2} - \mathbb{E}\delta_{m,n}^{meas^2}}{\mathbb{E}\delta_{m,n}^{meas^2}} \right| + \left| \frac{\mathbb{E}\delta_{m,n}^{meas^2} - \delta_m^2}{\mathbb{E}\delta_{m,n}^{meas^2}} \right| \right) \frac{\mathbb{E}\delta_{m,n}^{meas^2}}{\delta_m^2}$$

and

$$\sup_{n \in I_{\varepsilon'}} \left| \frac{\mathbb{E}\delta_{m,n}^{meas^2} - \delta_m^2}{\delta_m^2} \right| = \frac{\varepsilon'}{1 - \varepsilon'}, \quad \sup_{n \in I_{\varepsilon'}} \frac{\mathbb{E}\delta_{m,n}^{meas^2}}{\delta_m^2} = \frac{1}{1 - \varepsilon'},$$

we conclude that for $\varepsilon' = \frac{3}{16}\varepsilon \leq 1/4$

$$\begin{aligned} & \mathbb{P} \left(\left| \frac{\delta_{m,n(m,\delta)}^{meas}{}^2 - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon \right) \\ & \geq 1 - \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas^2} - \mathbb{E}\delta_{m,n}^{meas^2}}{\mathbb{E}\delta_{m,n}^{meas^2}} \right| > \varepsilon(1 - \varepsilon') - \frac{\varepsilon'}{1 - \varepsilon'} \right) - \mathbb{P}(n(m, \delta) \notin I_{\varepsilon'}) \\ (17) \quad & \geq 1 - \mathbb{P} \left(\sup_{n \in I_{\frac{3}{16}\varepsilon}} \left| \frac{\delta_{m,n}^{meas^2} - \mathbb{E}\delta_{m,n}^{meas^2}}{\mathbb{E}\delta_{m,n}^{meas^2}} \right| > \varepsilon/2 \right) - \mathbb{P}(n(m, \delta) \notin I_{\frac{3}{16}\varepsilon}). \end{aligned}$$

Thus it remains to show that the both terms with negative sign tend to zero.

PROPOSITION 4.7. *For every $\varepsilon > 0$ there holds*

$$\sup_{\delta_0 \geq \delta > 0} \mathbb{P}(n(m, \delta) \in I_{\varepsilon}(m, \delta)) \rightarrow 1$$

for $m \rightarrow \infty$.

Proof. For m large enough it is $\lfloor (1 + \varepsilon)\mu_m^\delta \rfloor \geq (1 + \varepsilon/2)\mu_m^\delta$ and

$$\begin{aligned} & \{n(m, \delta) \in I_{\varepsilon}(m, \delta)\} = \{|n(m, \delta) - \mu_m^\delta| \leq \varepsilon\mu_m^\delta\} \\ & \supseteq \left\{ \frac{ms_{m,n}^2}{c_m^2 n} > \delta^2, \forall n < (1 - \varepsilon)\mu_m^\delta \right\} \\ & \quad \cap \left\{ \frac{ms_{m,n}^2}{c_m^2 n} \leq \delta^2, \text{ for } n = \lfloor (1 + \varepsilon)\mu_m^\delta \rfloor \right\} \\ & = \left\{ ms_{n,m}^2 > \frac{n}{\mu_m^\delta}, \forall n < (1 - \varepsilon)\mu_m^\delta \right\} \cap \left\{ s_{n,m}^2 \leq \frac{n}{\mu_m^\delta}, \text{ for } n = \lfloor (1 + \varepsilon)\mu_m^\delta \rfloor \right\} \\ & \supseteq \left\{ |s_{n,m}^2 - \mathbb{E}[\delta_{11}^{(m)2}]| \leq \varepsilon/2 \mathbb{E}[\delta_{11}^{(m)2}], \forall n \geq 2 \right\}, \quad \square \end{aligned}$$

and the claim follows by Proposition 4.3.

For the first term in (17) we will need the following proposition.

PROPOSITION 4.8. *For (X_l) i.i.d, $l = 1, \dots, m$, with $\mathbb{E}X_l = 0$, $\mathbb{E}X_l^2 = 1$ and $\mathbb{E}X_l^4 < \infty$ and $(u_j)_{j \leq m}, (e_j)_{j \leq m} \subset \mathbb{R}^m$ orthonormal bases and $(\lambda_j)_{j \leq m} \in \mathbb{R}^+$, it holds that*

$$\mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 - 1 \right) \right|^2 \leq \max_{j \leq m} \lambda_j^2 (\mathbb{E} X_1^4 + 5)m.$$

Proof. By Jensen's inequality,

$$\begin{aligned} & \left(\mathbb{E} \left[\sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 - 1 \right) \right] \right)^2 \\ & \leq \mathbb{E} \left[\sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 - 1 \right) \right]^2 \\ & = \sum_{j, j'=1}^m \lambda_j \lambda_{j'} \left(\mathbb{E} \left[\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 \left(\sum_{l'=1}^m X_{l'}(u_{j'}, e_{l'}) \right)^2 \right] \right. \\ & \quad \left. - 2 \mathbb{E} \left[\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 \right] + 1 \right) \\ & = \sum_{j, j'=1}^m \lambda_j \lambda_{j'} \left(\sum_{l, l', l'', l'''=1}^m \mathbb{E} [X_l X_{l'} X_{l''} X_{l'''}] (u_j, e_l)(u_j, e_{l'}) (u_{j'}, e_{l''}) (u_{j'}, e_{l'''}) \right. \\ & \quad \left. + 2 (\mathbb{E} [X_1^2])^2 - 1 \right) \\ & = \sum_{j, j'=1}^m \lambda_j \lambda_{j'} \left(\mathbb{E} X_1^4 \sum_{l=1}^m (u_j, e_l)^2 (u_{j'}, e_l)^2 + (\mathbb{E} [X_1^2])^2 \sum_{\substack{l, l'=1 \\ l \neq l'}}^m (u_j, e_l)^2 (u_{j'}, e_{l'})^2 \right. \\ & \quad \left. + 2 (\mathbb{E} [X_1^2])^2 \sum_{\substack{l, l'=1 \\ l \neq l'}}^m (u_j, e_l)(u_j, e_{l'}) (u_{j'}, e_l)(u_{j'}, e_{l'}) - 1 \right). \end{aligned}$$

With

$$\sum_{\substack{l'=1 \\ l' \neq l}}^m (u_{j'}, e_{l'})^2 = 1 - (u_{j'}, e_l)^2$$

and

$$\sum_{\substack{l'=1 \\ l' \neq l}}^m (u_j, e_{l'}) (u_{j'}, e_{l'}) = (u_j, u_{j'}) - (u_j, e_l)(u_{j'}, e_l)$$

we further deduce that

$$\begin{aligned}
 & \left(\mathbb{E} \left[\left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m X_l(u_j, e_l) \right)^2 - 1 \right) \right| \right] \right)^2 \\
 &= \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \left(\mathbb{E} X_1^4 \sum_{l=1}^m (u_j, e_l)^2 (u_{j'}, e_l)^2 + \sum_{l=1}^m (u_j, e_l)^2 (1 - (u_{j'}, e_l)^2) \right. \\
 &\quad \left. + 2 \sum_{l=1}^m (u_j, e_l) (u_{j'}, e_l) ((u_j, u_{j'}) - (u_j, e_l) (u_{j'}, e_l)) - 1 \right) \\
 &= \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \left(\mathbb{E} X_1^4 \sum_{l=1}^m (u_j, e_l)^2 (u_{j'}, e_l)^2 + 1 - \sum_{l=1}^m (u_j, e_l)^2 (u_{j'}, e_l)^2 \right. \\
 &\quad \left. + 2 \left((u_j, u_{j'})^2 - \sum_l (u_j, e_l)^2 (u_{j'}, e_l)^2 \right) - 1 \right) \\
 &\leq \max_{j \leq m} \lambda_j^2 \left(\sum_{l=1}^m \sum_{j,j'=1}^m |\mathbb{E} X_1^4 - 3| (u_j, e_l)^2 (u_{j'}, e_l)^2 + 2 \sum_{j,j'=1}^m (u_j, u_{j'})^2 \right) \\
 &\leq \max_{j \leq m} \lambda_j^2 (\mathbb{E} X_1^4 + 5) m,
 \end{aligned}$$

□

Finally, it is

$$\begin{aligned}
 M_n^{(m)} &:= n \frac{\delta_{m,n}^{meas2} - \mathbb{E} \delta_{m,n}^{meas2}}{\mathbb{E} \delta_{m,n}^{meas2}} \\
 &= n \frac{\sum_{j=1}^m \lambda_j \left(\sum_{l=1}^m \sum_{i=1}^n \frac{\delta_{il}^{(m)}}{\sqrt{n}} (u_j^{(m)}, e_l^{(m)}) \right)^2 - \mathbb{E} \delta_{11}^{(m)2} \sum_{j=1}^m \lambda_j}{\mathbb{E} \delta_{11}^{(m)2} \sum_{j=1}^m \lambda_j} \\
 &= \frac{n}{\sum_{j'=1}^m \lambda_{j'}} \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m \sum_{i=1}^n \frac{\delta_{il}^{(m)}}{\sqrt{n \mathbb{E} \delta_{11}^{(m)2}}} (u_j^{(m)}, e_l^{(m)}) \right)^2 - 1 \right).
 \end{aligned}$$

It is easy to verify that $(M_n^{(m)})_{n \in \mathbb{N}}$ is a martingale adapted to the filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ generated by the measurement errors, $\mathcal{F}_n := \sigma(\delta_{ij}^{(m)}, i \leq n, j \leq m)$ for every fixed $m \in \mathbb{N}$. Now assume that Assumption 2.9.2 with $p \geq 2$ holds true. With $n_- := (1 + \frac{3}{16}\varepsilon)\mu_m^\delta$, $n_+ := (1 + \frac{3}{16}\varepsilon)\mu_m^\delta$, we obtain via the Kolmogorov-Doob-inequality

$$\begin{aligned}
 \mathbb{P} \left(\sup_{n \in I_{\frac{3}{16}\varepsilon}} \left| \frac{\delta_{m,n}^{meas2} - \mathbb{E} \delta_{m,n}^{meas2}}{\mathbb{E} \delta_{m,n}^{meas2}} \right| \geq \frac{\varepsilon}{2} \right) &= \mathbb{P} \left(n_- \sup_{n \in I_{\frac{3}{16}\varepsilon}} \left| \frac{\delta_{m,n}^{meas2} - \mathbb{E} \delta_{m,n}^{meas2}}{\mathbb{E} \delta_{m,n}^{meas2}} \right| \geq \frac{n_- \varepsilon}{2} \right) \\
 &\leq \mathbb{P} \left(\sup_{n \in I_{\frac{3}{16}\varepsilon}} |M_n^{(m)}| \geq \frac{n_- \varepsilon}{2} \right) \leq \frac{4 \mathbb{E} [M_{n_+}^{(m)2}]}{\varepsilon^2 n_-^2}.
 \end{aligned}$$

With $X_l := \sum_i \delta_{ij}^{(m)} / \sqrt{n \mathbb{E} \delta_{ij}^{(m)2}}$ Proposition 4.8 yields

$$\begin{aligned}
\frac{4\mathbb{E}\left[M_{n_+}^{(m)2}\right]}{n_-^2\varepsilon^2} &= \frac{4n_+^2}{n_-^2\varepsilon^2} \frac{\max_{j\leq m} \lambda_j^2 (\mathbb{E}X_1^4 + 5)m}{(\sum_j \lambda_j)^2} \\
&= \frac{4n_+^2}{\varepsilon^2 n_-^2} \frac{\max_{j\leq m} \sigma_j^{-4}}{\min_{j\leq m} \sigma_j^{-4}} \left(\frac{\mathbb{E}\delta_{11}^{(m)4}}{n_+ (\mathbb{E}\delta_{11}^{(m)2})^2} + 3\frac{n_+ - 1}{n_+} + 5 \right) \frac{1}{m} \\
&= \frac{n_+^2}{\varepsilon^2 n_-^2} \kappa(P_m)^4 \left(\frac{C_d}{n_+} + 3\frac{n_+ - 1}{n_+} + 5 \right) \frac{1}{m} \rightarrow 0
\end{aligned}$$

as $m \rightarrow \infty$. In the following we write u_j and e_j for $u_j^{(m)}$ and $e_j^{(m)}$. Under Assumption 2.9.1, the Kolmogorov-Doob-inequality yields

$$\mathbb{P} \left(\sup_{n \in I_{\frac{3}{16}\varepsilon}} \left| \frac{\delta_{m,n}^{meas2} - \mathbb{E}\delta_{m,n}^{meas2}}{\mathbb{E}\delta_{m,n}^{meas2}} \right| \geq \frac{\varepsilon}{2} \right) \leq \frac{\mathbb{E}|M_{n_+}^{(m)}|}{\varepsilon n_-}.$$

We set $S_m := \frac{M_{n_+}^{(m)}}{n_+} \sum_{j=1}^m \lambda_j$ and $Z_l^{(m)} := \sum_{i=1}^n \delta_{il}^{(m)} / \sqrt{n_+ \mathbb{E}\delta_{11}^{(m)2}}$ ($Z_l^{(m)}, j = 1, \dots, m, m \in \mathbb{N}$ are i.i.d.). For $K > 0$ we truncate

$$\begin{aligned}
V_l^{(m)} &:= Z_l^{(m)} \chi_{\{|Z_l^{(m)}| \leq K\}} - \mathbb{E} \left[Z_l^{(m)} \chi_{\{|Z_l^{(m)}| \leq K\}} \right] \\
W_l^{(m)} &:= Z_l^{(m)} \chi_{\{|Z_l^{(m)}| > K\}} - \mathbb{E} \left[Z_l^{(m)} \chi_{\{|Z_l^{(m)}| > K\}} \right].
\end{aligned}$$

Then $\mathbb{E}V_l^{(m)} = \mathbb{E}W_l^{(m)} = 0 = V_l^{(m)}W_l^{(m)}$ and therefore

$$\begin{aligned}
&\mathbb{E}|S_m| \\
&= \mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m Z_l^{(m)}(u_j, e_l) \right)^2 - 1 \right) \right| \\
&\leq \mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m V_l^{(m)}(u_j, e_l) \right)^2 - \mathbb{E} \left[V_1^{(1)2} \right] \right) \right| + \mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\sum_{l=1}^m W_l^{(m)}(u_j, e_l) \right)^2 \right| \\
&\quad + 2\mathbb{E} \left| \sum_{j=1}^m \lambda_j \sum_{\substack{l, l'=1 \\ l \neq l'}}^m V_l^{(m)} W_{l'}^{(m)}(u_j, e_l)(u_j, e_{l'}) \right| + \left| 1 - \mathbb{E} \left[V_1^{(1)2} \right] \right| \sum_{j=1}^m \lambda_j.
\end{aligned}$$

Since $\mathbb{E} \left[V_1^{(1)4} \right] < \infty$, by Proposition 4.8 above and Jensen's inequality,

$$\begin{aligned}
 & \mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m V_l^{(m)}(u_j, e_l) \right)^2 - \mathbb{E} [V_1^{(1)2}] \right) \right| \\
 & \leq \sqrt{\mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\left(\sum_{l=1}^m V_l^{(m)}(u_j, e_l) \right)^2 - \mathbb{E} [V_1^{(1)2}] \right) \right|^2} \\
 & \leq \|P_m^+\|^2 \sqrt{\mathbb{E} [V_1^{(1)4}] + 5\sqrt{m}}.
 \end{aligned}$$

For the second term,

$$\begin{aligned}
 \mathbb{E} \left| \sum_{j=1}^m \lambda_j \left(\sum_{l=1}^m W_l^{(m)}(u_j, e_l) \right)^2 \right| & \leq \mathbb{E} \left| \|P_m^+\|^2 \sum_{l,l'=1}^m W_l^{(m)} W_{l'}^{(m)} \sum_{j=1}^m (u_j, e_l)(u_j, e_{l'}) \right| \\
 & = \|P_m^+\|^2 \mathbb{E} \left| \sum_{l,l'=1}^m W_l^{(m)} W_{l'}^{(m)}(e_l, e_{l'}) \right| \\
 & = m \|P_m^+\|^2 \mathbb{E} [W_1^{(1)2}].
 \end{aligned}$$

For the third term we calculate the variance,

$$\begin{aligned}
& \left(\mathbb{E} \left| \sum_{j=1}^m \lambda_j \sum_{\substack{l,l'=1 \\ l \neq l'}}^m V_l^{(m)} W_{l'}^{(m)}(u_j, e_l)(u_j, e_{l'}) \right| \right)^2 \\
& \leq \mathbb{E} \left| \sum_{j=1}^m \lambda_j \sum_{\substack{l,l'=1 \\ l \neq l'}}^m V_l^{(m)} W_{l'}^{(m)}(u_j, e_l)(u_j, e_{l'}) \right|^2 \\
& \leq \mathbb{E} \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \sum_{\substack{l,l'=1 \\ l \neq l'}}^m \sum_{\substack{l'',l'''=1 \\ l'' \neq l'''}}^m V_l^{(m)} W_{l'}^{(m)} V_{l''}^{(m)} W_{l'''}^{(m)}(u_j, e_l)(u_j, e_{l'})(u_{j'}, e_{l''})(u_{j'}, e_{l'''}) \\
& = \mathbb{E} \left[V_1^{(1)2} \right] \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \sum_{\substack{l,l'=1 \\ l \neq l'}}^m (u_j, e_l)(u_j, e_{l'})(u_{j'}, e_l)(u_{j'}, e_{l'}) \\
& = \mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right] \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \sum_{l=1}^m (u_j, e_l)(u_{j'}, e_l) \left((u_j, u_{j'}) - (u_j, e_l)(u_{j'}, e_{l'}) \right) \\
& = \mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right] \sum_{j,j'=1}^m \lambda_j \lambda_{j'} \left((u_j, u_{j'})^2 - \sum_{l=1}^m (u_j, e_l)^2 (u_{j'}, e_l)^2 \right) \\
& \leq \mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right] \sum_{j,j'=1}^m \lambda_j \lambda_{j'} (u_j, u_{j'})^2 = \mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right] \sum_{j=1}^m \lambda_j^2 \\
& \leq \mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right] \|P_m^+\|^4 m.
\end{aligned}$$

Altogether,

$$\begin{aligned}
& \frac{2\mathbb{E}|M_{n_+}^{(m)}|}{\varepsilon n_-} \\
& \leq \frac{2n_+}{\varepsilon n_-} \frac{1}{\sum_j \lambda_j} \mathbb{E}|S_m| \\
& \leq \frac{2n_+}{\varepsilon n_-} \frac{1}{\sum_j \lambda_j} \left(\|P_m^+\|^2 \sqrt{\mathbb{E} \left[V_1^{(1)4} \right]} + 5\sqrt{m} + m \|P_m^+\|^2 \mathbb{E} \left[W_1^{(1)2} \right] \right. \\
& \quad \left. + \sqrt{\mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right]} \|P_m^+\|^2 \sqrt{m} |1 - \mathbb{E} V^2| m \|P_m^+\|^2 \right) \\
& \leq \frac{2n_+ \kappa(P_m)^2}{\varepsilon n_- \sqrt{m}} \left(\sqrt{\mathbb{E} \left[V_1^{(1)4} \right]} + 5 + \sqrt{\mathbb{E} \left[V_1^{(1)2} \right] \mathbb{E} \left[W_1^{(1)2} \right]} \right) \\
& \quad + \frac{2n_+ \kappa(P_m)^2}{\varepsilon n_-} \left(\mathbb{E} \left[W_1^{(1)2} \right] + \left| 1 - \mathbb{E} \left[V_1^{(1)2} \right] \right| \right). \quad \square
\end{aligned}$$

The claim follows with $\lim_{K \rightarrow \infty} \mathbb{E} \left[V_1^{(1)2} \right] = 1$, $\lim_{K \rightarrow \infty} \mathbb{E} \left[W_1^{(1)2} \right] = 0$ and $\sup_m \kappa(P_m)^2 < \infty$.

We come to the main proof

Proof. We set

$$\Omega_m := \left\{ \left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ P_m \hat{y} \right\| \leq \frac{\tau + C_0}{2C_0} \delta_m^{disc} \right\}.$$

Then,

$$(18) \quad \begin{aligned} \left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{y} \right\| \chi_{\Omega_m} &\leq \left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ P_m \hat{y} \right\| \chi_{\Omega_m} + \left\| P_m^+ P_m \hat{y} - \hat{y} \right\| \chi_{\Omega_m} \\ &\leq \frac{\tau + 3C_0}{2C_0} \delta_m^{disc}. \end{aligned}$$

By Algorithm 2 it is

$$\begin{aligned} &\alpha_m \\ &:= \left\{ q^k, k \in \mathbb{N}_0, \left\| K R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\| \leq 2\tau \delta_m^{disc} \right\} \\ &= \left\{ q^k, k \in \mathbb{N}_0, \left\| K R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\| \leq \frac{4\tau C_0}{\tau + 3C_0} \frac{\tau + 3C_0}{2C_0} \delta_m^{disc} \right\} \end{aligned}$$

and because of $\frac{4\tau C_0}{\tau + 3C_0} > C_0$, (18) and $\lim_{m \rightarrow \infty} \delta_m^{disc} = 0$, it follows that

$$\lim_{m \rightarrow \infty} \left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \chi_{\Omega_m} = 0$$

by Theorem 4.17 and Remark 4.18 from [8]. With the same reasoning it follows that there is a $L' \in \mathbb{R}$ such that

$$\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - K^+ \hat{y} \right\| \chi_{\Omega_m} \leq L' \rho^{\frac{1}{\nu+1}} \delta_m^{disc \frac{\nu}{\nu+1}},$$

□

if there are $0 < \nu \leq \nu_0 - 1$ and $w \in \mathcal{X}$ with $K^+ \hat{y} = (K^* K)^{\nu/2} w$ and $\|w\| \leq \rho$. Lemma 4.6 implies that $\lim_{m \rightarrow \infty} \mathbb{P}(\Omega_m) = 1$, which concludes the proof.

5. Numerical Demonstration. We provide numerical experiments to complement the theoretical analysis. Three model examples, i.e. **phillips** (mildly ill-posed, smooth), **gravity** (severely ill-posed, medium smooth) and **shaw** (severely ill-posed, non smooth), are taken from the open source MATLAB package Regutools [11]. The problems cover a variety of settings, e.g., different solution smoothness and degree of ill-posedness. These examples are discretisations of Fredholm/Volterra integral equations of the first kind, by means of either the Galerkin approximation with piecewise constant basis functions or quadrature rules. We approximate our infinite-dimensional K with one of the above examples with dimension $m_\infty \gg 1$. The number of measurements channels m is then always chosen such that $m \ll m_\infty$. In most of the examples we use discretisation by box functions as follows, compare to Lemma 2.5. With $k = m_\infty/m$ we set

$$P_m : \mathbb{R}^{m_\infty} \rightarrow \mathbb{R}^m$$

$$\begin{pmatrix} y_{(i-1)k+1} \\ \dots \\ y_{(i-1)k+k} \end{pmatrix} \mapsto \frac{1}{\sqrt{k}} (y_{(i-1)k+1} + \dots + y_{(i-1)k+k}) e_i$$

where $i = 1, \dots, m$ and e_1, \dots, e_m is the canonical basis of \mathbb{R}^m . In Subsection 5.3 we will also consider discretisation by hat functions to give an example with nonorthogonal discretisation. We chose a shifted generalised Pareto distribution for the distribution of the measurement error, e.g. $\delta_{ij}^{(m)} = Z_{ij}^{(m)} - EZ_{ij}^{(m)}$, where $Z_{ij}^{(m)}$ are i.i.d and follow a generalised Pareto distribution (`gprnd`(l, σ, θ, m, n) in Matlab, with $l = 1/3$, $\sigma = \sqrt{(1-l)^2(1-2l)\|\hat{y}\|}$ and $\theta = 0$). This distribution is highly non symmetric with a heavy tail. The above choices for the parameters imply that $\mathbb{E}\delta_{ij}^{(m)2} = \|\hat{y}\|$ and $\mathbb{E}|\delta_{ij}^{(m)}|^3 = \infty$. Thus the error fulfills Assumption 2.9.1 in all the examples. The parameter τ in the definition of the discrepancy principle is set to $\tau = 1.2$. All the statistical quantities are computed for 100 independent runs, and the results are presented as box plots.

5.1. Convergence of finite-dimensional residuum approach. First we visualise the convergence of the discrepancy principle with the finite-dimensional residuum approach, as stated in Corollary 1.1. We use discretisation by box functions as presented above and set $m_\infty = 4000$ and $m = 5, 10, 20$. For each m we plot in Figure 1 the resulting relative errors $\left\| R_{\alpha_m, n}^{(m)} \bar{Y}_n^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ for $n = 10, \dots, 10^9$ repetitions. For m fix, the relative errors first decrease steadily, and then saturate (at $\|\hat{x} - (P_m K)^+ P_m K \hat{x}\|$), as the number of repetitions n grows. The saturation level decreases rapidly while m grows, confirming the convergence of the approach. It is notable, that for all examples a fairly small number of measurement channels is sufficient to yield good approximations.

5.2. (Semi-)Convergence of infinite-dimensional residuum approach. Now we come to the discrepancy principle with the infinite-dimensional residuum approach, as stated in Corollary 1.2. Again we chose discretisation by box functions for the measurements with $m_\infty = 4000$ and this time we set $m = 20, 50, 100$. For each m we plot in the right column of Figure 1 the resulting relative errors $\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ for varying upper bound δ_m^{disc} from Assumption 3.1. More precisely we chose the latter in relation to the exact discretisation error $d_m := \|\hat{y} - P_m^+ P_m \hat{y}\|$. In particular we also consider $\delta_m^{disc} < d_m$ and we exhibit a semi-convergence. Strictly speaking, the last two choices ($d_m/2$ and $d_m/4$) for δ_m^{disc} violate Assumption 3.1 and we thus illustrate the sensitiveness to underestimation of the true discretisation error. It is notable that for the choice $\delta_m^{disc} = d_m/2$ (e.g. underestimation of the discretisation error by a factor 1/2) the relative errors are still decreasing. This is explained by the fact, that the estimation in (11) is quite coarse. Together with the choice $\tau = 1.2$ this yields, that it still holds that the true unknown error $\left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{y} \right\|$ fulfills $\left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{y} \right\| < 2\tau \delta_m^{disc}$. For the choice $\delta_m^{disc} = d_m/4$ the errors then diverge. The semi-convergence is in contrast to the saturation observed in the left column of Figure 1 and illustrates the fundamental difference, that for the finite-dimensional approach no quantitative knowledge of the discretisation error is required, while for the infinite-dimensional approach it is.

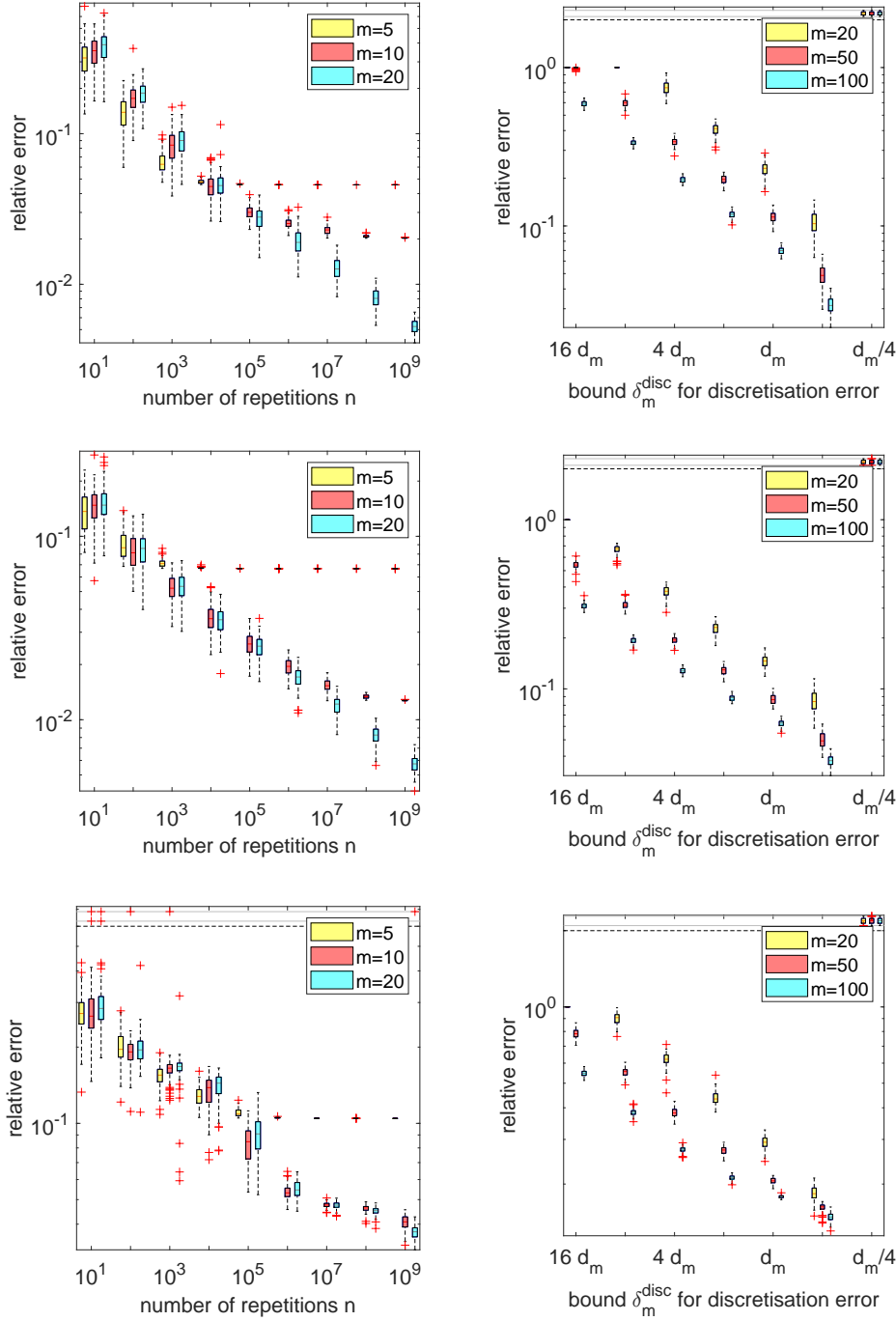


FIG. 1. Results of approach (2) and (6) with the discrepancy principle as implemented in Algorithm 1 (left column) or 2 (right column) respectively, for 'phillips' (first row), 'gravity' (second row) and 'shaw' (third row), visualised as boxplots for 100 independent runs. Left column: Relative errors $\|R_{\alpha_m, n}^{(m)} \bar{Y}_n^{(m)} - \hat{x}\| / \|\hat{x}\|$ against number of repetitions n for several numbers of measurement channels m . Right column: Relative errors $\|R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{\text{disc}})}^{(m)} - \hat{x}\| / \|\hat{x}\|$ against bound for the discretisation error δ_m^{disc} for several numbers of measurement channels m . δ_m^{disc} is chosen in relation to the exact discretisation error $d_m := \|\hat{y} - P_m^+ P_m \hat{y}\|$.

5.3. Comparison of the both approaches. We now compare the both approaches directly. We consider discretisation by box functions with $m_\infty = 4000$ and $m = 50, 100, 200$ and discretisation by hat functions (compare to Proposition 2.6). The latter is precisely implemented as follows. With $k = \frac{m_\infty - 1}{m - 1}$ we set

$$P_m : \mathbb{R}^{m_\infty} \rightarrow \mathbb{R}^m$$

$$\begin{pmatrix} y_{(i-1)k+1} \\ \dots \\ y_{(i+1)k+1} \end{pmatrix} \mapsto \frac{1}{\sqrt{\sum_{j=1}^{2k+1} a_j^2}} (a_1 y_{(i-1)k+1} + \dots + a_{2k+1} y_{(i+1)k+1}) e_i$$

where $i = 2, \dots, m - 1$ and

$$a_i := \begin{cases} (i-1)/k & i \leq k+1, \\ 1 - (i-k-1)/k & i \geq k+1. \end{cases}$$

For the boundaries we set,

$$\begin{pmatrix} y_1 \\ \dots \\ y_{k+1} \end{pmatrix} \mapsto \frac{1}{\sqrt{\sum_{i=k+1}^{2k+1} a_i^2}} (a_{k+1} y_1 + \dots + a_{2k+1} y_{k+1}) e_1$$

and

$$\begin{pmatrix} y_{m_\infty - (k+1)} \\ \dots \\ y_{m_\infty} \end{pmatrix} \mapsto \frac{1}{\sqrt{\sum_{i=1}^{k+1} a_i^2}} (a_1 y_{m_\infty - (k+1)} + \dots + a_{k+1} y_{m_\infty}) e_m.$$

Here we use $m_\infty = 4132$ and $m = 18, 28, 52$. We first applied Algorithm 2 with exact upper bound $\delta_m^{disc} = \|\hat{y} - P_m^+ P_m \hat{y}\|$. The (random) stopping index $n(m, \delta_m^{disc})$ from Algorithm 2 is then used as the number of repetitions n in Algorithm 1. We plot in Figure 2 the relative errors of the both approaches for growing number of measurement channels m . We observe the stated convergence as m grows. Moreover, the errors of the approach with finite-dimensional residuum are even slightly better than the ones of the approach with infinite-dimensional approach in all the examples. This gives numerical evidence, that also the first approach is order optimal in various settings.

6. Conclusion. In this work, we have analysed linear inverse problems under unknown white noise. We presented two approaches for the solution. In both cases, we used multiple discretised measurements to prove convergence in probability against the true solution, as the number of repetitions and the number of measurement channels tend to infinity. The first approach neither required knowledge of the arbitrary error distribution, nor quantitative knowledge of the quality of the discretisation to obtain convergence. For the second approach we also proved an optimal convergence rate, under additional knowledge of the discretisation error.

We want to pronounce two important outstanding questions. Firstly, the discretisation considered in this article entered the problem through discretised measurements. In particular, this is determined by the practical problem and the way

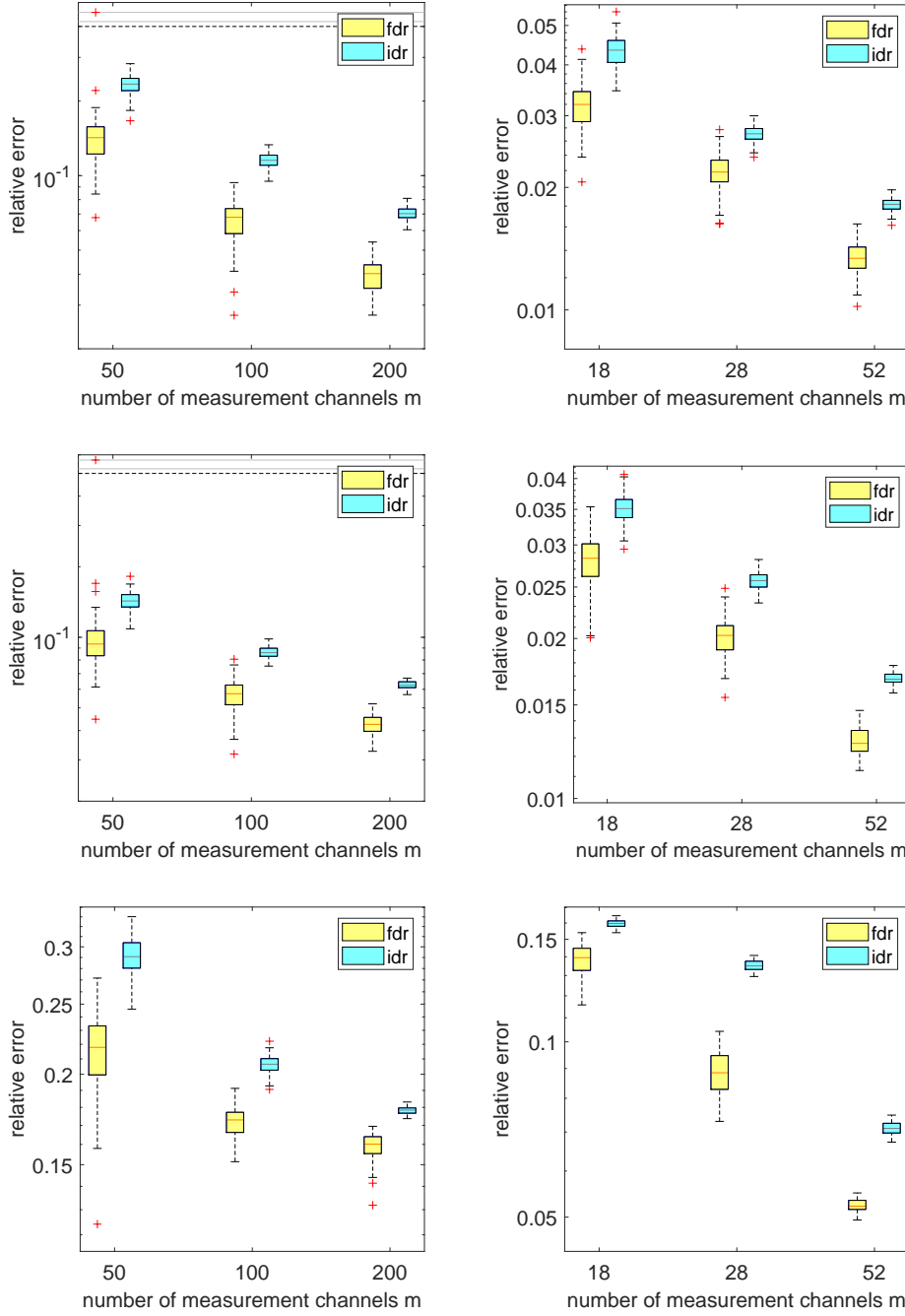


FIG. 2. Direct comparison of both approaches (2) (fdr) and (6) (idr) with discrepancy principle as implemented in Algorithm 1 and 2 for 'phillips' (first line), 'gravity' (second line) and 'shaw' (third line). For the discretisation of the measurements either box functions (first column) or hat functions (second column) are used. Concretely, the relative errors $\left\| R_{\alpha} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ (fdr) and $\left\| R_{\alpha} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ (idr) are plotted against the number of measurement channels m , where δ_m^{disc} is chosen to be the exact discretisation error $\left\| \hat{y} - P_m^+ P_m \hat{y} \right\|$ and $n(m, \delta_m^{disc})$ is calculated with Algorithm 2.

the data is measured or acquired. In order to solve the problem numerically, as in the preceding section, one also has to discretise the true unknown \hat{x} . In contrast to the measurements, here there is more freedom to choose the numerical discretisation, since one is basically only limited by computational power. It therefore is of high interest to find an optimal choice for that. Secondly, it might come as a surprise that in all the numerical examples the approach with finite-dimensional residuum (fdr) gives slightly better results than the one with infinite-dimensional residuum (idr), even though the theoretical results do only guarantee the optimality of the latter one. Thus an important open question is to derive natural and verifiable conditions, which rigorously guarantee optimality of the first approach.

REFERENCES

- [1] I. BABUŠKA AND J. OSBORN, *Eigenvalue problems*, (1991).
- [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] N. BISSANTZ, T. HOHAGE, A. MUNK, AND F. RUYMGAART, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 2610–2636.
- [4] G. BLANCHARD AND P. MATHÉ, *Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration*, Inverse problems, 28 (2012), p. 115011.
- [5] L. CAVALIER, *Inverse problems in statistics*, in Inverse problems and high-dimensional estimation, Springer, 2011, pp. 3–96.
- [6] L. CAVALIER AND A. TSYBAKOV, *Sharp adaptation for inverse problems with random noise*, Probability Theory and Related Fields, 123 (2002), pp. 323–354.
- [7] D. L. DONOHO, *Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition*, Applied and computational harmonic analysis, 2 (1995), pp. 101–126.
- [8] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.
- [9] H. GARDE AND N. HYVÖNEN, *Mimicking relative continuum measurements by electrode data in two-dimensional electrical impedance tomography*, arXiv preprint arXiv:2001.10604, (2020).
- [10] A. GUT, *Probability: a graduate course*, vol. 75, Springer Science & Business Media, 2013.
- [11] P. C. HANSEN, *Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems*, Numerical algorithms, 6 (1994), pp. 1–35.
- [12] P. C. HANSEN, *Discrete inverse problems: insight and algorithms*, vol. 7, Siam, 2010.
- [13] B. HARRACH, T. JAHN, AND R. POTTHAST, *Beyond the Bakushinskii veto: Regularising linear inverse problems without knowing the noise distribution*, Numerische Mathematik, 145 (2020), pp. 581–603.
- [14] K. ITO AND B. JIN, *Inverse problems: Tikhonov theory and algorithms*, World Scientific, 2015.
- [15] T. JAHN AND B. JIN, *On the discrepancy principle for stochastic gradient descent*, Inverse Problems, 36 (2020), p. 095009, <https://doi.org/10.1088/1361-6420/abaa58>, <https://doi.org/10.1088/1361-6420/abaa58>.
- [16] J. KAIPIO AND E. SOMERSALO, *Statistical inverse problems: discretization, model reduction and inverse crimes*, Journal of computational and applied mathematics, 198 (2007), pp. 493–504.
- [17] P. MATHÉ AND S. PEREVERZEV, *Regularization of some linear ill-posed problems with discretized random noisy data*, Mathematics of Computation, 75 (2006), pp. 1913–1929.
- [18] P. MATHÉ AND S. V. PEREVERZEV, *Optimal discretization of inverse problems in Hilbert scales. regularization and self-regularization of projection methods*, SIAM Journal on Numerical Analysis, 38 (2001), pp. 1999–2021.
- [19] P. MATHÉ AND S. V. PEREVERZEV, *Discretization strategy for linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), p. 1263.
- [20] P. MATHÉ AND S. V. PEREVERZEV, *Geometry of linear ill-posed problems in variable hilbert scales*, Inverse problems, 19 (2003), p. 789.
- [21] V. A. MOROZOV, *The error principle in the solution of operational equations by the regularization method*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 8 (1968), pp. 295–309.
- [22] S. M. RUMP, *Verified bounds for singular values, in particular for the spectral norm of a matrix*

- and its inverse*, BIT Numerical Mathematics, 51 (2011), pp. 367–384.
- [23] G. WAHBA, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM Journal on Numerical Analysis, 14 (1977), pp. 651–667.