

Regularising linear inverse problems under unknown non-Gaussian white noise allowing repeated measurements

BASTIAN HARRACH[†] AND TIM JAHN[‡]

Institute of Mathematics, Goethe-University Frankfurt, Frankfurt am Main, Germany

AND

ROLAND POTTHAST[§]

German Weather Service, Offenbach am Main, Germany

[Received on 6 August 2021]

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 reduce and 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.

Keywords: statistical inverse problems; discretisation; white noise; discrepancy principle.

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 Harrach *et al.* (2020) and Jahn (2021a) 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,

[†]Email: harrach@math.uni-frankfurt.de

[‡]Corresponding author. Email: jahn@math.uni-frankfurt.de

[§]Email: Roland.Potthast@dwd.de

$$Y_{ij} := l_j(\hat{y}) + \delta_{ij} \quad (1.1)$$

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

$$\begin{pmatrix} Y_{i1} - l_1(\hat{y}) \\ \dots \\ Y_{im} - l_m(\hat{y}) \end{pmatrix}_{i \in \mathbb{N}} \subset \mathbb{R}^m$$

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 exists 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

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

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

$$\arg \min_{x \in \mathcal{X}} \left\| \begin{pmatrix} l_1(Kx) \\ \dots \\ l_m(Kx) \end{pmatrix} - \bar{Y}_n^{(m)} \right\|_{\mathbb{R}^m}^2 + \alpha \|x\|_{\mathcal{X}}^2. \quad (1.2)$$

The main question of this work is whether the unique minimiser of (1.2), 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

$$\delta_{m,n}^{est} := \sqrt{\frac{m}{n} s_{m,n}^2}, \quad (1.3)$$

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. The estimator $s_{m,n}^2$ for the unknown variance $E(Y_{ij} - l_j(\hat{y}))^2$ is natural in our general setting. If one has more information about the structure of the discretisation, e.g. in regression problems where the unknown function is measured along a grid, more specified choices may also be reasonable. See Rice *et al.* (1984) and Dette *et al.* (1998), where the variance is estimated in such settings with only one measurement on each channel (i.e. for $n = 1$). In Dai *et al.* (2015) different methods are compared to each other for repeated measurements on each channel $n > 1$. In particular, it is shown that our choice is asymptotically optimal (for $n, m \rightarrow \infty$), but that there are better choices for finite sample sizes given that higher moments of the measurement error exist. From a deterministic view point, in order to guarantee convergence it would arguably be necessary to assure that the measurement error tends to 0, i.e. that $\delta_{m,n}^{est} \rightarrow 0$ in probability (or a.s. or in root mean square), which holds if and only if $m/n \rightarrow 0$ (see Proposition 5.2). This will be a central assumption in most of this manuscript. In the lens of classical results from the statistical side this however seems to be an unnaturally strong condition, since in many special cases it is sufficient to have that the (overall) measurement error stays bounded, i.e. that $n = m \rightarrow \infty$ (Vogel (2002)) or even that only the component-wise measurement error converges to 0, i.e. that merely $m, n \rightarrow \infty$ without any specific

relation between m and n (Cavalier (2011)). We will show in Section 4 below that somehow surprisingly the condition $m/n \rightarrow 0$ is in essence necessary to guarantee convergence in our general setting.

One of the most natural and popular strategies to determine the regularisation parameter α in (1.2) is the discrepancy principle Morozov (1968), which constitutes in solving

$$\left\| \begin{pmatrix} l_1(KR_\alpha^{(m)} \bar{Y}_n^{(m)}) \\ \dots \\ l_m(KR_\alpha^{(m)} \bar{Y}_n^{(m)}) \end{pmatrix} - \bar{Y}_n^{(m)} \right\| \approx \delta_{m,n}^{est} \quad (1.4)$$

(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 for $\alpha_{m,n}$ determined by the discrepancy principle (1.4) there holds

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

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) and for arbitrary unknown error distributions. This might be surprising in view of the Bakushinskii veto (Bakushinskii (1984)), which states that quantitative a priori knowledge about the noise is a crucial requirement for solving an inverse problem. We stress that Corollary 1.1 does not give a convergence rate. In order to obtain a rate additional smoothness assumptions (relative to the forward operator K) have to be imposed on the true solution \hat{x} and the relation of K and the discretisation will play a crucial role. This is a topic of actual research and postponed to a later work. Other than that we want to present an alternative approach, which allows to deduce rates in very general settings. However, note that in what follows the rates are deduced by a classical worst-case error analysis and are not optimal in the statistical setting. Whether the discrepancy principle can be modified to attain optimal rates (in the statistical setting) in our general frame work is beyond the scope of this work. We will discuss this in more detail in Section 4.

The main idea of the alternative approach is to first construct from the measured data in \mathbb{R}^m continuous measurements in the Hilbert space \mathcal{Y} , see e.g. Garde & Hyvönen (2021). For that we solve the following optimisation problem

$$\min_{y \in \mathcal{Y}} \left\| \begin{pmatrix} l_1(y) \\ \dots \\ l_m(y) \end{pmatrix} - \bar{Y}_n^{(m)} \right\|, \quad (1.5)$$

which has an unique solution with minimal norm (due to Moore-Penrose) denoted by $Z_n^{(m)}$ in the following. We restrict to discretisations for which (1.5) is well-conditioned, see Assumption 3.1. For general discretisations one would need to add an additional regularisation term. Then, instead of (1.2) we solve the following optimisation problem with infinite-dimensional residuum (idr)

$$\arg \min_{x \in \mathcal{X}} \left\| Kx - Z_n^{(m)} \right\|_{\mathcal{Y}}^2 + \alpha \|x\|_{\mathcal{X}}^2. \quad (1.6)$$

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

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

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

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

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

$$\left\| KR\alpha Z_{n(m, \delta_m^{disc})}^{(m)} - Z_{n(m, \delta_m^{disc})}^{(m)} \right\| \approx 2\delta_m^{disc}, \quad (1.7)$$

with $R\alpha Z_{n(m, \delta_m^{disc})}^{(m)}$ the unique solution of (1.6) (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 3.2). 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 (1.7). 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

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

for some constant L' .

It is a long standing dilemma that 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 Harrach *et al.* (2020) 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 semi-discretised measurements under arbitrary unknown white noise. It is widely known that discretisation has a regularising effect, see for example Mathé & Pereverzev (2001) and Hansen (2010) for the discretisation in the deterministic setting, O'Sullivan (1986), Mathé & Pereverzev (2001), Mathé & Pereverzev (2003a) and Mathé & Pereverzev (2006) for

the statistical frequentist setting and Kaipio & Somersalo (2007) and Ito & Jin (2015) for the Bayesian approach. In applications related to machine learning often one considers discretisation by random sampling, see e.g. De Vito *et al.* (2006) or Bauer *et al.* (2007). 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. Fairly often, inverse problems under white noise (see e.g. Donoho (1995) and Cavalier & Tsybakov (2002)) are treated the first way and the white noise is modelled as a Hilbert space process operating on \mathcal{Y} , see Bissantz *et al.* (2007) and Cavalier (2011). The major challenge of this modelling is that then the measurements are not elements of \mathcal{Y} . This implies that one has to restrict to sufficiently smoothing operators and to include correction terms in the convergence 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, e.g. cross validation Wahba (1977) or Lepski's balancing principle Mathé & Pereverzev (2003b), even though a modified discrepancy principle could be applied Blanchard & Mathé (2012). These technical difficulties are not present in the semi-discretised setting considered here. Among the first results on the discrepancy principle in such a setting we want to mention Vogel (2002), where a convergence rate analysis is given under the assumption that the singular value decomposition is known. There the regularisation parameter is determined not for the random residual as in (1.4) but for its squared expectation. While this gives some important insight such a choice is clearly not implementable. Results for the truly data-driven implementable version (1.4) are presented in Blanchard *et al.* (2018b) and Blanchard *et al.* (2018a), where optimal rates are deduced for polynomially ill-posed problems under Gaussian white noise. We will compare our results to these in more detail in Section 4. In particular we show that the existence of the fourth moment of the error distribution is a crucial requirement for the latter references. A major difference to most of the aforementioned references is that there the variance of the measurement error (respectively the noise level of the white noise process) is assumed to be known. This is justified by the fact that usually little attention is put on the behaviour of the solution as the discretisation dimension grows and in fact the error distribution is assumed to be independent of the size of the discretisation. Here we explicitly allow the error distribution to vary with the size of the discretisation and thus we make the estimation part of the analysis.

To put it in a nut shell, the main 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 that is known to be challenging in stochastic regularisation even for strongly L^2 -bounded noise, see Harrach *et al.* (2020) and Jahn & Jin (2020). 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 5.2 and Lemma 5.3, do not depend on the chosen regularisation or parameter choice rule. Finally, if one neither has information about the noise level nor is one able to repeat a measurement solely so-called heuristic parameter choice rules could be used. The term heuristic is referring to the fact that convergence results only hold under a restricted noise analysis. Here we want to mention the quasi-optimality criterion as one the most popular heuristic rules, see e.g. Tikhonov & Glasko (1965) and Kindermann *et al.* (2018) for results under (almost surely) bounded noise. See also Bauer & Reiß (2008), where an analysis of the quasi-optimality criterion under white noise in a Bayesian setting is presented.

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. In Section

4 we compare the results in detail to existing ones. The proofs are deferred to Section 5 and we conclude with a numerical study in Section 6 and some final remarks in Section 7.

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)

$$P_m : \mathcal{Y} \rightarrow \mathbb{R}^m, y \mapsto \begin{pmatrix} l_1^{(m)}(y) \\ \dots \\ l_m^{(m)}(y) \end{pmatrix}, \quad (2.1)$$

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 $P_m P_m^*$ is bijective and thus P_m has a singular value decomposition with exactly m (non-zero) singular values.

From now on we consider general 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) $(F_\alpha)_{\alpha > 0}$ are piecewise continuous real valued functions on $[0, \|K\|^2]$ with

$$\lim_{\alpha \rightarrow 0} \sup_{\varepsilon \leq \lambda \leq \|K\|^2} |F_\alpha(\lambda) - 1/\lambda| = 0 \quad (2.2)$$

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 exists 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 exists 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 2.1 Assumption 2.1 coincides with the classical ones in Engl *et al.* (1996) up to (2.2), 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 these 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 2.2 (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 2.2, starting with the one from the introduction.

PROPOSITION 2.1 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 2.2 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 2.2 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 2.2 is fulfilled with $A = Id$.

PROPOSITION 2.3 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 2.2 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 2.3 (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 2.4 Assume that K is injective, the discretisation fulfills Assumption 2.2, the error is accordingly to Assumption 2.3 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

$$\left\| P_m K R_\alpha^{(m)} \bar{Y}_n^{(m)} - \bar{Y}_n^{(m)} \right\| \approx \delta_{m,n}^{est} \quad (2.3)$$

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 $\bar{Y}_n^{(m)} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} Y_{i1}^{(m)} \\ \vdots \\ Y_{im}^{(m)} \end{pmatrix}$;
 - 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** $\left\| \begin{pmatrix} l_1^{(m)}(K R_{q^k}^{(m)} \bar{Y}_n^{(m)}) \\ \vdots \\ l_m^{(m)}(K R_{q^k}^{(m)} \bar{Y}_n^{(m)}) \end{pmatrix} - \bar{Y}_n^{(m)} \right\| > \tau \delta_{m,n}^{est}$ **do**
 - 7: $k = k + 1$;
 - 8: **end while**
 - 9: $\alpha_{m,n} = q^k$;
-

Algorithm 1 terminates (with a 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 Harrach *et al.* (2020). We now extend the assumptions of the error in the introduction.

ASSUMPTION 2.4 (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 2.4.1 and 2.4.2 is that for the latter the error distribution may vary with m , to the cost of a uniform moment condition.

REMARK 2.2 Assumption 2.4.2 guarantees that the error distribution does not degenerate too much. It is trivially fulfilled if e.g. $\delta_{ij}^{(m)} \stackrel{d}{=} c_m X$, 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 Harrach *et al.* (2020)).

THEOREM 2.5 Assume that K is injective with dense range and that the discretisation fulfills Assumption 2.2 and that the error is accordingly to Assumption 2.4 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 2.5 and Proposition 2.1. We conclude the section with a remark regarding Assumption 2.4.

REMARK 2.3 As already mentioned Assumption 2.4 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 $p > 1$

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

as $m \rightarrow \infty$. Thus Assumption 2.4 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

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

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 (1.6). 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

$$2\delta_m^{disc} \approx \left\| \hat{y} - P_m^+ P_m \hat{y} \right\| + \left\| P_m^+ P_m \hat{y} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\| \geq \left\| \hat{y} - P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} \right\|. \quad (3.1)$$

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

$$\limsup_{m \rightarrow \infty} \kappa(P_m) := \limsup_{m \rightarrow \infty} \|P_m\| \|P_m^+\| = \limsup_{m \rightarrow \infty} \frac{\max_{j=1, \dots, m} \sigma_j^{(m)}}{\min_{j=1, \dots, m} \sigma_j^{(m)}} \leq \limsup_{m \rightarrow \infty} \frac{C_m}{c_m} < \infty. \quad (3.2)$$

Often the stability assumption (3.2) can be guaranteed by an angle condition for the unique $\eta_j^{(m)} \in \mathcal{Y}$ that fulfill $l_j^{(m)}(y) = (\eta_j, y)$ for all $y \in \mathcal{Y}$.

PROPOSITION 3.1 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. It would be desirable to also have a simple criterion to guarantee that δ_m^{disc} tends to 0. For $P_m = (l_1(\cdot) \ \dots \ l_m(\cdot))^T$ (the l_j do not depend on m) this could be guaranteed e.g. when $(l_j)_{j \in \mathbb{N}}$ is complete, because then $\mathcal{N}(P_m) \supset \mathcal{N}(P_{m+1}) \supset \dots$ converges monotonically to 0 and $P_m^+ P_m = P_{\mathcal{N}(P_m)^\perp}$ is the orthogonal projection onto the orthogonal complement of $\mathcal{N}(P_m)$. A straight forward generalisation of completeness to discretisation schemes $P_m = (l_1^{(m)}(\cdot) \ \dots \ l_m^{(m)}(\cdot))^T$ would be to presume, that for all $y \in \mathcal{Y} \setminus \{0\}$ there exists a $\varepsilon > 0$

such that $\|P_m y\| \geq \varepsilon$ for m large enough. The following counter example however shows that this is not sufficient to guarantee that the discretisation error tends to 0.

REMARK 3.1 Let $(v_j)_{j \in \mathbb{N}}$ be an orthonormal basis of \mathcal{Y} . Set $l_j^{(m)}(y) = (y, v_j)$ for $j = 2, \dots, m$ and $l_1^{(m)}(y) = (y, v_1/\sqrt{2} + v_{m+1}/\sqrt{2})$. For $y \neq 0$ we set $\varepsilon = |(y, v_j)|/2$ with $j = \min\{j' : (y, v_{j'}) \neq 0\}$. Then clearly $\|P_m y\| \geq \varepsilon$ for m large enough. But, it holds that $\mathcal{N}(P_m) = \langle v_1/\sqrt{2} - v_{m+1}/\sqrt{2}, v_{m+2}, v_{m+3}, \dots, \rangle$ and thus

$$v_1 - P_m^+ P_m v_1 = P_{\mathcal{N}(P_m)} v_1 = v_1/\sqrt{2} \not\rightarrow 0$$

for $m \rightarrow \infty$.

We now show that Assumption 3.1 is fulfilled for various popular discretisation schemes. We start with the example from the introduction.

PROPOSITION 3.2 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 exists a l_j such that $l_j(y) \neq 0$, and well-conditioned that is

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

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 5 for the definition of the singular value decomposition.

PROPOSITION 3.3 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 3.4 Consider $\mathcal{X} = \mathcal{Y} = L^2(0, 1)$ and $\eta_j^{(m)}$ the box functions from Proposition 2.2. 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 3.5 Consider $\mathcal{X} = \mathcal{Y} = L^2(0, 1)$ and $\eta_j^{(m)}$ the hat functions from Proposition 2.3. 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 holds that

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

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 3.6 Assume that K is injective, the discretisation fulfills Assumption 3.1, the error is accordingly to Assumption 2.3 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

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

$$\text{for } n(m, \delta_m^{disc}) = \lceil \frac{m}{c_m^2 \delta_m^{disc2}} \rceil.$$

REMARK 3.2 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 $n(m, \delta_m^{disc}) := \min \left\{ n' \geq 1 : \frac{ms_{m,n'}^2}{n'c_m^2} \leq \delta_m^{disc2} \right\}$ 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 3.7 Assume that K is injective with dense range and that the discretisation fulfills Assumption 3.1 and that the error is accordingly to Assumption 2.4, with $p \geq 2$ in the case of 2.4.2 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 exist 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

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

for some constant L' .

Now Corollary 1.2 in the introduction is an easy consequence of Theorem 3.7 and Proposition 3.2.

4. Discussion

In this section we discuss the above results in more detail in the light of classical results for statistical inverse problems. Classical results are usually formulated for white noise with intensity σ^2 . With averaging multiple measurements we control the size of the white noise, in fact it holds that $\sigma^2 \asymp 1/n$. Many convergence results require that the size of the discretised measurements is constant for m , i.e. that $\sigma^2 = 1/m$. Thus, the classical results hold for $n = m$ and the condition $m/n \rightarrow 0$ for the convergence results in Section 2 seems very strong. Our first result shows that in our general setting this condition is necessary to ensure convergence for a priori regularisation methods. Hereby, note that in the classical

deterministic theory a general ill-posed linear problem $K\hat{x} = \hat{y}$ with K some ill-posed linear operator and $(y^\delta)_{\delta>0}$ a sequence of (deterministic) measurements fulfilling $\|\hat{y} - y^\delta\| \leq \delta$ can be solved using any filter based regularisation fulfilling Assumption 2.1 together with a proper a priori choice rule $\alpha = \alpha(\delta)$ (e.g. $\alpha(\delta) = \delta$). In particular, such a choice depends only on the noise level δ and as such is independent of K . Now we assume that in our statistical setting the number of measurement channels equals the number of repetitions, i.e. $m = n$. Further assume that $\alpha = \alpha_m$ is any possible a priori parameter choice rule that converges monotonically to 0 as $m \rightarrow \infty$.

PROPOSITION 4.1 There exist a compact operator $K : l^2(\mathbb{N}) \rightarrow l^2(\mathbb{N})$, an element $\hat{y} \in l^2(\mathbb{N})$, a discretisation scheme $P_m : l^2(\mathbb{N}) \rightarrow \mathbb{R}^m$ and an error model $(\delta_{ij})_{ij \in \mathbb{N}}$ such that

$$\mathbb{E} \|R_{\alpha_m}^{(m)} \bar{Y}_m^{(m)} - K^+ \hat{y}\|^2 \rightarrow \infty$$

as $m \rightarrow \infty$, where $\bar{Y}_n^{(m)} = \sum_{i=1}^m (Y_{i1} \ \dots \ Y_{im})^T / m$ with $Y_{ij} = (P_m \hat{y})_j + \delta_{ij}$ and $(R_\alpha^{(m)})_{\alpha>0}$ is the Tikhonov regularisation.

Note that in the above proposition it was important that we fixed the a priori choice rule before the ill-posed problem given through K . If we restrict to certain classes, e.g. to mildly ill-posed problems (i.e. the singular values of K fulfill $\sigma_j^2 \asymp j^{-q}$ for some $q > 0$), one can give a priori parameter choice rules which converge for $m = n$.

We now compare our results in detail to recent results for the discrepancy principle and ultimately show that here the condition $m/n \rightarrow 0$ is necessary even if one restricts to mildly ill-posed problems. In Blanchard *et al.* (2018a) and Blanchard *et al.* (2018b) order-optimal L^2 -rates are given for the discrepancy principle under Gaussian noise and sufficiently unsmooth data. In these articles the implementation of the discrepancy principle differs from ours (see Algorithm 1) as there in essence the hyperparameter τ_m depends on the number of measurement channels m , whereas we choose a constant $\tau > 1$ as in the classical deterministic theory. Precisely, there the regularisation parameter is essentially determined as

$$\inf_{\alpha>0} \|P_m K R_\alpha^{(m)} \bar{Y}_n^{(m)} - \bar{Y}_n^{(m)}\| \leq \kappa_{m,n}$$

with $|\kappa_{m,n}^2 - \frac{m}{n}| = \mathcal{O}\left(\frac{\sqrt{m}}{n}\right)$ (note that $\frac{m}{n} = \mathbb{E}[\delta_{m,n}^{est,2}] = \mathbb{E}\|\bar{Y}_n^{(m)} - P_m \hat{y}\|^2$ is the (expected) squared noise if $\mathbb{E}[\delta_{ij}^2] = 1$). Apart from the fact that we consider more general discretisation schemes, the main difference to the aforementioned results is that we allow for general unknown error distributions. If one instead of L^2 -convergence asks only for convergence in probability it seems to us that the assumption of Gaussian noise in Blanchard *et al.* (2018a) and Blanchard *et al.* (2018b) could be relaxed to arbitrary distributions obeying a finite fourth moment. In fact, under that relaxed assumption one can show that the oscillation of the residual is of a comparably small order and then the choice $\kappa_{m,n}$, due to the correct leading order, can capture the smoothness of \hat{x} more accurately (and exactly up to saturation) than the plain discrepancy principle (i.e. the choice $\tau > 1$) would do and thus gives better convergence rates. However, this procedure seems not to be stable if higher moments do not exist. From the following result one can directly deduce that then the analysis in Blanchard *et al.* (2018a) and Blanchard *et al.* (2018b) breaks down. In particular it shows using a counter example that for both choices the discrepancy principle does not converge in any commonly used mode, if n/m does not converge to 0.

THEOREM 4.2 Let $K : l^2(\mathbb{N}) \rightarrow l^2(\mathbb{N})$ be diagonal with singular values $\sigma_j^2 = j^{-q}$ with $q > 1$ and singular basis $(v_j)_{j \in \mathbb{N}}$. Let P_m be the discretisation along the singular values, i.e. $(P_m v_j)_l = (v_j, v_l)$ for $j \in \mathbb{N}$ and $l = 1, \dots, m$. Let $Y_{ij} = (P_m \hat{y})_j + \delta_{ij}$ be i.i.d. measurements of $\hat{y} \in l^2(\mathbb{N})$ (which will be specified below),

where δ_{ij} has density f_ε (with $\varepsilon < 2/11$ as given in Proposition 5.4 below) and consider $(R_\alpha)_{\alpha>0}$ the spectral cut-off regularisation. Let $c > 0$. If $\hat{y} = \sum_{j=1}^{\infty} j^{-\frac{q}{2}-1} v_j$ and the regularisation parameter $\alpha_{m,n}$ is determined with Algorithm 1 and $\tau > 1$, then there exists $C > 0$ such that

$$\lim_{\substack{m,n \rightarrow \infty \\ m/n \geq c}} \mathbb{P} \left(\|R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y}\| \geq C \right) = 1.$$

If $\hat{y} = 0$ and the regularisation parameter $\alpha_{m,n}$ is determined with Algorithm 1, where in line 6 the right hand side $\tau \delta_{m,n}^{est}$ is replaced with $\sqrt{\frac{m+\sqrt{m}}{n}}$, then there exists $p_\varepsilon > 0$ such that for any $L > 0$ it holds that

$$\liminf_{\substack{m,n \rightarrow \infty \\ m/n \geq c}} \mathbb{P} \left(\|R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - K^+ \hat{y}\| \geq L \right) > p_\varepsilon.$$

A modulation of the discrepancy principle which yields optimal rates (in probability) for linear problems requiring only a finite second moment is studied in Jahn (2021b). The analysis there however is restricted to spectral cut-off regularisation.

We finish this section with a comment on the way we measure smoothness. As already mentioned in the introduction, existing results usually pay little attention on the behaviour of the solution as the discretisation dimension m tends to ∞ . Consequently, the source conditions allowing to perform a convergence rate analysis are formulated in the discretised setting. I.e. smoothness of \hat{x} is not measured relative to the infinite-dimensional problem given by K (as it is here), but to the discretised one given by $P_m K$. In the latter case a standard worst-case analysis would yield a convergence rate for the approach with finite-dimensional residuum (which again would not be optimal) and one could compare the rates of the both approaches with finite-dimensional and infinite-dimensional residuum respectively. In Section 6 this is done numerically with problems from the open source MATLAB package Regutools (Hansen (1994)). There the approach with finite-dimensional residuum gives slightly better rates and is hence preferable, due to its better stability properties (e.g. convergence without knowledge of a discretisation error). The following example however shows that through discretisation the smoothness of \hat{x} may be substantially deteriorated, in which case the approach with infinite-dimensional residuum would perform better.

Let $K : l^2(\mathbb{N}) \rightarrow l^2(\mathbb{N})$ be a diagonal operator with $K v_j = j^{-1} v_j$. Let $\hat{x} = \sum_{j=1}^{\infty} j^{-2} v_j$ and consider the discretisation $P_m : l^2(\mathbb{N}) \rightarrow \mathbb{R}^m$ with

$$(P_m y)_j = l_j^{(m)}(y) := \begin{cases} (y, v_j) & \text{for } j = 1, \dots, m-1, \\ e^{-m} v_m + \sqrt{1 - e^{-2m}} v_{\lceil e^m \rceil} & \text{for } j = m. \end{cases}$$

It holds that $\|l_j^{(m)}\| = 1$ for all $m \in \mathbb{N}, j = 1, \dots, m$, that $P_m^* P_m y \rightarrow y$ as $m \rightarrow \infty$ for all $y \in l^2(\mathbb{N})$ and that $\kappa(P_m) = 1$, thus the discretisation $(P_m)_{m \in \mathbb{N}}$ fulfills Assumptions 2.2 and 3.1. Let $\hat{x} := \sum_{j=1}^{\infty} j^{-2} v_j$. Then there exist $\nu, \rho > 0$ and $w \in l^2(\mathbb{N})$ with $\|w\| \leq \rho$ and $\hat{x} = (K^* K)^{\frac{\nu}{2}} w$ (a possible choice would be $\nu = 1$ and $\rho = \sqrt{\sum_{j=1}^{\infty} j^{-2}} = \frac{\pi}{\sqrt{6}}$). I.e. \hat{x} obeys smoothness ν, ρ relative to K . However, the following proposition shows that \hat{x} obeys asymptotically only a much worse smoothness relative to $P_m K$ (even though $(P_m K)^* P_m K \rightarrow K^* K$ uniformly as $m \rightarrow \infty$ by Lemma 5.1 below).

PROPOSITION 4.3 Let \hat{x} and K, P_m be given as above. Let $\nu_m, \rho_m > 0$ and $w_m \in l^2(\mathbb{N})$ be such that $P_{\mathcal{N}(P_m K)^\perp} \hat{x} = ((P_m K)^* P_m K)^{\frac{\nu_m}{2}} w_m$ with $\|w_m\| = \rho_m$. Then there exist $c, \varepsilon > 0$ such that either

$$\begin{aligned}
& \lim_{m \rightarrow \infty} v_m = 0, \quad \lim_{m \rightarrow \infty} \rho_m = \infty, \\
\text{or } & \liminf_{m \rightarrow \infty} v_m > 0, \quad \liminf_{m \rightarrow \infty} \rho_m \geq e^{\varepsilon m}, \\
\text{or } & \limsup_{m \rightarrow \infty} \rho_m = \infty, \quad \limsup_{m \rightarrow \infty} v_m \leq c \frac{\log(m)}{m}
\end{aligned}$$

holds.

5. Proofs

In this section we collect the proofs. We will need the singular value decomposition of an injective compact operator A (see Cavalier (2011)): 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 $A^* u_l = \sigma_l v_l$.

5.1 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 5.1 Assume that K is injective and that Assumption 2.2 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 semidefinite. 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 exists 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

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

Because $A_m \rightarrow A$ pointwise there exists 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 Babuška & Osborn (1991). 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 Babuška & Osborn (1991), there exists 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 exists a $M \in \mathbb{N}$ with $M = \sum_{i=1}^{i^*} |I_i|$ for some $i^* \in \mathbb{N}$ such that

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

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

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

for m sufficiently large. \square

5.1.1 *Proof of Proposition 2.1.* It holds that $\sup_{m \in \mathbb{N}} \|P_m y\| = \sup_{m \in \mathbb{N}} \sum_{j=1}^m l_j(y)^2 < \infty$, 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 $P_m^* P_m y \rightarrow Ay$ with $A = P_\infty^* P_\infty$ and A is injective because of the completeness condition.

5.1.2 *Proof of Proposition 2.2.* 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.

5.1.3 *Proof of Proposition 2.3.* 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

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

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 holds that

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

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 $m \rightarrow \infty$.

5.1.4 *Proof of Theorem 2.4.* We will need the following proposition for the convergence proofs.

PROPOSITION 5.1 Assume that Assumption 2.2 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 exists 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

$$\|R_{\alpha}\| \leq \sqrt{C_R C_F / \alpha}, \quad (5.1)$$

see e.g. Engl *et al.* (1996) or Proposition 1 of Harrach *et al.* (2020). Therefore it follows that

$$\left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \right\|^2 \mathbb{E} \delta_m^{meas2} \leq \left(\left\| R_{\alpha(\delta_{m,n}^{est})}^{(m)} \right\| \delta_{m,n}^{est} \right)^2 C_d \leq C_d C_R C_F \frac{\delta_{m,n}^{est2}}{\alpha(\delta_{m,n}^{est})} \rightarrow 0 \quad (5.2)$$

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

$$\lim_{m \rightarrow \infty} \left\| \hat{x} - P_{\mathcal{N}(P_m K)^{\perp}} \hat{x} \right\| = \lim_{m \rightarrow \infty} \left\| P_{\mathcal{N}(P_m K)} \hat{x} \right\| = 0 \quad (5.3)$$

by Proposition 5.1. Finally, for any $\varepsilon > 0$ by Lemma 5.1.2 there exists 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

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

By Lemma 5.1.1, (2.2) and since $\alpha(\delta_{m,n}^{est}) \rightarrow 0$ for $m, n \rightarrow \infty, m/n \rightarrow 0$

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

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 (5.2) and (5.3).

5.1.5 Proof of Theorem 2.5. By the nature of white noise we cannot expect the error to concentrate along a certain direction, in contrast to Harrach *et al.* (2020). 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 5.2 Assume that the error fulfills Assumption 2.4. 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 2.4.2, 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 Gut (2013) 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 2.4.1, by the Kolmogorov-Doob-inequality

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

It holds that

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

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 $\mathbb{E} |\sum_j X_j^{(m)}| \rightarrow 0$ as $m \rightarrow \infty$. Let $\varepsilon' > 0$. By dominated convergence and integrability of $X_j^{(m)}$ there exists $M > 0$ large enough such that for $Y_j^{(m)} := X_j^{(m)} \mathcal{X}_{\{|X_j^{(m)}| \leq M\}}$ and $Z_j^{(m)} := X_j^{(m)} \mathcal{X}_{\{|X_j^{(m)}| > M\}}$ it

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

$$\begin{aligned}
 \mathbb{E} \left| \sum_{j=1}^m X_j^{(m)} \right| &\leq \mathbb{E} \left| \sum_{j=1}^m Y_j^{(m)} - \mathbb{E}Y_j^{(m)} \right| + \mathbb{E} \left| \sum_{j=1}^m Z_j^{(m)} - \mathbb{E}Z_j^{(m)} \right| \\
 &\leq \sqrt{\mathbb{E} \left| \sum_{j=1}^m Y_j^{(m)} - \mathbb{E}Y_j^{(m)} \right|^2} + \sum_{j=1}^m \mathbb{E} |Z_j^{(m)} - \mathbb{E}Z_j^{(m)}| \\
 &\leq \sqrt{m\mathbb{E} |Y_1^{(1)} - \mathbb{E}Y_1^{(1)}|^2} + 2m\mathbb{E}|Z_1^{(1)}| \leq \sqrt{m2M\mathbb{E}|X_1^{(1)}|} + 2m\varepsilon,
 \end{aligned} \tag{5.4}$$

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 5.2 Assume that the error model is accordingly to Assumption 2.4. Then there holds

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

Proof. It holds that

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

Thus by Proposition 5.2 it suffices to show that

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

Let us first assume that Assumption 2.4.1 holds true. Then, by Markov's inequality

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

Now with

$$X_{jn}^{(m)} := \left(\frac{\sum_{i=1}^n \delta_{ij}^{(m)}}{\sqrt{n\mathbb{E}\delta_{i1}^{(m)2}}} \right)^2 - 1$$

it holds that $(X_{jn}^{(m)})_{j=1, \dots, m}, m \in \mathbb{N}$ are i.i.d and $\mathbb{E}X_{jn}^{(m)} = 0, \mathbb{E}|X_{jn}^{(m)}| = 2 < \infty$. We proceed similarly as at the end of the proof of Proposition 5.2 and show that

$$\mathbb{E} \left| \frac{1}{m} \sum_{j=1}^m X_{jn}^{(m)} \right| \rightarrow 0 \quad (5.5)$$

as $m \rightarrow \infty$ (uniformly in $n \in \mathbb{N}$), where we face an additional technical difficulty due to the dependence on n . Let $\varepsilon > 0$ and Z be a standard Gaussian (thus $\mathbb{E}Z^2 = 1$ in particular). Then for M large enough it holds that

$$\mathbb{E} \left[\mathcal{X}_{\{|Z^2-1| \geq M\}} \right] \leq \frac{\varepsilon}{4} \quad (5.6)$$

$$\mathbb{E} \left[Z^2 \mathcal{X}_{\{|Z^2-1| < M\}} \right] \geq \mathbb{E}[Z^2] - \frac{\varepsilon}{4} = 1 - \frac{\varepsilon}{4}. \quad (5.7)$$

By the standard central limit theorem for real valued random variables it holds that

$$\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{i1}^{(1)2}]}} \rightarrow Z$$

weakly as $n \rightarrow \infty$. Since

$$f_1 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \mathcal{X}_{\{|x^2-1| \geq M\}},$$

$$f_2 : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2 \mathcal{X}_{\{|x^2-1| < M\}}$$

are bounded functions whose set of discontinuities has Lebesgue measure 0 it holds that

$$\mathbb{E} \left[f_p \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{i1}^{(1)2}]}} \right) \right] \rightarrow \mathbb{E}[f_p(Z)]$$

as $n \rightarrow \infty$ for $p = 1, 2$ by Portmanteaus lemma (see e.g. Klenke (2013)). Thus by (5.6) there exists a n^* such that

$$\begin{aligned} \mathbb{E} \left[f_1 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] &\leq \mathbb{E}[f_1(Z)] + \left| \mathbb{E} \left[f_1 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) - f_1(Z) \right] \right| \leq \frac{\varepsilon}{2} \\ \mathbb{E} \left[f_2 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] &\geq \mathbb{E}[f_2(Z)] - \left| \mathbb{E} \left[f_2 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) - f_2(Z) \right] \right| \geq 1 - \frac{\varepsilon}{2} \end{aligned}$$

for all $n \geq n^*$ and $p = 1, 2$. We again set $Y_{jn}^{(m)} := X_{jn}^{(m)} \chi_{\{|X_{jn}^{(m)}| \leq M\}}$ and $Z_{jn}^{(m)} := X_{jn}^{(m)} \chi_{\{|X_{jn}^{(m)}| > M\}}$ and define

$$f_3 := \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2 \chi_{\{|x^2 - 1| \geq M\}}.$$

Then

$$\begin{aligned} \mathbb{E}|Z_{1n}^{(n)}| &\leq \mathbb{E} \left[f_3 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] + \mathbb{E} \left[f_1 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] \\ &= \mathbb{E} \left[\left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right)^2 \right] - \mathbb{E} \left[f_2 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] \\ &\quad + \mathbb{E} \left[f_1 \left(\frac{\sum_{i=1}^n \delta_{i1}^{(1)}}{\sqrt{n\mathbb{E}[\delta_{11}^{(1)2}]}} \right) \right] \\ &\leq 1 - (1 - \frac{\varepsilon}{2}) + \frac{\varepsilon}{2} = \varepsilon, \end{aligned} \tag{5.8}$$

for all $n \geq n^*$, where we used that $f_2(x) = f_3(x) = x^2$ in the second step. With the same argumentation as in (5.4),

$$\begin{aligned}
\mathbb{E} \left| \sum_{j=1}^m X_{jn}^{(m)} \right| &\leq \mathbb{E} \left| \sum_{j=1}^m Y_{jn}^{(m)} - \mathbb{E} Y_{jn}^{(m)} \right| + \mathbb{E} \left| \sum_{j=1}^m Z_{jn}^{(m)} - \mathbb{E} Z_{jn}^{(m)} \right| \\
&\leq \sqrt{\mathbb{E} \left| \sum_{j=1}^m Y_{jn}^{(m)} - \mathbb{E} Y_{jn}^{(m)} \right|^2} + \sum_{j=1}^m \mathbb{E} \left| Z_{jn}^{(m)} - \mathbb{E} Z_{jn}^{(m)} \right| \\
&\leq \sqrt{m \mathbb{E} \left| Y_{1n}^{(1)} - \mathbb{E} Y_{1n}^{(1)} \right|^2} + 2m \mathbb{E} |Z_{1n}^{(1)}| \leq \sqrt{m 2M \mathbb{E} |X_{1n}^{(1)}|} + 2m\varepsilon \\
&\leq \sqrt{4mM} + 2m\varepsilon,
\end{aligned}$$

for all $n \geq n^*$, where we used that $\mathbb{E} |X_{1n}^{(1)}| \leq 2$ and (5.8) in the last step. Thus $\mathbb{E} |\sum_{j=1}^m X_{jm}^{(m)}|/m \leq 3\varepsilon$ for m, n large enough and sending ε to 0 proves the claim (5.5).

Now assume that Assumption 2.4.2 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

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

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. \square

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

PROPOSITION 5.3 For all $\varepsilon > 0$, there exist $m_0 \in \mathbb{N}$ 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 5.1.2 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

$$\begin{aligned} \left\| (P_m K R_\alpha^{(m)} - Id) P_m K \hat{x} \right\|^2 / \alpha &= \sum_{j=1}^m \left(F_\alpha(\sigma_j^{(m)^2}) \sigma_j^{(m)^2} - 1 \right)^2 \frac{\sigma_j^{(m)^2}}{\alpha} (\hat{x}, v_j^{(m)})^2 \\ &\leq \left(\sup_{\lambda > 0} \lambda^{\frac{v_0}{2}} |F_\alpha(\lambda) \lambda - 1| \right)^2 \|\hat{x}\|^2 \sum_{j=1}^M \frac{\sigma_j^{(m)^{2(1-v_0)}}}{\alpha} \\ &\quad + \left(\sup_{\lambda > 0} \lambda^{\frac{1}{2}} |F_\alpha(\lambda) \lambda - 1| \right)^2 \frac{\sum_{j=M+1}^m (\hat{x}, v_j^{(m)})^2}{\alpha} \\ &\leq C_{v_0}^2 M \sigma_M^{(m)^{2(1-v_0)}} \|\hat{x}\|^2 \alpha^{v_0-1} + C_1^2 \sum_{l=M+1}^m (\hat{x}, v_l^{(m)})^2 \\ &\leq 2C_{v_0}^2 M \sigma_M^{(\infty)^{2(1-v_0)}} \|\hat{x}\|^2 \alpha^{v_0-1} + \varepsilon/2 \leq \varepsilon \end{aligned}$$

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 5.1.1 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 5.1,

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

for m large enough and by Lemma 5.1.2

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

for m sufficiently large. So Lemma 5.1.1 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,

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

Proposition 5.3 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 (5.1)

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

for m large enough. Putting it all together yields

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

for m sufficiently large, which together with $\lim_{\substack{m,n \rightarrow \infty \\ m/n \rightarrow 0}} \mathbb{P}(\Omega_{m,n}) = 1$ finishes the proof.

5.2 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.

5.2.1 Proof of Proposition 3.1. It holds that $\kappa(P_m) = \kappa(P_m|_{\mathcal{N}(P_m)^\perp})$. 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

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

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 Rump (2011) it holds that

$$1 - c \leq \sigma_j \left(\frac{A_m}{\|\eta_1^{(m)}\|^2} \right) \leq 1 + c, \quad (5.9)$$

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.

5.2.2 Proof of Proposition 3.2. The bounds c_m, C_m follow directly from Proposition 3.1. 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 exists 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, $\delta_m^{\text{disc}} = \|P_{\mathcal{N}(P_m)} y\| = \sqrt{\sum_{j=m+1}^{\infty} (y, w_j)^2} \rightarrow 0$ as $m \rightarrow \infty$.

5.2.3 Proof of Proposition 3.3. 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 $(v_j)_{j \in \mathbb{N}}$ is an orthonormal basis the claim follows with Proposition 3.1.

5.2.4 Proof of Proposition 3.4. The choice $c_m = C_m = 1$ follows from Proposition 3.1 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 $\xi_t \in [\frac{j-1}{m}, \frac{j}{m})$.

5.2.5 *Proof of Proposition 3.5.* It holds that

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

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 3.1. Let $y_m \in \mathcal{N}(P_m)^\perp$ be the interpolating spline of continuously differentiable \hat{y} . By the mean value theorem there exist $\xi_t, \zeta_t \in [\frac{j-1}{m-1}, \frac{j}{m-1})$ such that

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

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

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

5.2.6 *Proof of Theorem 3.6.* We use the bias-variance decomposition

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

as $m \rightarrow \infty$.

5.2.7 *Proof of Theorem 3.7.* The proof of Theorem 3.7 is more technical than the one of Theorem 2.5 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 5.3 Assume that the discretisation fulfills Assumption 2.2 and the error is accordingly to Assumption 2.4 with $p \geq 2$ in the case of Assumption 2.4.2. 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

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

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

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

$$\text{with } \bar{Y}_{n(m,\delta)}^{(m)} = \frac{1}{n(m,\delta)} \sum_{i=1}^{n(m,\delta)} \begin{pmatrix} Y_{i1}^{(m)} & \dots & Y_{im}^{(m)} \end{pmatrix}^T \text{ and } \delta_m := \delta \sqrt{\sum_{j=1}^m \frac{c_m^2}{m \sigma_j^2}}.$$

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 $\mathbb{E} \|P_m^+ \bar{Y}_{n(m,\delta)}^{(m)} - P_m^+ P_m \hat{y}\|^2$ since $c_m \leq \sigma_j^{(m)}$. We define

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

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

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

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

$$\mathbb{E} \delta_{m,n}^{meas2} = \sum_{j=1}^m \lambda_j \mathbb{E} \left(\sum_{l=1}^n \sum_{i=1}^n \frac{\delta_{il}^{(m)}}{n} (u_j^{(m)}, e_l^{(m)}) \right)^2 = \frac{\mathbb{E} \delta_{11}^{(m)2}}{n} \sum_{j=1}^m \lambda_j$$

and

$$\begin{aligned} \mathbb{P} \left(\left| \frac{\delta_{m,n(m,\delta)}^{meas2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon \right) &\geq \mathbb{P} \left(\left| \frac{\delta_{m,n(m,\delta)}^{meas2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon, n(m, \delta) \in I_{\varepsilon'} \right) \\ &\geq \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon, n(m, \delta) \in I_{\varepsilon'} \right) \\ &\geq 1 - \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas2} - \delta_m^2}{\delta_m^2} \right| > \varepsilon \right) - \mathbb{P}(n(m, \delta) \notin I_{\varepsilon'}). \end{aligned}$$

Since

$$\left| \frac{\delta_{m,n}^{meas2} - \delta_m^2}{\delta_m^2} \right| \leq \left(\left| \frac{\delta_{m,n}^{meas2} - \mathbb{E} \delta_{m,n}^{meas2}}{\mathbb{E} \delta_{m,n}^{meas2}} \right| + \left| \frac{\mathbb{E} \delta_{m,n}^{meas2} - \delta_m^2}{\mathbb{E} \delta_{m,n}^{meas2}} \right| \right) \frac{\mathbb{E} \delta_{m,n}^{meas2}}{\delta_m^2}$$

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 $\varepsilon' = \frac{3}{16} \varepsilon \leq 1/4$

$$\begin{aligned} & \mathbb{P} \left(\left| \frac{\delta_{m,n(m,\delta)}^{meas^2} - \delta_m^2}{\delta_m^2} \right| \leq \varepsilon \right) \\ & \geq 1 - \mathbb{P} \left(\sup_{n \in I_{\varepsilon'}} \left| \frac{\delta_{m,n}^{meas^2} - \mathbb{E} \delta_{m,n}^{meas^2}}{\mathbb{E} \delta_{m,n}^{meas^2}} \right| > \varepsilon(1 - \varepsilon') - \frac{\varepsilon'}{1 - \varepsilon'} \right) - \mathbb{P}(n(m, \delta) \notin I_{\varepsilon'}) \\ & \geq 1 - \mathbb{P} \left(\sup_{n \in I_{\frac{3}{16}\varepsilon}} \left| \frac{\delta_{m,n}^{meas^2} - \mathbb{E} \delta_{m,n}^{meas^2}}{\mathbb{E} \delta_{m,n}^{meas^2}} \right| > \varepsilon/2 \right) - \mathbb{P}(n(m, \delta) \notin I_{\frac{3}{16}\varepsilon}). \end{aligned} \quad (5.10)$$

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

PROPOSITION 5.4 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 holds that $\lfloor (1 + \varepsilon)\mu_m^\delta \rfloor \geq (1 + \varepsilon/2)\mu_m^\delta$ and

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

and the claim follows by Proposition 5.2. □

For the first term in (5.10) we will need the following proposition.

PROPOSITION 5.5 For $(X_l)_{l \in \mathbb{N}}$ i.i.d. 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

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

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

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

□

Finally, it holds that

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

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 $\mathcal{F}_n := \sigma \left(\delta_{ij}^{(m)}, i \leq n, j \leq m \right)$ for every fixed $m \in \mathbb{N}$. Now assume that Assumption 2.4.2 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

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

With $X_l := \sum_i \delta_{ij}^{(m)} / \sqrt{n \mathbb{E}\delta_{ij}^{(m)2}}$ Proposition 5.5 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 2.4.1, the Kolmogorov-Doob-inequality yields

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

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

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

Since $\mathbb{E} \left[V_1^{(1)4} \right] < \infty$, by Proposition 5.5 above and Jensen's inequality

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

For the second term

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

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,

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

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

$$\begin{aligned} \left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{y} \right\| \chi_{\Omega_m} &\leq \left\| P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - P_m^+ P_m \hat{y} \right\| \chi_{\Omega_m} + \left\| P_m^+ P_m \hat{y} - \hat{y} \right\| \chi_{\Omega_m} \\ &\leq \frac{\tau + 3C_0}{2C_0} \delta_m^{disc}. \end{aligned} \quad (5.11)$$

By Algorithm 2 it holds that

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

and because of $\frac{4\tau C_0}{\tau + 3C_0} > C_0$, (5.11) 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 Engl *et al.* (1996). With the same reasoning it follows that there exists a $L' \in \mathbb{R}$ such that

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

if there are $0 < v \leq v_0 - 1$ and $w \in \mathcal{X}$ with $K^+ \hat{y} = (K^* K)^{v/2} w$ and $\|w\| \leq \rho$. Lemma 5.3 implies that $\lim_{m \rightarrow \infty} \mathbb{P}(\Omega_m) = 1$, which concludes the proof. □

5.3 Proofs for Section 4

We start with the proof of Proposition 4.1.

Proof of Proposition 4.1. For simplicity we assume that $\hat{x} = 0 = \hat{y}$. Let $K : l^2(\mathbb{N}) \rightarrow l^2(\mathbb{N})$ be the diagonal operator with $K v_j = \sigma_j v_j := \alpha_j^{\frac{1}{2}} v_j$ (with $(v_j)_{j \in \mathbb{N}}$ the canonical basis) and let P_m be the discretisation along the singular basis, i.e. $(P_m y)_j = l_j^{(m)}(y) = (y, v_j)$ for all $j = 1, \dots, m$ with $m \in \mathbb{N}$ and $y \in l^2(\mathbb{N})$. Then

$$\begin{aligned}
 \mathbb{E}\|R_\alpha^{(m)} \bar{Y}_m^{(m)} - K^+ \hat{y}\|^2 &= \sum_{j=1}^m \frac{\sigma_j^2}{(\alpha_m + \sigma_j^2)} \mathbb{E} \left(\bar{Y}_m^{(m)} - P_m \hat{y}, v_j \right)^2 = \mathbb{E}[\delta_{11}^2] \sum_{j=1}^m \frac{(\alpha_j^{\frac{1}{4}})^2}{\left(\alpha_m + (\alpha_j^{\frac{1}{4}})^2\right)^2} \\
 &= \mathbb{E}[\delta_{11}^2] \sum_{j=1}^m \frac{\sqrt{\alpha_j}}{(\alpha_m + \sqrt{\alpha_j})^2} \geq \mathbb{E}[\delta_{11}^2] \sum_{j=1}^m \frac{\sqrt{\alpha_m}}{(\alpha_m + \sqrt{\alpha_m})^2} \\
 &\geq \frac{\mathbb{E}[\delta_{11}^2]}{\sqrt{\alpha_m}} \sum_{j=1}^m \frac{1}{(\sqrt{\alpha_m} + 1)^2} \geq \frac{m \mathbb{E}[\delta_{11}^2]}{\alpha_m (1 + \sqrt{\alpha_1})^2} \rightarrow \infty
 \end{aligned}$$

as $m \rightarrow \infty$. □

Proof of Theorem 4.2. We need the following auxiliary result, which we will afterwards use to confirm that the oscillations of the residual are too strong when the error distribution lacks of higher moments.

LEMMA 5.4 For $0 < \varepsilon < 1$ let $(\delta_{ij})_{i,j \in \mathbb{N}}$ be i.i.d. with density

$$f_\varepsilon(x) = \frac{c_\varepsilon}{|x|^{3+\varepsilon}} \mathcal{X}_{\{(-\infty, -b_\varepsilon] \cup [b_\varepsilon, \infty)\}}(x),$$

where $c_\varepsilon = \frac{\varepsilon^{1+\frac{\varepsilon}{2}}}{2(2+\varepsilon)^{\frac{\varepsilon}{2}}}$ and $b_\varepsilon = \sqrt{\frac{\varepsilon}{2+\varepsilon}}$. Then there exist $p_\varepsilon, \kappa_\varepsilon > 0$ such that

$$\mathbb{P} \left(\sum_{j=1}^m \left(\left(\frac{\sum_{i=1}^n \delta_{ij}}{\sqrt{n}} \right)^2 - 1 \right) \geq \kappa_\varepsilon m^{\frac{2-\varepsilon}{2+\varepsilon}} \right) \geq p_\varepsilon$$

for all $cn \leq m$ with m, n large enough.

Proof of Lemma 5.4. Straight forward computations show that f_ε is indeed the density of a probability distribution and that $\mathbb{E}[\delta_{ij}] = 0$ and $\mathbb{E}[\delta_{ij}^2] = 1$ and $\mathbb{P}(\delta_{ij} \geq x) = \frac{c_\varepsilon}{2+\varepsilon} \frac{1}{x^{2+\varepsilon}} = \mathbb{P}(|\delta_{ij}| \geq x)/2$ hold for $x \geq b_\varepsilon$.

Thus we may apply Corollary 1.1.2. from Vinogradov (1994) and obtain that there exist constants K_1, K_2, K_3 such that for any $n \in \mathbb{N}$ and all $y \geq \sqrt{K_1 n \log(n)}$ there holds

$$\begin{aligned}
 &\left| \mathbb{P} \left(\sum_{i=1}^n \delta_{i1} \geq y \right) - \frac{c_\varepsilon}{2+\varepsilon} \frac{n}{y^{2+\varepsilon}} \right| \\
 &\leq K_2 \frac{n}{y^{2+\varepsilon}} \left(\left(\frac{\sqrt{n}}{y} \right)^{\frac{2+\varepsilon}{3+\varepsilon}} + e^{-K_3 \left(\frac{y^2}{n} \right)^{\frac{1}{3+\varepsilon}}} \right) + n \sup_{x \geq y/3} \left| \mathbb{P}(\delta_{11} > x) - \frac{c_\varepsilon}{2+\varepsilon} \frac{1}{x^{2+\varepsilon}} \right| \\
 &= K_2 \frac{n}{y^{2+\varepsilon}} \left(\left(\frac{\sqrt{n}}{y} \right)^{\frac{2+\varepsilon}{3+\varepsilon}} + e^{-K_3 \left(\frac{y^2}{n} \right)^{\frac{1}{3+\varepsilon}}} \right)
 \end{aligned}$$

Therefore, by symmetry of the distribution of δ_{ij} we have for all $x \geq K_2 \log(n)$

$$\begin{aligned}
\mathbb{P}\left(\left(\frac{\sum_{i=1}^n \delta_{i1}}{\sqrt{n}}\right)^2 \geq x\right) &= \mathbb{P}\left(\left|\sum_{i=1}^n \delta_{i1}\right| \geq \sqrt{xn}\right) = 2\mathbb{P}\left(\sum_{i=1}^n \delta_{i1} \geq \sqrt{xn}\right) \\
&= \frac{2c_\varepsilon}{2+\varepsilon} \frac{n}{(\sqrt{nx})^{2+\varepsilon}} + \mathcal{O}\left(K_2 \frac{n}{(\sqrt{nx})^{2+\varepsilon}} \left(\left(\frac{\sqrt{n}}{\sqrt{nx}}\right)^{\frac{2+\varepsilon}{3+\varepsilon}} + e^{-K_3 \left(\frac{(\sqrt{nx})^2}{n}\right)^{\frac{1}{3+\varepsilon}}}\right)\right) \\
&= \frac{2c_\varepsilon}{2+\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}} x^{1+\frac{\varepsilon}{2}}} + \frac{1}{n^{\frac{\varepsilon}{2}} x^{1+\frac{\varepsilon}{2}}} \mathcal{O}\left(x^{-\frac{2+\varepsilon}{3+\varepsilon}} + e^{-x^{\frac{1}{3+\varepsilon}}}\right) \\
&= \frac{2c_\varepsilon}{2+\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}} x^{1+\frac{\varepsilon}{2}}} (1 + \mathcal{O}(1)) \quad (\text{for } x \rightarrow \infty). \tag{5.12}
\end{aligned}$$

Let $x_n \geq K_2 \log(n) + 1$ (will be specified later) and $X_{jn} := \left(\frac{\sum_{i=1}^n \delta_{ij}}{\sqrt{n}}\right)^2$. We truncate and split the sum in two parts

$$\sum_{j=1}^m \left(\left(\frac{\sum_{i=1}^n \delta_{ij}}{\sqrt{n}} \right)^2 - 1 \right) = \sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} + \sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}}.$$

The second term contains only the extremes of the sum and we will show that here both parts will contribute to the overall sum (note that when sufficiently high moments exist (at least a fourth moment) one could show that the overall sum is dominated by the first part). We first treat the second term. Since $(X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} \geq x_n - 1 > 0$ (for n sufficiently large) we have that for $t_{n,m} \geq x_n - 1$

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} > t_{n,m}\right) &\geq \mathbb{P}\left(\max_{j=1, \dots, m} (X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} > t_{n,m}\right) \\
&= 1 - \mathbb{P}\left(\max_{j=1, \dots, m} (X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} \leq t_{n,m}\right) \\
&= 1 - \mathbb{P}\left((X_{1n} - 1) \mathcal{X}_{\{X_{1n} > x_n\}} \leq t_{n,m}\right)^m = 1 - (1 - \mathbb{P}(X_{1n} - 1 \geq t_{n,m}))^m \\
&= 1 - \left(1 - \frac{2c_\varepsilon}{2+\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}} t_{n,m}^{1+\frac{\varepsilon}{2}}}\right)^m (1 + \mathcal{O}(1)). \tag{5.13}
\end{aligned}$$

For the remaining term we need the first three moments

$$\begin{aligned}
\mu_n &:= \mathbb{E}\left[(X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}}\right], \\
\sigma_n &:= \sqrt{\mathbb{E}\left[\left((X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} - \mu_n\right)^2\right]}, \\
\rho_n &:= \mathbb{E}\left[\left|(X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} - \mu_n\right|^3\right].
\end{aligned}$$

We claim that if $x_n/\log(n) \rightarrow 0$

$$\mu_n = -\frac{4c_\varepsilon(1+\mathcal{O}(1))}{(2+\varepsilon)\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}}x_n^{\frac{\varepsilon}{2}}} \quad (5.14)$$

$$\sigma_n^2 \geq (1+\mathcal{O}(1))2c_\varepsilon \frac{2-\varepsilon}{2+\varepsilon} \frac{x_n^{1-\frac{\varepsilon}{2}}}{n^{\frac{\varepsilon}{2}}} \quad (5.15)$$

$$\rho_n \leq (1+\mathcal{O}(1))3c_\varepsilon \frac{4-\varepsilon}{2+\varepsilon} \frac{x_n^{2-\frac{\varepsilon}{2}}}{n^{\frac{\varepsilon}{2}}}, \quad (5.16)$$

for $n \rightarrow \infty$. We will prove the assertions (5.14) - (5.16) with (5.12) and Theorem 12.1 of Gut (2013). Note that $\varepsilon < 2$ and X_{jn} is positive. For (5.14) because of $\mathbb{E}[X_{jn} - 1] = 0$ there holds

$$\begin{aligned} \mu_n &= \mathbb{E} \left[(X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} \right] = -\mathbb{E} \left[(X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} \right] = -\mathbb{E} \left[X_{jn} \mathcal{X}_{\{X_{jn} > x_n\}} \right] + \mathbb{E} \left[\mathcal{X}_{\{X_{jn} > x_n\}} \right] \\ &= -\int_{x_n}^{\infty} \mathbb{P}(X_{jn} > t) dt + \mathbb{P}(X_{jn} > x_n) = \frac{2c_\varepsilon(1+\mathcal{O}(1))}{2+\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}}} \left(-\int_{x_n}^{\infty} \frac{1}{t^{1+\frac{\varepsilon}{2}}} dt + \frac{1}{x_n^{1+\frac{\varepsilon}{2}}} \right) \\ &= -\frac{4c_\varepsilon(1+\mathcal{O}(1))}{(2+\varepsilon)\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}}x_n^{\frac{\varepsilon}{2}}} \quad (\text{for } n \rightarrow \infty). \end{aligned}$$

Further, we obtain

$$\begin{aligned} \mathbb{E} \left[(X_{jn} - 1)^2 \mathcal{X}_{\{X_{jn} \leq x_n\}} \right] &\geq \mathbb{E} \left[(X_{jn} - 1)^2 \mathcal{X}_{\{K_1 \log(n) \leq X_{jn} \leq x_n\}} \right] = 2 \int_{K_1 \log(n)-1}^{x_n-1} t \mathbb{P}(X_{jn} - 1 > t) dt \\ &= \frac{4c_\varepsilon(1+\mathcal{O}(1))}{2+\varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}}} \int_{K_1 \log(n)}^{x_n} t^{-\frac{\varepsilon}{2}} dt = (1+\mathcal{O}(1))2c_\varepsilon \frac{2-\varepsilon}{2+\varepsilon} \frac{x_n^{1-\frac{\varepsilon}{2}}}{n^{\frac{\varepsilon}{2}}} \end{aligned}$$

for $n \rightarrow \infty$, which together with (5.14) yields (5.15). Finally, a similar reasoning proves

$$\mathbb{E} \left[|X_{jn} - 1|^3 \mathcal{X}_{\{X_{jn} \leq x_n\}} \right] \geq (1+\mathcal{O}(1))3c_\varepsilon \frac{4-\varepsilon}{2+\varepsilon} \frac{x_n^{2-\frac{\varepsilon}{2}}}{n^{\frac{\varepsilon}{2}}},$$

which in turn implies (5.16).

Now

$$\begin{aligned} \mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} \leq t \right) &= \mathbb{P} \left(\frac{\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} - m\mu_n}{\sqrt{m\sigma_n^2}} \sqrt{m\sigma_n^2} + m\mu_n \leq t \right) \\ &= \mathbb{P} \left(\frac{\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} - m\mu_n}{\sqrt{m\sigma_n^2}} \leq \frac{t - m\mu_n}{\sqrt{m\sigma_n^2}} \right). \end{aligned}$$

By The Berry-Esseen Theorem (Theorem 6.1 in Gut (2013)), there exist $C > 0$ such that

$$\mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} \leq t \right) \leq \Phi \left(\frac{t - m\mu_n}{\sqrt{m\sigma_n^2}} \right) + C \frac{\rho_n}{\sigma_n^3 \sqrt{m}},$$

where Φ is the cumulative distribution function of a standard Gaussian random variable. With (5.14)-(5.16) we see that

$$\begin{aligned} \frac{m\mu_n}{\sqrt{m\sigma_n^2}} &= -(1 + \mathcal{O}(1)) \frac{1}{\varepsilon} \sqrt{\frac{8c_\varepsilon}{4 - \varepsilon^2}} \frac{m^{\frac{1}{2}}}{n^{\frac{\varepsilon}{4}} x_n^{\frac{1}{2} + \frac{\varepsilon}{4}}}, \\ \frac{\rho_n}{\sigma_n^3 \sqrt{m}} &= (1 + \mathcal{O}(1)) 3(4 - \varepsilon) \sqrt{\frac{2 + \varepsilon}{8c_\varepsilon(2 - \varepsilon)^3}} \frac{n^{\frac{\varepsilon}{4}} x_n^{\frac{1}{2} + \frac{\varepsilon}{4}}}{m^{\frac{1}{2}}}. \end{aligned}$$

Therefore, for

$$x_n := \left(\frac{\eta \sqrt{m}}{n^{\frac{\varepsilon}{4}}} \right)^{\frac{1}{\frac{1}{2} + \frac{\varepsilon}{4}}} = \eta^{\frac{4}{2 + \varepsilon}} \frac{m^{\frac{2}{2 + \varepsilon}}}{n^{\frac{\varepsilon}{2 + \varepsilon}}},$$

there exist constants $\beta_\varepsilon, \gamma_\varepsilon > 0$ such that

$$\mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} \leq x_n\}} \leq 2m\mu_n \right) \leq \Phi \left(-\frac{\beta_\varepsilon}{\eta} \right) + \gamma_\varepsilon \eta$$

for all m, n large enough, since with the above choice because of $n \leq cm$ and $\varepsilon < 1$ there holds $x_n / \log(n) \rightarrow 0$. Note that the right hand side does not depend on m and n . There holds

$$-m\mu_n = (1 + \mathcal{O}(1)) \frac{4c_\varepsilon}{\eta^{\frac{2\varepsilon}{2 + \varepsilon}} (2 + \varepsilon) \varepsilon} \frac{m^{\frac{2}{2 + \varepsilon}}}{n^{\frac{\varepsilon}{2 + \varepsilon}}},$$

thus for η small we have $-m\mu_n > x_n$. Consequently, with $t_{m,n} = -3m\mu_n$ (5.13) becomes

$$\begin{aligned} \mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \mathcal{X}_{\{X_{jn} > x_n\}} > t_{n,m} \right) &\geq 1 - \left(1 - \frac{2c_\varepsilon}{2 + \varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}} t_{n,m}^{\frac{1 + \varepsilon}{2}}} (1 + \mathcal{O}(1)) \right)^m \\ &= 1 - \left(1 - \frac{2c_\varepsilon}{2 + \varepsilon} \frac{1}{n^{\frac{\varepsilon}{2}} \left((1 + \mathcal{O}(1)) 3m^{\frac{2}{2 + \varepsilon}} n^{-\frac{\varepsilon}{2 + \varepsilon}} \eta^{-\frac{2\varepsilon}{2 + \varepsilon}} \right)^{1 + \frac{\varepsilon}{2}}} \right)^m \\ &= 1 - \left(1 - (1 + \mathcal{O}(1)) \frac{2c_\varepsilon}{2 + \varepsilon} \frac{\eta^\varepsilon}{m} \right)^m \geq 1 - e^{-(1 + \mathcal{O}(1)) \frac{2c_\varepsilon}{2 + \varepsilon} \eta^\varepsilon} \\ &\geq \zeta_\varepsilon \eta^\varepsilon, \end{aligned}$$

for some constant $\zeta_\varepsilon > 0$, when m, n and η are sufficiently large respectively small. Putting all together we obtain

$$\begin{aligned}
 & \mathbb{P} \left(\sum_{j=1}^m \left(\left(\frac{\sum_{i=1}^n \delta_{ij}}{\sqrt{n}} \right)^2 - 1 \right) \geq -m\mu_n \right) \\
 & \geq \mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \chi_{\{X_{jn} \leq x_n\}} \geq 2m\mu_n, \sum_{j=1}^m (X_{jn} - 1) \chi_{\{X_{jn} > x_n\}} \geq -3m\mu_n \right) \\
 & \geq 1 - \mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \chi_{\{X_{jn} \leq x_n\}} < 2m\mu_n \right) - \mathbb{P} \left(\sum_{j=1}^m (X_{jn} - 1) \chi_{\{X_{jn} > x_n\}} < -3m\mu_n \right) \\
 & \geq 1 - \left(\Phi \left(-\frac{\beta_\varepsilon}{\eta} \right) + \gamma_\varepsilon \eta \right) - (1 - \zeta_\varepsilon \eta^\varepsilon) = \zeta_\varepsilon \eta^\varepsilon - \gamma_\varepsilon \eta - \Phi \left(-\frac{\beta_\varepsilon}{\eta} \right) \geq \zeta_\varepsilon \eta^\varepsilon - \gamma_\varepsilon \eta - e^{-\frac{\beta_\varepsilon^2}{\eta^2}} \geq p_\varepsilon
 \end{aligned}$$

for some $p_\varepsilon > 0$ for $\eta = \eta_\varepsilon$ sufficiently small and fixed (since $\varepsilon < 1$). Finally, the assertion follows with

$$-m\mu_n = (1 + \mathcal{O}(1)) \frac{4c_\varepsilon c^{\frac{\varepsilon}{2+\varepsilon}} m^{\frac{2}{2+\varepsilon}}}{\eta^{\frac{2\varepsilon}{2+\varepsilon}} (2 + \varepsilon) \varepsilon n^{\frac{\varepsilon}{2+\varepsilon}}} \geq \kappa_\varepsilon m^{\frac{2-\varepsilon}{2+\varepsilon}},$$

which holds for some $\kappa_\varepsilon > 0$ and all $n \leq m/c$ with m, n large enough. \square

We come to the main proof of Theorem 4.2 and first look at the case where $\tau > 1$ and $\hat{x} = K^+ \hat{y} = \sum_{j=1}^\infty j^{-1} v_j$. Set

$$C := \sqrt{\sum_{j=1}^\infty j^{-2}}_{j^{-q} \leq \frac{qc(\tau-1)^2}{\pi^2}}$$

Now, for $\alpha > 0$ we have

$$\begin{aligned}
 \left\| \left(P_m K R_\alpha^{(m)} - Id_{\mathbb{R}^m} \right) \bar{Y}_n^{(m)} \right\| & \leq \left\| \left(P_m K R_\alpha^{(m)} - Id_{\mathbb{R}^m} \right) P_m \hat{y} \right\| + \left\| \left(P_m K R_\alpha^{(m)} - Id_{\mathbb{R}^m} \right) \left(\bar{Y}_n^{(m)} - P_m \hat{y} \right) \right\| \\
 & \leq \sqrt{\sum_{j=1}^m (\hat{y}, v_j)^2}_{j^{-q} < \alpha} + \|\bar{Y}_n^{(m)} - P_m \hat{y}\| = \sqrt{\sum_{j=1}^m j^{-q-2}}_{j^{-q} < \alpha} + \|\bar{Y}_n^{(m)} - P_m \hat{y}\| \\
 & \leq \sqrt{\alpha \sum_{j=1}^\infty j^{-q-2}} + \|\bar{Y}_n^{(m)} - P_m \hat{y}\| \leq \sqrt{\frac{\alpha \pi^2}{6}} + \|\bar{Y}_n^{(m)} - P_m \hat{y}\|.
 \end{aligned}$$

This, together with the defining relation of the discrepancy principle

$$\tau \delta_{m,n} < \left\| \left(P_m K R_{\alpha_{m,n}/q}^{(m)} - Id_{\mathbb{R}^m} \right) \bar{Y}_n^{(m)} \right\|,$$

Lemma 5.2, Proposition 5.2 and $\mathbb{E}[\delta_{11}^2] = 1$ ultimately yields

$$\mathbb{P} \left(\alpha_{m,n} \geq \frac{q(\tau-1)^2 m}{\pi^2 n} \right) \rightarrow 1$$

as $m \rightarrow \infty$ (uniformly in $n \in \mathbb{N}$). Consequently, we have that

$$\begin{aligned} \mathbb{P}\left(\|R_{\alpha_{m,n}}\bar{Y}_n^{(m)} - \hat{x}\| \geq C\right) &= \mathbb{P}\left(\sqrt{\sum_{\substack{j=1 \\ \alpha_{m,n} \leq j^{-q}}}^m \frac{(\bar{Y}_n^{(m)} - P_m \hat{y})^2}{j^q} + \sum_{\substack{j=1 \\ \alpha_{m,n} > j^{-q}}}^{\infty} (\hat{x}, e_j)^2} \geq C\right) \\ &\geq \mathbb{P}\left(\sqrt{\sum_{\substack{j=1 \\ \alpha_{m,n} > j^{-q}}}^{\infty} j^{-2}} \geq C, \alpha_{m,n} \geq \frac{q}{\pi^2} \frac{m}{n}\right) = \mathbb{P}\left(\alpha_{m,n} \geq \frac{q(\tau-1)^2 m}{\pi^2 n}\right) \rightarrow 1 \end{aligned}$$

as $m, n \rightarrow \infty$ with $m/n \geq c$, where we used the definition of C in the third step. Thus Theorem 4.1 is proved in the case that $\tau > 1$.

Now we discuss the case, where the right hand side $\tau \delta_{m,n}$ in line 6 of Algorithm 1 is replaced with $\sqrt{\frac{1+\sqrt{m}}{n}}$ and where $\hat{x} = K^+ \hat{y} = 0$. Let $\varepsilon < \frac{2}{11}$ in the definition of the density of the error distribution of the δ_{ij} . Then there exists γ with $\frac{5}{6} < \gamma < \frac{2-\varepsilon}{2+\varepsilon}$ and by Lemma 5.4 there exist $p_\varepsilon > 0$ such that

$$\mathbb{P}\left(\sum_{j=m^{\frac{1}{2}+1}}^m \left(\frac{\sum_{i=1}^n \delta_{ij}}{n}\right)^2 \geq \frac{m-m^{\frac{1}{2}}}{n} + \frac{(m-m^{\frac{1}{2}})^\gamma}{n}\right) \geq p_\varepsilon \quad (5.17)$$

for all m, n large enough with $n \leq m$. Since

$$m - m^{\frac{1}{2}} + (m - m^{\frac{1}{2}})^\gamma \geq m - m^{\frac{1}{2}} + \frac{1}{2^\gamma} m^\gamma = m + m^{\frac{1}{2}} + \frac{m^{\frac{1}{2}}}{2^\gamma} (m^{\frac{\gamma}{2}-1} - 2^\gamma) > m + m^{\frac{1}{2}}$$

for m sufficiently large there holds for $\alpha_{m,n}^* := (m^{\frac{1}{2}})^{-q} = \sigma_{\sqrt{m}}^2$

$$\mathbb{P}\left(\left\|P_m K R_{\alpha_{m,n}^*} \bar{Y}_n^{(m)} - \bar{Y}_n^{(m)}\right\| = \sum_{j=m^{\frac{1}{2}+1}}^m \left(\frac{\sum_{i=1}^n \delta_{ij}}{n}\right)^2 > \frac{m+m^{\frac{1}{2}}}{n} = (\tau_m \delta_{n,m}^{est})^2\right) \geq p_\varepsilon$$

for all m, n large enough with $n \leq m$. By monotonicity of the spectral cut-off regularisation we deduce that

$$\mathbb{P}\left(\|P_m K R_{\alpha}^{(m)} \bar{Y}_n^{(m)} - \bar{Y}_n^{(m)}\| > \tau_m \delta_{n,n}^{est}, \forall \alpha_m^* \leq \alpha \leq \|K\|^2\right) \geq p_\varepsilon,$$

thus $\mathbb{P}(\alpha_{m,n} \leq \alpha_m^*) \geq p_\varepsilon$ for all m, n large enough with $n \leq m$. Further,

$$\begin{aligned}
 & \mathbb{P} \left(\left\| R_{\alpha_{m,n}}^{(m)} \bar{Y}_n^{(m)} - \hat{x} \right\|^2 \geq \frac{m^{\frac{1+q}{2}}}{n} \right) \geq \mathbb{P} \left(\sum_{j=1}^{\frac{m}{2}} j^q \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 \geq \frac{m^{\frac{1+q}{2}}}{n}, \alpha_{m,n} \leq \alpha_m^* \right) \\
 & \geq \mathbb{P} \left(\sum_{j=\frac{m}{2}}^{\frac{m}{2}} j^q \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 \geq \frac{m^{\frac{1+q}{2}}}{n}, \alpha_{m,n} \leq \alpha_m^* \right) \\
 & \geq \mathbb{P} \left(\frac{m^{\frac{q}{2}}}{2^q} \sum_{j=\frac{m}{2}}^{\frac{m}{2}} \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 \geq \frac{m^{\frac{1+q}{2}}}{n}, \alpha_{m,n} \leq \alpha_m^* \right) \\
 & \geq 1 - \mathbb{P}(\alpha_{m,n} > \alpha_m^*) - \mathbb{P} \left(\sum_{j=\frac{m}{2}}^{\frac{m}{2}} \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 < \frac{2^q m^{\frac{1}{2}}}{n} \right) \geq p_\varepsilon - \mathbb{P} \left(\sum_{j=\frac{m}{2}}^{\frac{m}{2}} \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 < \frac{2^q m^{\frac{1}{2}}}{n} \right)
 \end{aligned}$$

for all m, n large enough with $n \leq m$. Finally, with (5.5) it follows that

$$\mathbb{P} \left(\sum_{j=\frac{m}{2}}^{\frac{m}{2}} \left(\frac{\sum_{i=1}^n \delta_{ij}}{n} \right)^2 < \frac{2^q m^{\frac{1}{2}}}{n} \right) \rightarrow 0$$

as $m \rightarrow \infty$ (uniformly in $n \in \mathbb{N}$), which finishes the proof of Theorem 4.2 since $q > 1$. \square

Proof of Proposition 4.3. Let $\sigma_1^{(m)}, \dots, \sigma_m^{(m)}$ and $v_1^{(m)}, \dots, v_m^{(m)}$ denote the singular values respectively vectors of KP_m . Clearly, $v_j^{(m)} = v_j$ and $\sigma_j^{(m)} = j^{-1}$ for all $m \in \mathbb{N}$ and $j = 1, \dots, m-1$. Moreover, the ansatz $v_m^{(m)} = av_m + bv_{\lceil e^m \rceil}$ in

$$(P_m K)^* P_m K v_m^{(m)} = \sigma_m^{(m)2} v_m^{(m)}$$

yields

$$\sigma_m^{(m)} = \sqrt{m^{-2} e^{-2m} + \lceil e^m \rceil^{-2} (1 - e^{-2m})} \quad \text{and} \quad v_m^{(m)} = \frac{m^{-1} e^{-m} v_m + \lceil e^m \rceil^{-1} \sqrt{1 - e^{-2m}} v_{\lceil e^m \rceil}}{\sigma_m^{(m)}}.$$

Now

$$\begin{aligned}
 \sum_{j=1}^{m-1} (\hat{x}, v_j) v_j + (\hat{x}, v_m^{(m)}) v_m^{(m)} &= \sum_{j=1}^m (\hat{x}, v_j^{(m)}) v_j^{(m)} = P_{\mathcal{N}(P_m K)^\perp} \hat{x} = ((P_m K)^* P_m K)^{\frac{v_m}{2}} w_m \\
 &= \sum_{j=1}^m \sigma_j^{(m) v_m} (w_m, v_j^{(m)}) v_j^{(m)} = \sum_{j=1}^{m-1} \sigma_j^{v_m} (w_m, v_j) v_j + \sigma_m^{(m) v_m} (w_m, v_m^{(m)}) v_m^{(m)}.
 \end{aligned}$$

Consequently,

$$\begin{aligned} \rho_m \geq |(w_m, v_m^{(m)})| &\stackrel{!}{=} \frac{(\hat{x}, v_m^{(m)})}{\sigma_m^{(m)v_m}} = \frac{\left(\sum_{j=1}^{\infty} j^{-2} v_j, m^{-1} e^{-m} v_m + \lceil e^m \rceil^{-1} \sqrt{1 - e^{-2m}} v_{\lceil e^m \rceil}\right)}{\sigma_m^{(m)1+v}} \\ &= \frac{m^{-3} e^{-m} + \lceil e^m \rceil^{-3} \sqrt{1 - e^{-2m}}}{(m^{-2} e^{-2m} + \lceil e^m \rceil^{-2} (1 - e^{-2m}))^{\frac{1+v_m}{2}}} \geq \frac{m^{-3} e^{-m}}{(2e^{-2m})^{\frac{1+v_m}{2}}} = \frac{e^{v_m m}}{2^{\frac{1+v_m}{2}} m^3} \end{aligned}$$

and the proof of the proposition can be finished by a case-by-case analysis. \square

6. 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` Hansen (1994). 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 2.2. 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 6.3 we will also consider discretisation by hat functions to give an example with non-orthogonal discretisation. We chose a shifted generalised Pareto distribution for the distribution of the measurement error, i.e. $\delta_{ij}^{(m)} = Z_{ij}^{(m)} - EZ_{ij}^{(m)}$, where $Z_{ij}^{(m)}$ are i.i.d and follow a generalised Pareto distribution (`gprnd(l, sigma, theta, 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 2.4.1 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.

6.1 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 6.2 the resulting relative errors $\left\| R_{\alpha_m}^{(m)} \bar{Y}_n^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ 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.

6.2 (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 6.2 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 (3.1) is quite coarse. Together with the choice $\tau = 1.2$ this implies 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 6.2 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.

6.3 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 2.3). 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,

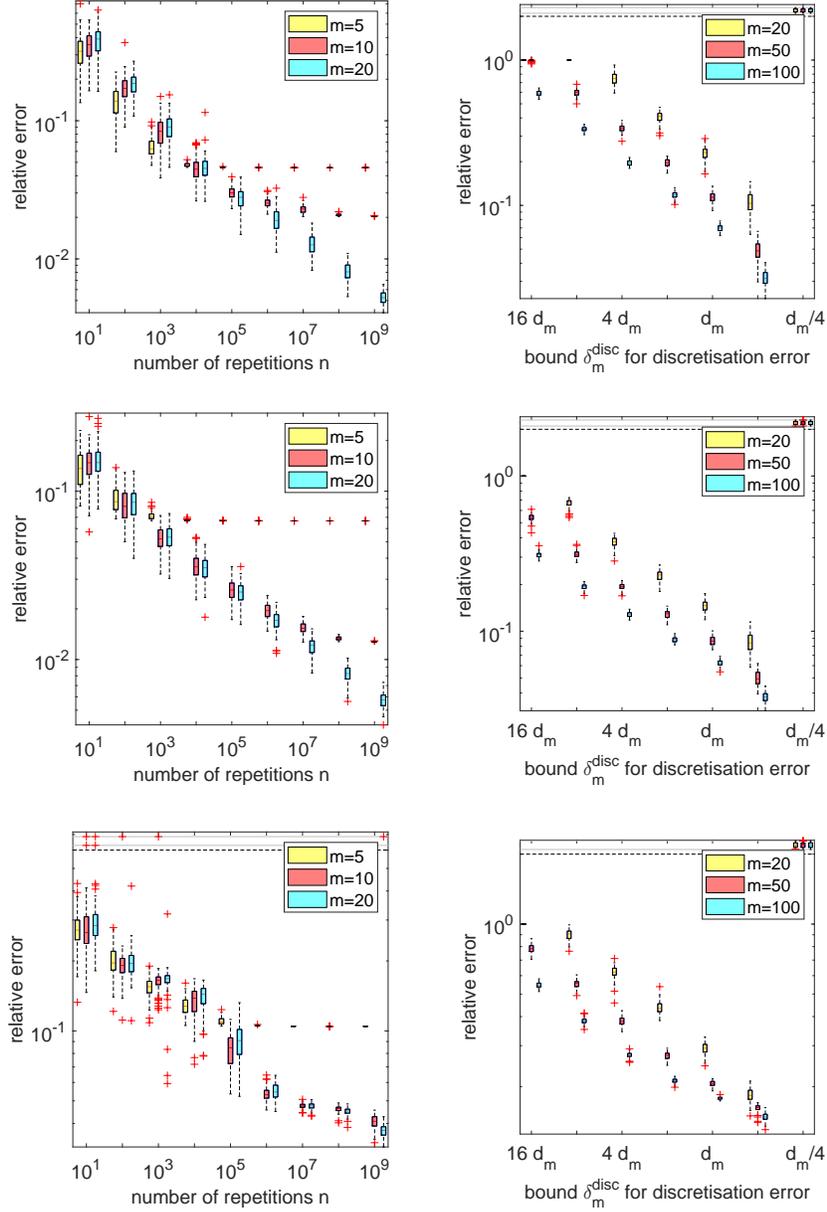


FIG. 1. Results of approach (1.2) and (1.6) with the discrepancy principle as implemented in Algorithm 1 (left column) or 2 (right column) respectively, for 'phillips' (first row), 'gravity' (second row) and 'shaw' (third row) visualised as boxplots for 100 independent runs. *Left column:* Relative errors $\left\| R_{\alpha_m, n}^{(m)} \bar{Y}_n^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ against number of repetitions n for several numbers of measurement channels m . *Right column:* Relative errors $\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ 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 $d_m := \|\hat{y} - P_m^+ P_m \hat{y}\|$.

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

and

$$\begin{pmatrix} y_{m_\infty-(k+1)} \\ \dots \\ y_{m_\infty} \end{pmatrix} \mapsto \frac{1}{\sqrt{\sum_{i=1}^{k+1} a_i^2}} (a_1y_{m_\infty-(k+1)} + \dots + a_{k+1}y_{m_\infty}) e_m.$$

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 6.3 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 indicates that here no smoothness got lost through discretisation in contrast to Proposition 4.3.

7. 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 a convergence rate under additional knowledge of the discretisation error. We want to pronounce some important outstanding questions. First, one could drop the simplification that one has an equal number of measurements on each measurement channel and try to distribute a fixed total number of measurements on the measurement channels in an optimal way (see also Mathé & Pereverzev (2017)). Further, the discretisation considered in this article entered the problem through discretised measurements. In particular, this is determined by the practical problem and the way the data is measured or acquired. In order to solve the problem numerically 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. Also, it would be desirable to better understand the interplay between the discretised and the infinite-dimensional problem, e.g. regarding the smoothness of the true solution relative to the former and the latter respectively. Hereby an important open question is to derive natural and verifiable conditions that rigorously guarantee convergence rates also for the approach with finite-dimensional residuum. Finally, it is worth investigating whether it is possible to modify the discrepancy principle to attain optimal convergence rates (in the statistical setting) in our general framework (see Jahn (2021b)).

REFERENCES

BABUŠKA, I. & OSBORN, J. (1991) Eigenvalue problems.

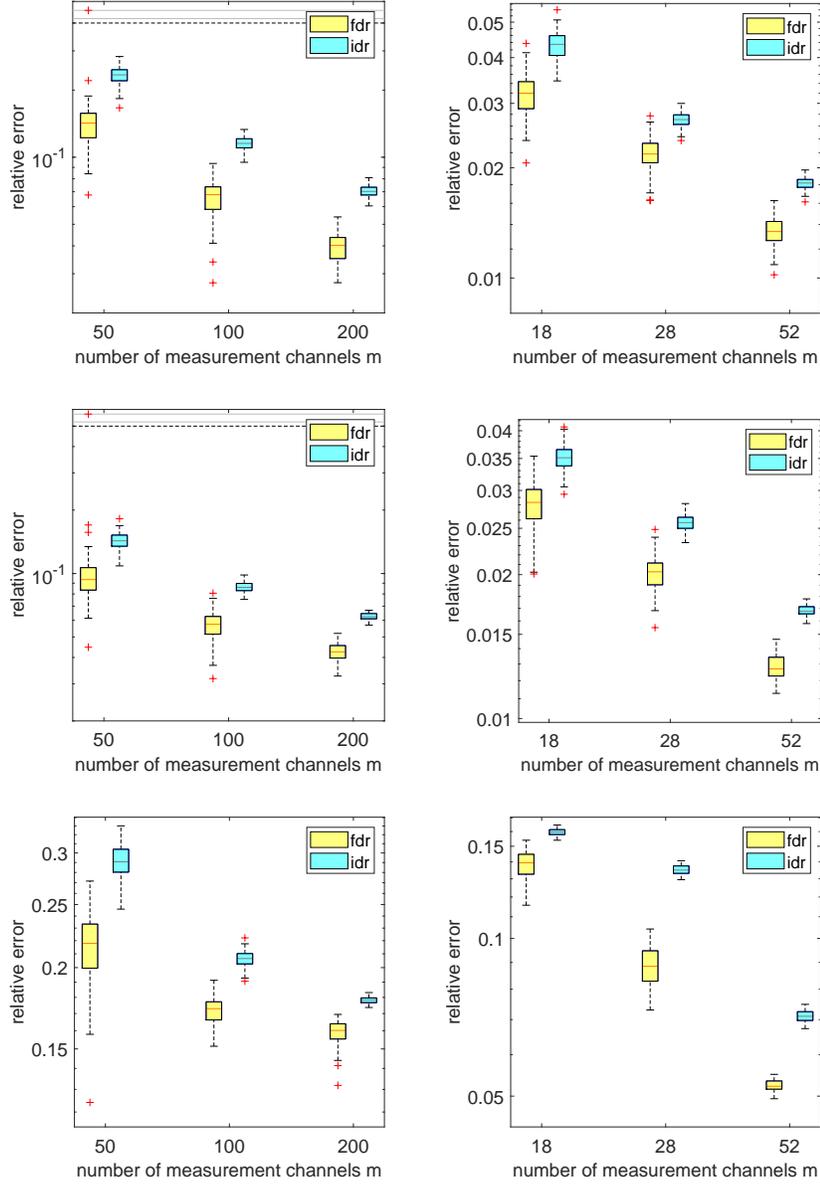


FIG. 2. Direct comparison of both approaches (1.2) (fdr) and (1.6) (idr) with discrepancy principle as implemented in Algorithm 1 and 2 for 'phillips' (first line), 'gravity' (second line) and 'shaw' (third line). For the discretisation of the measurements either box functions (first column) or hat functions (second column) are used. The relative errors $\left\| R_{\alpha_m, n(m, \delta_m^{disc})} \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ (fdr) and $\left\| R_{\alpha_m} P_m^+ \bar{Y}_{n(m, \delta_m^{disc})}^{(m)} - \hat{x} \right\| / \|\hat{x}\|$ (idr) are plotted against the number of measurement channels m , where δ_m^{disc} is chosen to be the exact discretisation error $\|\hat{y} - P_m^+ P_m \hat{y}\|$ and $n(m, \delta_m^{disc})$ is calculated with Algorithm 2.

- BAKUSHINSKII, A. (1984) Remarks on the choice of regularization parameter from quasioptimality and relation tests. *Zh. Vychisl. Mat. i Mat. Fiz.*, **24**, 1258–1259.
- BAUER, F., PEREVERZEV, S. & ROSASCO, L. (2007) On regularization algorithms in learning theory. *Journal of complexity*, **23**, 52–72.
- BAUER, F. & REISS, M. (2008) Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, **24**, 055009.
- BISSANTZ, N., HOHAGE, T., MUNK, A. & RUYMGAART, F. (2007) Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal on Numerical Analysis*, **45**, 2610–2636.
- BLANCHARD, G., HOFFMANN, M. & REISS, M. (2018a) Early stopping for statistical inverse problems via truncated svd estimation. *Electronic Journal of Statistics*, **12**, 3204–3231.
- BLANCHARD, G., HOFFMANN, M. & REISS, M. (2018b) Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, **6**, 1043–1075.
- BLANCHARD, G. & MATHÉ, P. (2012) Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse problems*, **28**, 115011.
- CAVALIER, L. (2011) Inverse problems in statistics. *Inverse problems and high-dimensional estimation*. Springer, pp. 3–96.
- CAVALIER, L. & TSYBAKOV, A. (2002) Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, **123**, 323–354.
- DAI, W., MA, Y., TONG, T. & ZHU, L. (2015) Difference-based variance estimation in nonparametric regression with repeated measurement data. *Journal of Statistical Planning and Inference*, **163**, 1–20.
- DE VITO, E., ROSASCO, L. & CAPONNETTO, A. (2006) Discretization error analysis for tikhonov regularization. *Analysis and Applications*, **4**, 81–99.
- DETTE, H., MUNK, A. & WAGNER, T. (1998) Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**, 751–764.
- DONOHO, D. L. (1995) Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and computational harmonic analysis*, **2**, 101–126.
- ENGL, H. W., HANKE, M. & NEUBAUER, A. (1996) *Regularization of inverse problems*, vol. 375. Springer Science & Business Media.
- GARDE, H. & HYVÖNEN, N. (2021) Mimicking relative continuum measurements by electrode data in two-dimensional electrical impedance tomography. *Numerische Mathematik*, **147**, 579–609.
- GUT, A. (2013) *Probability: a graduate course*, vol. 75. Springer Science & Business Media.
- HANSEN, P. C. (1994) Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems. *Numerical algorithms*, **6**, 1–35.
- HANSEN, P. C. (2010) *Discrete inverse problems: insight and algorithms*, vol. 7. Siam.
- HARRACH, B., JAHN, T. & POTTHAST, R. (2020) Beyond the Bakushinskii veto: Regularising linear inverse problems without knowing the noise distribution. *Numerische Mathematik*, **145**, 581–603.
- ITO, K. & JIN, B. (2015) *Inverse problems: Tikhonov theory and algorithms*. World Scientific.
- JAHN, T. (2021a) A modified discrepancy principle to attain optimal convergence rates under unknown noise. *Inverse Problems (in press)*, <https://doi.org/10.1088/1361-6420/ac1775>.
- JAHN, T. (2021b) Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise. *arXiv preprint arXiv:2104.06184*.
- JAHN, T. & JIN, B. (2020) On the discrepancy principle for stochastic gradient descent. *Inverse Problems*, **36**, 095009.
- KAIPIO, J. & SOMERSALO, E. (2007) Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of computational and applied mathematics*, **198**, 493–504.
- KINDERMANN, S., PEREVERZEV, S. & PILIPENKO, A. (2018) The quasi-optimality criterion in the linear functional strategy. *Inverse Problems*, **34**, 075001.

- KLENKE, A. (2013) *Probability theory: a comprehensive course*. Springer Science & Business Media.
- MATHÉ, P. & PEREVERZEV, S. V. (2006) Regularization of some linear ill-posed problems with discretized random noisy data. *Mathematics of Computation*, **75**, 1913–1929.
- MATHÉ, P. & PEREVERZEV, S. V. (2001) Optimal discretization of inverse problems in Hilbert scales. regularization and self-regularization of projection methods. *SIAM Journal on Numerical Analysis*, **38**, 1999–2021.
- MATHÉ, P. & PEREVERZEV, S. V. (2003a) Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, **19**, 1263.
- MATHÉ, P. & PEREVERZEV, S. V. (2003b) Geometry of linear ill-posed problems in variable hilbert scales. *Inverse problems*, **19**, 789.
- MATHÉ, P. & PEREVERZEV, S. V. (2017) Complexity of linear ill-posed problems in hilbert space. *Journal of Complexity*, **38**, 50–67.
- MOROZOV, V. A. (1968) The error principle in the solution of operational equations by the regularization method. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, **8**, 295–309.
- O'SULLIVAN, F. (1986) A statistical perspective on ill-posed inverse problems. *Statistical science*, 502–518.
- RICE, J. *et al.* (1984) Bandwidth choice for nonparametric regression. *Annals of statistics*, **12**, 1215–1230.
- RUMP, S. M. (2011) Verified bounds for singular values, in particular for the spectral norm of a matrix and its inverse. *BIT Numerical Mathematics*, **51**, 367–384.
- TIKHONOV, A. N. & GLASKO, V. B. (1965) Use of the regularization method in non-linear problems. *USSR Computational Mathematics and Mathematical Physics*, **5**, 93–107.
- VINOGRADOV, V. (1994) *Refined large deviation limit theorems*, vol. 315. CRC Press.
- VOGEL, C. R. (2002) *Computational methods for inverse problems*. SIAM.
- WAHBA, G. (1977) Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis*, **14**, 651–667.