OPEN ACCESS
IOP Publishing

Inverse Problems

Inverse Problems **37** (2021) 095008 (23pp)

https://doi.org/10.1088/1361-6420/ac1775

A modified discrepancy principle to attain optimal convergence rates under unknown noise

Tim Jahn[*]

Institute of Mathematics, Goethe-University Frankfurt, Germany

E-mail: timjahn2@gmx.de

Received 5 March 2021, revised 9 June 2021
Accepted for publication 23 July 2021
Published 11 August 2021

Abstract

We consider a linear ill-posed equation in the Hilbert space setting. Multiple independent unbiased measurements of the right-hand side are available. A natural approach is to take the average of the measurements as an approximation of the right-hand side and to estimate the data error as the inverse of the square root of the number of measurements. We calculate the optimal convergence rate (as the number of measurements tends to infinity) under classical source conditions and introduce a modified discrepancy principle, which asymptotically attains this rate.

Keywords: statistical inverse problems, discrepancy principle, convergence, optimality, spectral cut-off

(Some figures may appear in colour only in the online journal)

1. Introduction

So we aim to solve $K\hat{x} = \hat{y}$, where K is compact with dense range and $\hat{x}$ and $\hat{y}$ are elements of infinite-dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. The exact data $\hat{y}$ is unknown, but we have access to multiple and unbiased i.i.d. measurements $Y_1, \ldots, Y_n$ with unknown arbitrary distribution and finite variance ($\mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$). Note that at this point the measurements are infinite-dimensional objects (e.g. functions), we will later discretise along the singular vectors of the operator K. Repeating and averaging the measurement process is a standard engineering practice to estimate and reduce random uncertainties, see [5, 18, 23] for introducing monographs on

[*]Author to whom any correspondence should be addressed.

the subject of error analysis from a practical view point. In the given setting, a natural estimator of the unknown data $\hat{y}$ is the sample mean

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

The compactness of K implies that the equation is ill-posed, so that one cannot rely on classical direct methods like LR- or UR-decomposition to determine the (generalised) inverse of K. Regularisation is needed, and the inverse is replaced with a family of related but continuous approximations, e.g. Tikhonov or spectral cut-off regularisation. The particular choice of the approximation has to be based inevitably on knowledge of an upper bound of the true error $\delta_n^{\text{true}} := \|\bar{Y}_n - \hat{y}\|$, as the famous result of Bakushinskii [2] states. While the exact value of δ_n^{true} is clearly not given due to randomness, its variance depends mainly on the number of measurements,

$$\mathbb{E}\left[\delta_n^{\text{true}\,2}\right] = \mathbb{E}\|\bar{Y}_n - \hat{y}\|^2 = \frac{\mathbb{E}\|Y_1 - \hat{y}\|^2}{n}.$$

Thus

$$\delta_n^{\text{est}} := \frac{1}{\sqrt{n}} \quad \text{or} \quad \delta_n^{\text{est}} := \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\|Y_i - \bar{Y}_n\|^2}}{\sqrt{n}}$$

are natural estimators of the unknown true error δ_n^{true}. So a natural approach for the solution of the equation is to use the mean $\bar{Y}_n$ and the estimated data error δ_n^{est} together with a deterministic regularisation method. Indeed, in [11] it was verified, that the approach converges in a suitable sense for a large class of regularisation methods. See also [12, 20], where this approach was extended to settings involving white or Poissonian noise. The rate of convergence of a given regularisation is known to depend on a certain smoothness of the unknown solution $\hat{x}$ relative to the operator K. Classical convergence rates for deterministic noise (i.e. in a setting where one knows an upper bound for the norm of the noise) are deduced by a worst case error analysis. In our setting however, the noise, though random, typically excludes many 'bad' directions for a fixed unknown error distribution. So as it is typical under random noise, see e.g. [3, 8] or [17], the optimal rates obtained here should be substantially better than the ones one would expect from a deterministic worst case error analysis. Indeed, we show that the optimal rates here are better than for the deterministic worst case. The main result of this work then constitutes of a modified discrepancy principle, which yields (almost) the best possible rate for arbitrary unknown error distributions.

Denote by $(\sigma_j, u_j, v_j)_{j\in\mathbb{N}}$ the singular value decomposition of K (i.e. $(u_j)_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{Y}$ (note that K is assumed to have dense range), $(v_j)_{j\in\mathbb{N}}$ an orthonormal basis of $\mathcal{N}(K)^\perp$, $(\sigma_j)_{j\in\mathbb{N}}$ a monotone to 0 converging sequence of positive numbers and it holds that $Kv_j = \sigma_j u_j$). In the following we will restrict to the spectral cut-off regularisation and to mildly ill-posed problems, i.e. we assume that there exists $q > 0$ such that $\sigma_j^2 \asymp j^{-q}$. Thus the reconstruction will be based on the projections (Y_i, u_j) for $i, j \in \mathbb{N}$. The unbiasedness assumption reads $\mathbb{E}[(Y_i, u_j)] = (\hat{y}, u_j)$ for all $i, j \in \mathbb{N}$ and we moreover assume that there are $C_p > 0, p > 1$ with $\mathbb{E}[\left(Y_i - \hat{y}, u_j\right)^2] \leqslant C_p j^{-p}$. Later we will always consider only a finite number of components for a fixed number of measurements n. Spectral cut-off at truncation level k for the component wise averages then yields the following estimator for $\hat{x}$

$$\bar{X}_k^n := \sum_{j=1}^{k} \frac{1}{\sigma_j} \left(\bar{Y}_n, u_j\right) v_j. \tag{1.1}$$

In order to find a reasonable reconstruction the truncation level has to be determined dependent on the (estimated) noise level, which depends on the number of measurements n.

The rest of the paper is organised as follows. In the following section 2 we state the three main theorems, which are proven in section 3. The main result is accompanied by numerical experiments in section 4 and the article ends with a short conclusion in section 5.

2. Main results

We derive convergence rates with respect to classical Hölder-type source conditions

$$\hat{x} \in \mathcal{X}_{\nu,\rho} := \{(K^*K)^{\nu/2}\xi \mid \xi \in \mathcal{X}, \|\xi\| \leqslant \rho\} = \left\{\sum_{j=1}^{\infty} \sigma_j^{\nu}(\xi, v_j)v_j \mid \xi \in \mathcal{X}, \|\xi\| \leqslant \rho\right\}. \tag{2.1}$$

If $\hat{x} \in \mathcal{X}_{\nu,\rho}$, we say that $\hat{x}$ obeys smoothness (ν, ρ) relative to K. Via (1.1) a whole class of estimators indexed by $k \in \mathbb{N}$ is defined, which is also known under the term projection estimators (with respect to the singular value decomposition, see [9]). The first result gives the optimal error bound for our estimators (1.1) on $\mathcal{X}_{\nu,\rho}$, where we measure performance by the integrated mean squared error (also called the minimax (L^2)-risk in this context).

Theorem 2.1. *Let $\hat{y} := K\hat{x}$. Assume that $Y_1, Y_2, \ldots$ are i.i.d. for $i = 1, 2, \ldots$ with $\mathbb{E}[Y_1] = \hat{y}$. Moreover, assume that there are $q > 0, p > 1$ with $\sigma_j^2 \asymp j^{-q}$ and $\mathbb{E}\left(Y_1 - \hat{y}, u_j\right)^2 \asymp j^{-p}$. Then there holds*

$$\inf_{k\geqslant 1} \sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}} \mathbb{E}\|\bar{X}_k^n - \hat{x}\|^2 \asymp \begin{cases} \dfrac{1}{n} & q - p < -1 \\ \dfrac{\log(n\rho)}{n} & q - p = -1 \\ \rho^{\frac{q+1-p}{(\nu+1)q+1-p}} \left(\dfrac{1}{n}\right)^{\frac{\nu}{\nu+1-\frac{p-1}{q}}} & q - p > -1 \end{cases}.$$

In particular, for the a priori *choice*

$$k_n \asymp \begin{cases} (\rho n)^{\frac{1}{\nu q}} & q - p \leqslant -1 \\ (\rho n)^{\frac{1}{(1+\nu)q+1-p}} & q - p > -1 \end{cases}$$

it holds that

$$\sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}} \mathbb{E}\|\bar{X}_{k_n}^n - \hat{x}\|^2 \asymp \inf_{k\geqslant 1} \sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}} \mathbb{E}\|\bar{X}_k^n - \hat{x}\|^2.$$

Note that under additional assumption, one can show that the rate from theorem 2.1 is (up to a constant factor) the optimal rate for all possible estimators, not just for projection estimators (1.1). See e.g. [4, 22] for the case where $\left((Y_1 - \hat{y}, u_j)\right)_{j\in\mathbb{N}}$ are independent and Gaussian.

In view of the fact that the optimal worst case error bound for deterministic noise level $1/\sqrt{n}$ under the source condition $\hat{x} \in \mathcal{X}_{\nu,\rho}$ has order $(1/\sqrt{n})^{\frac{\nu}{\nu+1}}$, we see that the minimax risk attained by the oracle k_n is in all cases strictly better. In particular, for $q - p < -1$, the problem is in fact well-posed. However, the above choice k_n requires knowledge of both the smoothness ν and the decay of variances p. A plain use of the discrepancy principle [21] as an adaptive strategy to determine the truncation level would be to find $k = k(\bar{Y}_n, \delta_n^{\mathrm{est}})$, such that the size of the residual is approximately equal to the estimated noise level, i.e. by the relation

$$\sqrt{\sum_{j=k+1}^{\infty} (\bar{Y}_n, u_j)^2} \approx \delta_n^{\mathrm{est}}. \tag{2.2}$$

In [11] it was shown, that the choice (2.2) adapts to the unknown smoothness ν in the sense that asymptotically the optimal deterministic bound holds with a probability converging to 1. According to theorem 2.1, this is suboptimal. The reason is an intrinsic drawback of the plain discrepancy principle for statistical noise, which tempts to stop too late. We therefore consider in this work a modified version of the discrepancy principle, which also takes information about the stochastic nature of the noise into account.

We first formulate a simplified version of the main result to illustrate the approach. As already mentioned, the rate of convergence depends on certain smoothness properties of the real unknown solution $\hat{x}$ relative to the forward operator K, see e.g. (2.1). The general idea is to rescale the operator K with a weighting operator S such that the smoothness of $\hat{x}$ relative to the rescaled operator SK is better than the original one relative to K. In order to avoid distinction of several cases, let us assume for a moment that $q - p > -1$ additional to the assumptions of theorem 2.1. Moreover, we assume that p is known to us. The latter is a serious restriction, which will be dropped in the main result theorem 2.3 below. However, there are settings, where this knowledge is justified, see example 2.1 at the end of this section. For any $\varepsilon > 0$ with $p > 1 + \varepsilon$ we define the (linear and unbounded) weighting operator S as the linear extension of

$$\begin{aligned} S &: \mathcal{D}(S) \subset \mathcal{Y} \to \mathcal{Y} \\ u_j &\mapsto d_j u_j := j^{\frac{p-1-\varepsilon}{2}} u_j, \quad j \in \mathbb{N} \end{aligned}$$

on

$$\mathcal{D}(S) := \left\{ \sum_{j=1}^{\infty} \alpha_j u_j \ : \ \sum_{j=1}^{\infty} \alpha_j^2 d_j^2 < \infty \right\}.$$

Since $q - p > -1$, we directly see that $SK : \mathcal{X} \to \mathcal{Y}$ is compact, with singular values $d_j\sigma_j \asymp j^{-\frac{q+1+\varepsilon-p}{2}}$ and the same singular bases $(v_j)_{j\in\mathbb{N}}$ and $(u_j)_{j\in\mathbb{N}}$ as K. Now assume that $\hat{x}$ obeys smoothness (ν, ρ) relative to K, i.e. there exists $\xi \in \mathcal{Y}$ with $\hat{x} = (K^*K)^{\frac{\nu}{2}}\xi$ and $\|\xi\| \leqslant \rho$. Let $\nu' := \frac{q}{q+1+\varepsilon-p}\nu$. Note that $\nu' > \nu$, since we assumed that $p > 1 + \varepsilon$. Then

$$\hat{x} = (K^*K)^{\frac{\nu}{2}}\xi = \sum_{j=1}^{\infty} \sigma_j^{\nu}(\xi, v_j)v_j = \sum_{j=1}^{\infty} (d_j\sigma_j)^{\nu'} \frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}}(\xi, v_j)v_j = \left((SK)^*SK\right)^{\frac{\nu'}{2}}\xi',$$

with $\xi' = \sum_{j=1}^{\infty} \frac{\sigma_j^{\nu-\nu'}}{d_j}(\xi, v_j)v_j$ and

$$\|\xi'\|^2 = \sum_{j=1}^{\infty}\left(\frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}}\right)^2 (\xi, v_j)^2 \asymp \sum_{j=1}^{\infty} \frac{j^{-q(\nu-\nu')}}{j^{(p-1-\varepsilon)\nu'}}(\xi, v_j)^2$$

$$= \sum_{j=1}^{\infty} j^{-q\nu\left(1-\frac{q}{q+1+\varepsilon-p}\right)-\frac{p-1-\varepsilon}{q+1+\varepsilon-p}q\nu}(\xi, v_j)^2 = \sum_{j=1}^{\infty}(\xi, v_j)^2 \leqslant \rho^2,$$

therefore $\hat{x}$ obeys smoothness $(\nu', c\rho)$ relative to SK (with a constant $c > 0$). Moreover, the rescaled measurements $SY_1, SY_2, \ldots$ are unbiased estimators of $S\hat{y}$ with finite variance

$$\mathbb{E}\|SY_1 - S\hat{y}\|^2 = \sum_{j=1}^{\infty} d_j^2 \mathbb{E}(Y_1 - \hat{y}, u_j)^2 \asymp \sum_{j=1}^{\infty} j^{p-1-\varepsilon} j^{-p} = \sum_{j=1}^{\infty} j^{-(1+\varepsilon)} < \infty. \tag{2.3}$$

Our modification of the discrepancy principle is that we apply it not to the unscaled operator and measurements K and $\bar{Y}_n$, but to the rescaled ones SK and $S\bar{Y}_n$ (note that SK also has dense range). Consequently, the stopping index k_n is the solution of the equation

$$\sqrt{\sum_{j=k+1}^{\infty}(S\bar{Y}_n, u_j)^2} = \sqrt{\sum_{j=k+1}^{\infty} d_j^2(\bar{Y}_n, u_j)^2} \approx \delta_n^{\mathrm{est}\prime}, \tag{2.4}$$

with

$$\delta_n^{\mathrm{est}\prime} := \frac{1}{\sqrt{n}} \quad \text{or} \quad \delta_n^{\mathrm{est}\prime} := \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\|SY_i - S\bar{Y}_n\|^2}}{\sqrt{n}}.$$

The following theorem states that up to ε the optimal bound from theorem 2.1 holds with a probability converging to 1 as $n \to \infty$ using this strategy. Note that convergence in mean squared error cannot be expected for the discrepancy principle, see [11]. However, adding to the procedure a so-called 'emergency stop' (see [8, 11]) might allow to deduce rates in mean squared error, but we will leave this as a future work and focus on the rates in probability.

Theorem 2.2. *Let $\hat{y} = K\hat{x}$ and assume that $Y_1, Y_2, \ldots$ are i.i.d. for $i = 1, 2, \ldots$ with $\mathbb{E}[Y_1] = \hat{y}$. Moreover, assume that there are $q > 0, p > 1$ with $\sigma_j^2 \asymp j^{-q}$ and $\mathbb{E}\big(Y_1 - \hat{y}, u_j\big)^2 \asymp j^{-p}$ and $q > p - 1$. Let $\varepsilon > 0$ such that $p > 1 + \varepsilon$ and assume that $\hat{x} \in \mathcal{X}_{\nu,\rho}$. Then there exists $L > 0$, such that for k_n the solution of* (2.4) *there holds*

$$\mathbb{P}\left(\|\bar{X}_{k_n}^n - \hat{x}\| \leqslant L\rho^{\frac{1}{1+\nu'}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\right) \to 1 \tag{2.5}$$

as $n \to \infty$.

Theorem 2.2 is an immediate consequence of theorem 1.2.4 from [14] (which is a refined version of theorem 4 of [11]) applied to SK and $S\bar{Y}_n$. A quick calculation reveals, that $\frac{\nu'}{\nu'+1} = \frac{\nu}{\nu+1-\frac{p-1-\varepsilon}{q}}$, thus up to $\varepsilon > 0$ we get the optimal rate from theorem 2.1. Note however, that the smaller we choose ε, the slower will be the convergence to 1 in (2.5).

Now we generalise the above result in several ways. We relax the condition $\mathbb{E}(Y_1 - \hat{y}, u_j)^2 \asymp j^{-p}$ to $\mathbb{E}(Y_1 - \hat{y}, u_j)^2 \leqslant C_p j^{-p}$ (for some $C_p > 0$). Most importantly, the exponent p is no longer

Algorithm 1. Modified discrepancy principle with estimated data error.

Given measurements (Y_i, u_j) with $i = 1, \ldots, n$ and $j = 1, \ldots, \lfloor n^{1-\varepsilon_1} \rfloor$;
Estimate variances
Set $s_{j,n}^2 := \frac{1}{n-1}\sum_{i=1}^n \left(Y_i - \bar{Y}_n, u_j\right)^2$;
Calculate weights
Set $d_{1,n} := \sqrt{\min\left(\frac{\sum_{j'=1}^{\lfloor n^{1-\varepsilon_1}\rfloor} s_{j',n}^2}{s_{1,n}^2}, \frac{1}{\sigma_1^2}\right)}$;
for $j = 2, \ldots, \lfloor n \rfloor^{1-\varepsilon_1}$ **do**
    Set $d_{j,n} := \sqrt{\min\left(\frac{j^{-(1+\varepsilon_2)}}{s_{j,n}^2}\sum_{j'=1}^{\lfloor n^{1-\varepsilon_1}\rfloor} s_{j',n}^2, \frac{\sigma_{j-1}^2}{\sigma_j^2} d_{j-1,n}^2\right)}$;
end for
Apply discrepancy principle to rescaled measurements
Set $\delta_n^{\mathrm{est}\prime} := \sqrt{\frac{\sum_{j=1}^{\lfloor n^{1-\varepsilon_1}\rfloor} d_{j,n}^2 s_{j,n}^2}{n}}$;
$k = 0$
while $\sqrt{\sum_{j=k+1}^{\lfloor n^{1-\varepsilon_1}\rfloor} d_{j,n}^2\left(\bar{Y}_n, u_j\right)^2} > \delta_n^{\mathrm{est}\prime}$ **do**
    $k = k + 1$;
end while
$k_n = k$;

assumed to be known. Moreover, we account for the fact that in practice we can only measure a finite number of components. The component-wise variances $\mathbb{E}[(Y_1 - \hat{y}, u_j)^2]$ will be estimated from the multiple measurements and are then used to determine the (now random) rescaling weights $d_{j,n}$. Consequently, the weighting operator S_n is now also random and depends on the samples $Y_1, \ldots, Y_n$. The precise implementation of the modified discrepancy principle (with ad hoc unknown decay of the component-wise variances) is given in algorithm 1. Hereby, the factor $\sum_{j=1}^{m_n} s_{n,j}^2$ is in essence a normalisation by $\mathbb{E}\|Y_1 - \hat{y}\|^2$. The two arguments of the $\min(\cdot)$ function can be roughly interpreted as follows: the first one assures that the rescaled measurements still have finite variance (cf (2.3)), while the second one assures that the rescaled operator is still bounded. We state now the main result, which confirms that the optimal bound from theorem 2.1 holds with a probability converging to 1 as $n \to \infty$ (up to discretisation and $\varepsilon_2 > 0$ arbitrary small in the exponent) using this strategy.

Theorem 2.3. *Let $K\hat{x} = \hat{y}$ and $0 < \varepsilon_1, \varepsilon_2 < 1$. Assume that $Y_1, Y_2, \ldots$ are i.i.d. with $\mathbb{E}[Y_1] = \hat{y}$. Moreover, there are q, p with $q > p - 1 > 0$ and $C_p, C > 0$ with $\sigma_j^2 \asymp j^{-q}$ and $\mathbb{E}[\left(Y_1 - \hat{y}, u_j\right)^2] \leqslant C_p j^{-p}$ and $\mathbb{E}[\left(Y_1 - \hat{y}, u_j\right)^4]/\left(\mathbb{E}\left[\left(Y_1 - \hat{y}, u_j\right)^2\right]\right)^2 \leqslant C$ for all $j \in \mathbb{N}$. Assume that $\hat{x} \in \mathcal{X}_{\nu,\rho}$ for $\nu, \rho > 0$ and let k_n be the stopping index of the modified discrepancy principle as implemented in algorithm 1 with $0 < \varepsilon_1 < 1$ and $0 < \varepsilon_2 < p - 1$ and $Y_1, \ldots, Y_n$. Then there exists an $L > 0$ such that there holds*

$$\lim_{n\to\infty} \mathbb{P}\left(\|\bar{X}_{k_n}^n - \hat{x}\| \leqslant L \max\left(\rho^{\frac{q+1+\varepsilon_2-p}{(\nu+1)q+1+\varepsilon_2-p}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu}{\nu+1-\frac{p-1-\varepsilon_2}{q}}}, \rho\left(\frac{1}{\sqrt{n}}\right)^{-(1-\varepsilon_1)q\nu}\right)\right) = 1. \tag{2.6}$$

The proof of theorem 2.3 is substantially more difficult than the one of theorem 2.2, mostly due to the dependence of the rescaling operator S_n on the (realisations of) the measurements

$Y_1, \ldots, Y_n$. In particular, the $S_nY_1, \ldots, S_nY_n$ are not independent and hence we cannot simply apply the results from [11].

Remark 2.1. Algorithm 1 could be applied in a general setting, e.g. also to severely ill-posed problems. The weights $d_{j,n}$ are defined such that

$$d_{1,n}\sigma_1 \geqslant d_{2,n}\sigma_2 \geqslant \ldots .$$

Note that a chosen ε_2 fulfills the condition $\varepsilon_2 < p-1$, if $\lim_{j\to\infty}\lim_{n\to\infty} d_{j,n} = \infty$ and the latter can be checked to verify, that ε_2 was chosen sufficiently small. The assumption $q > p-1$ is only made for convenience and is not restrictive, since if $\mathbb{E}(Y_1 - \hat{y}, u_j)^2 \leqslant C_p j^{-p}$ there also holds that $\mathbb{E}(Y_1 - \hat{y}, u_j)^2 \leqslant C_p j^{-p'}$ for all $p' \leqslant p$.

The second argument of the maximum in (2.6) is a discretisation error due to the usage of only finitely many singular vectors. If the latter is negligible, i.e. if $\frac{\nu}{\nu+1} < (1-\varepsilon)q\nu$, the rate from theorem 2.3 is better than the deterministic worst-case rate from [11]. The additional assumption $\sup_{j\in\mathbb{N}} \frac{\mathbb{E}[(Y_1-\hat{y},u_j)^4]}{(\mathbb{E}[(Y_1-\hat{y},u_j)^2])^2} < \infty$ assures that the component distribution are not too degenerated. This is clearly fulfilled, if $\mathbb{E}(Y_1 - \hat{y}, u_j) \overset{d}{=} c_j Z$ for some Z with $\mathbb{E}[Z] = 0, E[Z^4] < \infty$ and $(c_j)_{j\in\mathbb{N}} \subset \mathbb{R}\backslash\{0\}$ (e.g. this holds under Gaussian noise). In particular no independence between the components is required. Our assumption of finite variance ($\mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$) excludes a direct application to white noise scenarios. The following example shows, how to adapt the approach for Hilbert–Schmidt operators under white noise.

Example 2.1. Consider the equation $A\hat{x} = \hat{z}$ for $A : \mathcal{X} \to \mathcal{Y}$ Hilbert–Schmidt and assume there is a $q > 1$ such that $\sigma_j(A)^2 \asymp j^{-q}$. Assume that the measurements are corrupted by i.i.d centered Hilbert-space processes $Z_1, Z_2, \ldots$ (operating on $\mathcal{Y}$). I.e. the $Z_i : \mathcal{Y} \to L^2(\Omega, \mathcal{A}, \mathbb{P})$ are bounded linear operators from $\mathcal{Y}$ to the space of square-integrable real-valued random variables (on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$), such that $\mathbb{E}(Z_1, z) = 0$. Moreover, Z_i has an arbitrary covariance operator $\mathbf{Cov}_Z : \mathcal{Y} \to \mathcal{Y}$, which is the bounded linear operator defined implicitly via the equation $(\mathbf{Cov}_Z\, z, z') = \mathbb{E}\left[(Z_1, z)(Z_1, z')\right]$ for all $z, z' \in \mathcal{Y}$ (the case where $\mathbf{Cov}_Z = \mathrm{Id}$ is denoted as white noise). Instead of $Ax = z$ we solve the symmetrised equation $K\hat{x} = \hat{y}$, with $K = A^*A$ and $y = A^*\hat{z}$. The symmetrised i.i.d. measurements $Y_1 = A^*(\hat{z} + Z_1), Y_2 = A^*(\hat{z} + Z_2), \ldots$ then fulfill $\mathbb{E}[Y_1] = A^*\hat{z} = \hat{y}$ and

$$\begin{aligned}\mathbb{E}(Y_1 - \hat{y}, u_j(K))^2 &= \mathbb{E}(A^*(\hat{z} + Z_1) - A^*\hat{z}, v_j(A))^2 = \sigma_j(A)^2\mathbb{E}(Z_1, u_j(A))^2 \\ &\asymp j^{-q}(\mathbf{Cov}_Z\, u_j(A), u_j(A))^2 \leqslant j^{-q}\|\mathbf{Cov}_Z\|^2 = j^{-q}.\end{aligned}$$

Since we assume to know the singular value decomposition, this allows to apply theorem 2.2. It should be noted, that for coloured noise, no prewhitening step or additional assumptions for the covariance operator are needed (in contrast to e.g. [8]).

All in all, the main contribution of this work is to answer the question of optimal adaptivity (in the minimax-sense) for the discrepancy principle in statistical inverse problems with multiple measurements, which was left open in the original work [11]. Moreover, in the light of example 2.1 the results may be compared to classical existing results for statistical inverse problems, usually using a white noise error model. In particular, they generalise results from [8, 15, 17], where modifications of the discrepancy principle are applied to the symmetrised equation in several ways. Firstly, the error distribution is arbitrary, secondly the noise level and the covariance structure need not to be known and thirdly, a self-similarity condition (assumption 3 in [17] and assumption 2.4 in [15]) for $\hat{x}$ is not needed. However it should be mentioned here that using symmetrisation is usually avoided, since the ill-posedness of the symmetrised equation $A^*Ax = A^*y$ is much worse than the one of the original equation $Ax = y$. Note that this does not contradict the (almost) order-optimality of the methods relying on the discrepancy principle mentioned above, but still may cause problems in practice. Because of this, under white noise one often relies on other methods, which do not depend on the residual, see e.g. [6] for *a priori* bounds [19], for the Lepski principle, or [16] for unbiased risk estimation, to only name a few. Finally in [7, 12, 13] recent modifications of the discrepancy principle, which are based on discretisation and not on symmetrisation, are investigated in white noise scenarios.

3. Proofs

In this section we present the proofs of the above statements.

3.1. Proof of theorem 2.1

Note that $(\hat{y}, u_j) = \sigma_j(\hat{x}, v_j) = \sigma_j^{1+\nu}(\xi, v_j)$. The bias-variance decomposition gives

$$\begin{aligned}\mathbb{E}\|\bar{X}_k^n - \hat{x}\|^2 &= \mathbb{E}\left[\sum_{j=1}^k \left(\frac{(\bar{Y}_n, u_j)}{\sigma_j} - (\hat{x}, v_j)\right)^2\right] + \sum_{j=k+1}^\infty (\hat{x}, v_j)^2 \\ &= \sum_{j=1}^k \sigma_j^{-2}\mathbb{E}(\bar{Y}_n - \hat{y}, u_j)^2 + \sum_{j=k+1}^\infty (\hat{x}, v_j)^2 \\ &= \frac{1}{n}\sum_{j=1}^k \sigma_j^{-2}\mathbb{E}(Y_1 - \hat{y}, u_j)^2 + \sum_{j=k+1}^\infty \sigma_j^{2\nu}(\xi, v_j)^2 \\ &\asymp \frac{1}{n}\sum_{j=1}^k j^{q-p} + \sum_{j=k+1}^\infty j^{-\nu q}(\xi, v_j)^2.\end{aligned}$$

Therefore it holds that

$$\begin{aligned}\sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}} \mathbb{E}\|\bar{X}_k^n - \hat{x}\|^2 &\asymp \frac{1}{n}\int_{j=1}^k x^{q-p}\,\mathrm{d}x + \rho k^{-\nu q} \\ &\asymp \begin{cases} \frac{1}{n} + \rho k^{-\nu q} & q-p<-1 \\ \frac{1}{n}\ \log(k) + \rho k^{-\nu q} & q-p=-1. \\ \frac{1}{n}k^{p-q+1} + \rho k^{-\nu q} & q-p>-1 \end{cases}\end{aligned}$$

The right-hand side is minimised by the choices

$$k = k_n \asymp \begin{cases} (\rho n)^{\frac{1}{\nu q}} & q-p \leqslant -1 \\ (\rho n)^{\frac{1}{(1+\nu)q+1-p}} & q-p>-1 \end{cases}.$$

Thus we obtain

$$\inf_{k\geqslant 1}\sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}}\mathbb{E}\|\bar{X}_k^n-\hat{x}\|^2\asymp\sup_{\hat{x}\in\mathcal{X}_{\nu,\rho}}\mathbb{E}\|\bar{X}_{k_n}^n-\hat{x}\|^2$$

$$\asymp\begin{cases}\frac{1}{n} & q-p<-1\\ \frac{\log(\rho n)}{n} & q-p=-1.\\ \rho^{\frac{q+1-p}{(\nu+1)q+1-p}}\left(\frac{1}{n}\right)^{\frac{\nu}{\nu+1-\frac{p-1}{q}}} & q-p>-1\end{cases}$$

3.2. Proof of theorem 2.3

Let $m_n := \lfloor n^{1-\varepsilon_1}\rfloor$ and $\hat{x}=(K^*K)^{\frac{\nu}{2}}\xi$ with $\|\xi\|\leqslant\rho$. For $j\in\mathbb{N}$ fixed it holds that $s_{j,n}^2\rightarrow\mathbb{E}(Y_1-\hat{y},u_j)^2$ (in probability, almost surely and in L^2). We denote the deterministic limit (for $n\rightarrow\infty$) of the random weights $d_{j,n}$ by

$$d_1:=\sqrt{\min\left(\frac{\mathbb{E}\|Y_1-\hat{y}\|^2}{\mathbb{E}(Y_1-\hat{y},u_1)^2},\frac{1}{\sigma_1^2}\right)},$$

$$d_j:=\sqrt{\min\left(\frac{j^{-(1+\varepsilon_2)}}{\mathbb{E}\big(Y_1-\hat{y},u_j\big)^2}\mathbb{E}\|Y_1-\hat{y}\|^2,\frac{\sigma_{j-1}^2}{\sigma_j^2}d_{j-1}^2\right)},\quad j>1.$$

The weights $d_{j,n}$ and d_j can be interpreted as belonging to weighting operators S_n and S respectively. Moreover, the assumption on the error distribution of the $(Y_1-\hat{y},u_j)$ imply that S can be seen as a deterministic limit (for $n\rightarrow\infty$) of the S_n in a suitable sense. This will ultimately allow to rephrase the increased smoothness relative to the deterministic rescaled limit operator SK instead of the random rescaled operator S_nK. In the following, S_n and S are not used explicitly, it suffices to stick to the weights $d_{j,n}$ and d_j. We start with the following auxiliary proposition, which summarises some of the properties of the sequence $(d_j)_{j\in\mathbb{N}}$.

Proposition 3.1. *There holds*

$$d_j\leqslant\frac{1}{\sigma_j},\tag{3.1}$$

$$\lim_{j\rightarrow\infty}d_j=\infty,\tag{3.2}$$

$$\inf_{j\in\mathbb{N}}d_j=:d>0.\tag{3.3}$$

Proof of proposition 3.1. Note that obviously $d_j>0$ for all $j\in\mathbb{N}$. First, (3.1) is fulfilled for $j=1$. For $j\geqslant 2$, we have

$$d_j\leqslant\frac{\sigma_{j-1}}{\sigma_j}d_{j-1}\leqslant\frac{\sigma_{j-1}}{\sigma_j}\frac{\sigma_{j-2}}{\sigma_{j-1}}d_{j-2}\leqslant\frac{\sigma_{j-1}}{\sigma_j}\frac{\sigma_{j-2}}{\sigma_{j-1}}\ldots\frac{\sigma_1}{\sigma_2}d_1\leqslant\frac{\sigma_{j-1}}{\sigma_j}\frac{\sigma_{j-2}}{\sigma_{j-1}}\ldots\frac{\sigma_1}{\sigma_2}\frac{1}{\sigma_1}=\frac{1}{\sigma_j}.\tag{3.4}$$

For (3.2) set $J := \sup\left\{ j \in \mathbb{N}\ :\ d_j = \sqrt{\frac{j^{-(1+\varepsilon_2)}}{\mathbb{E}(Y_1-\hat{y},u_j)^2}\mathbb{E}\|Y_1-\hat{y}\|^2}\right\}$. If $J=\infty$, the statement is proven since from $p>1+\varepsilon_2$ it follows that

$$\frac{j^{-(1+\varepsilon_2)}}{\mathbb{E}(Y_1-\hat{y},u_j)^2} \geqslant j^{p-(1+\varepsilon_2)}C_p^{-1} \to \infty$$

as $j\to\infty$. Otherwise, if $J<\infty$, there holds $d_j = \frac{\sigma_{j-1}}{\sigma_j}d_{j-1}$ for $j\geqslant J$ and thus

$$d_j = \frac{\sigma_{j-1}}{\sigma_j}d_{j-1} = \frac{\sigma_{j-1}}{\sigma_j}\frac{\sigma_{j-2}}{\sigma_{j-1}}\dots\frac{\sigma_J}{\sigma_{J+1}}d_J = \frac{\sigma_J}{\sigma_j}d_J \to \infty$$

as $j\to\infty$, since $\sigma_j\to 0$. Finally, (3.3) follows directly from (3.2). □

Now we first show that the true solution has at least smoothness $\nu' := \frac{q}{q+1+\varepsilon_2-p}\nu$ (relative to the rescaled limit operator SK). Since $\varepsilon_2 < p-1$ there holds $\nu'>\nu$. We use (the reverse of) (3.1) together with (3.3) and obtain

$$\frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}} \leqslant \frac{d_j^{\nu'-\nu}}{d_j^{\nu'}} = d_j^{-\nu} \leqslant d^{-\nu} \tag{3.5}$$

for all $j\in\mathbb{N}$. We express $\hat{x}$ with respect to the rescaled limit operator SK and obtain

$$\hat{x} = \sum_{j=1}^{\infty}\sigma_j^{\nu}(\xi,v_j)v_j = \sum_{j=1}^{\infty}\left(d_j\sigma_j\right)^{\nu'}\frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}}(\xi,v_j)v_j = \sum_{j=1}^{\infty}(d_j\sigma_j)^{\nu'}(\xi',v_j)v_j \tag{3.6}$$

with $\xi' := \sum_{j=1}^{\infty}\frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}}(\xi,v_j)v_j$. By (3.5), there holds $\|\xi'\|\leqslant d^{-\nu}\|\xi\|\leqslant d^{-\nu}\rho =: \rho'$.

The assumption $\sup_{j\in\mathbb{N}}\frac{\mathbb{E}\left[(Y_1-\hat{y},u_j)^4\right]}{\left(\mathbb{E}\left[(Y_1-\hat{y},u_j)^2\right]\right)^2}$ guarantees that we can estimate the variances $\mathbb{E}(Y_1-\hat{y},u_j)^2$ uniformly for $j=1,\dots,m_n$. To see this we use theorem 2 of [1] which states that $\mathbb{E}[|s_{j,n}^2-\mathbb{E}(Y_1-\hat{y},u_j)^2|^2]\leqslant\frac{4}{n}\mathbb{E}(Y_1-\hat{y},u_j)^4$. Let $\eta>0$. Then

$$\begin{aligned}
&\mathbb{P}\left(|s_{j,n}^2-\mathbb{E}(Y_1-\hat{y},u_j)^2|\leqslant\eta\mathbb{E}(Y_1-\hat{y},u_j)^2,\ \forall\ j\leqslant m_n\right)\\
&\quad\geqslant 1-\sum_{j=1}^{m_n}\mathbb{P}\left(|s_{j,n}^2-\mathbb{E}(Y_1-\hat{y},u_j)^2|>\eta\mathbb{E}(Y_1-\hat{y},u_j)^2\right)\\
&\quad\geqslant 1-\sum_{j=1}^{m_n}\frac{\mathbb{E}[|s_{j,n}^2-\mathbb{E}(Y_1-\hat{y},u_j)^2|^2]}{(\eta\mathbb{E}[(Y_1-\hat{y},u_j)^2])^2}\\
&\quad= 1-\frac{4m_n}{\eta^2 n}\sup_{j=1,\dots,m_n}\frac{\mathbb{E}[(Y_1-\hat{y},u_j)^4]}{(\mathbb{E}[(Y_1-\hat{y},u_j)^2])^2}\geqslant 1-\frac{4m_nC}{\eta^2 n}\\
&\quad\geqslant 1-\frac{4}{\eta^2}Cn^{-\varepsilon_1}\to 1
\end{aligned}$$

as $n \to \infty$, where we used Chebyshev's inequality in the second step. From that directly follows

$$\mathbb{P}\left(\frac{d_j}{2} \leqslant d_{j,n} \leqslant 2d_j,\ \forall\ j = 1, \ldots, m_n\right) \to 1, \tag{3.7}$$

$$\mathbb{P}\left(|\sqrt{n}\delta_n^{\text{est}} - \gamma| \leqslant \frac{\gamma}{2}\right) \to 1 \tag{3.8}$$

for

$$\delta_n^{\text{est}\prime} := \sqrt{\frac{\sum_{j=1}^{m_n} d_{j,n}^2 s_{j,n}^2}{n}} \tag{3.9}$$

from algorithm 1 and $\gamma := \sqrt{\sum_{j=1}^{\infty} d_j^2 \mathbb{E}(Y_1 - \hat{y}, u_j)^2}$ as $n \to \infty$.

We will distinguish two cases in the following. In the analysis we will often restrict to certain good events Ω_n which hold with a probability $\mathbb{P}(\Omega_n) \to 1$ as $n \to \infty$. Moreover, we will repeatedly use Markov/Chebyshev's inequality and that for i.i.d. real-valued random variables $Z_1, \ldots, Z_n$ with $\mathbb{E}[Z_i] = 0$ and $\mathbb{E}[Z_i^2] < \infty$ there holds $\mathbb{E}[\left(\sum_{i=1}^n Z_i\right)^2] = n\mathbb{E}[Z_1^2]$. Thus, e.g.

$$\mathbb{E}\left[(\bar{Y}_n - \hat{y}, u_j)^2\right] = \frac{1}{n}\mathbb{E}[(Y_1 - \hat{y}, u_j)^2]$$

for all $j, n \in \mathbb{N}$.

3.2.1. Case 1. We first assume that for all $k \in \mathbb{N}$ there exists $j_k \geqslant k$ such that $(\hat{y}, u_{j_k}) \neq 0$. Note that then also $(\hat{x}, v_{j_k}), (\xi, v_{j_k}) \neq 0$.

Lemma 3.1. *Assume that for all $k \in \mathbb{N}$ there exists $j_k \geqslant k$ such that $(\hat{y}, u_{j_k}) \neq 0$. Then for $\delta_n^{est\prime}$ from (3.9) there holds*

$$\mathbb{P}\left(\sqrt{\sum_{j=k_n}^{m_n} d_{j,n}^2 (\bar{Y}_n - \hat{y}, u_j)^2} \leqslant \frac{\delta_n^{est\prime}}{2}\right) \to 1$$

as $n \to \infty$.

Proof of lemma 3.1. We first show that there exists $(q_n)_{n\in\mathbb{N}}$ such that

$$\mathbb{P}(k_n \geqslant q_n) \to 1 \quad \text{and} \quad q_n \to \infty \tag{3.10}$$

as $n \to \infty$. For that it suffices to show that $\lim_{n\to\infty} \mathbb{P}(k_n \geqslant k) = 1$ for all $k \in \mathbb{N}$. By assumption there exists $j_k \geqslant k$ such that $(\hat{y}, u_{j_k}) \neq 0$. We set

$$\Omega_n := \left\{|\sqrt{n}\delta_n^{\text{est}\prime} - \gamma| \leqslant \frac{\gamma}{2},\ (\bar{Y}_n, u_{j_k})^2 \geqslant (\hat{y}, u_{j_k})^2/2,\ \frac{d_{j_k,n}}{2} \leqslant d_{j_k} \leqslant 2d_{j_k,n}\right\}. \tag{3.11}$$

Then for $n \geqslant \max(j_k, 32\gamma^2/(d_{j_k}(\hat{y}, u_{j_k}))^2)$,

$$\delta_n^{\mathrm{est}\prime}\chi_{\Omega_n} \leqslant \frac{2\gamma}{\sqrt{n}}\chi_{\Omega_n} < \sqrt{\frac{d_{j_k}^2(\hat{y}, u_{j_k})^2}{8}}\chi_{\Omega_n} \leqslant \sqrt{\frac{d_{j_k}^2(\bar{Y}_n, u_{j_k})^2}{2}}\chi_{\Omega_n}$$
$$\leqslant \sqrt{d_{j_k,n}^2(\bar{Y}_n, u_{j_k})^2} \leqslant \sqrt{\sum_{j=k+1}^{m_n} d_{j,n}^2(\bar{Y}_n, u_j)^2}.$$

Thus $k_n\chi_{\Omega_n} \geqslant k\chi_{\Omega_n}$ by algorithm 1 and (3.10) follows with $\lim_{n\to\infty}\mathbb{P}(\Omega_n) = 1$, which holds because of ((3.7) and (3.8)) and the law of large numbers. We come to the main proof. Let $\varepsilon > 0$ and q_n be such that $q_n \to \infty$ and

$$\mathbb{P}(k_n \geqslant q_n) \to 1 \tag{3.12}$$

as $n \to \infty$. Then,

$$\mathbb{P}\left(\sqrt{\sum_{j=k_n}^{m_n} d_{j,n}^2(\bar{Y}_n - \hat{y}, u_j)^2} \leqslant \frac{\delta_n^{\mathrm{est}\prime}}{2}\right)$$
$$\geqslant \mathbb{P}\left(\sqrt{\sum_{k=q_n}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} \leqslant \frac{\gamma}{8\sqrt{n}},\ \delta_n^{\mathrm{est}\prime} \geqslant \frac{\gamma}{2\sqrt{n}},\ 2d_j\right.$$
$$\left.\geqslant d_{j,n}\ \forall\ j \leqslant m_n,\ k_n \geqslant q_n\right)$$
$$\geqslant 1 - \mathbb{P}\left(\sqrt{\sum_{k=q_n}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} > \frac{\gamma}{8\sqrt{n}}\right) - \mathbb{P}\left(\delta_n^{\mathrm{est}\prime} < \frac{\gamma}{2\sqrt{n}}\right)$$
$$- \mathbb{P}\left(\exists\ j \leqslant m_n \text{ such that } d_{j,n} > 2d_j\right) - \mathbb{P}(k_n < q_n). \tag{3.13}$$

Now

$$\mathbb{P}\left(\sqrt{\sum_{j=q_n}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} > \frac{\gamma}{8\sqrt{n}}\right)$$
$$\leqslant \frac{64n}{\gamma^2}\sum_{j=q_n}^{m_n} d_j^2\mathbb{E}(\bar{Y}_n - \hat{y}, u_j)^2 = \frac{64}{\gamma^2}\sum_{j=q_n}^{m_n} d_j^2\mathbb{E}(Y_1 - \hat{y}, u_j)^2$$
$$= \frac{64}{\gamma^2}\sum_{j=q_n}^{m_n} \min\left(\frac{j^{-(1+\varepsilon_2)}}{\mathbb{E}(Y_1 - \hat{y}, u_j)^2}, \sigma_j^{-2}\right)\mathbb{E}(Y_1 - \hat{y}, u_j)^2$$
$$= \frac{64}{\gamma^2}\sum_{j=q_n}^{m_n} j^{-(1+\varepsilon_2)} \leqslant \frac{64}{\gamma^2}\sum_{j=q_n}^{m_n} j^{-(1+\varepsilon_2)} \to 0 \tag{3.14}$$

as $n \to \infty$, where we used $\sum_{j=1}^{\infty} j^{-(1+\varepsilon_2)} < \infty$ and $q_n \to \infty$ in the sixth step. Plugging (3.7), (3.8), (3.12) and (3.14) into (3.13) then yields

$$\mathbb{P}\left(\sqrt{\sum_{k=q_n}^{m_n} d_{j,n}^2(\bar{Y}_n - \hat{y}, u_j)^2} \leqslant \frac{\delta_n^{\mathrm{est}\prime}}{2}\right) \to 1$$

as $n \to \infty$ and the proof of lemma 3.1 is concluded. □

We start the main proof and decompose as usual into a data propagation error, approximation error and discretisation error

$$\|\bar{X}_{k_n}^n - \hat{x}\| \leqslant \sqrt{\sum_{j=1}^{k_n} \frac{(\bar{Y}_n - \hat{y}, u_j)^2}{\sigma_j^2}} + \sqrt{\sum_{j=k_n+1}^{m_n} (\hat{x}, v_j)^2} + \sqrt{\sum_{j=m_n+1}^{\infty} (\hat{x}, v_j)^2}.$$

We first consider the second term (approximation error). With the convention $\sum_{j=s}^{t} = 0$ for $s > t$, a standard application of Hölder's inequality for $p = \frac{\nu'+1}{\nu'}$ and $q = \nu' + 1$, (3.6) and the triangle inequality yield

$$\begin{aligned}
\sqrt{\sum_{j=k_n+1}^{m_n} (\hat{x}, u_j)^2} &= \sqrt{\sum_{j=k_n+1}^{m_n} (d_j\sigma_j)^{2\nu'}(\xi', v_j)^2} \\
&\leqslant \sqrt{\left(\sum_{j=k_n+1}^{m_n} (d_j\sigma_j)^{2(\nu'+1)}(\xi', v_j)^2\right)^{\frac{\nu'}{\nu'+1}} \left(\sum_{j=k_n+1}^{m_n} (\xi', v_j)^2\right)^{\frac{1}{\nu'+1}}} \\
&\leqslant \rho'^{\frac{1}{\nu'+1}} \left(\sqrt{\sum_{j=k_n+1}^{m_n} (d_j\sigma_j)^2(\hat{x}, v_j)^2}\right)^{\frac{\nu'}{\nu'+1}} \\
&= \rho'^{\frac{1}{\nu'+1}} \left(\sqrt{\sum_{j=k_n+1}^{m_n} d_j^2(\hat{y}, v_j)^2}\right)^{\frac{\nu'}{\nu'+1}} \\
&\leqslant \rho'^{\frac{1}{\nu'+1}} \left(\sqrt{\sum_{j=k_n+1}^{m_n} d_j^2(\bar{Y}_n, u_j)^2} + \sqrt{\sum_{j=k_n+1}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2}\right)^{\frac{\nu'}{\nu'+1}}.
\end{aligned}$$

Thus for

$$\Omega_n := \left\{\sqrt{\sum_{j=k_n}^{m_n} d_{j,n}^2(\bar{Y}_n - \hat{y}, u_j)^2} \leqslant \frac{\delta_n^{\mathrm{est}\prime}}{2},\ |\delta_n^{\mathrm{est}\prime} - \frac{\gamma}{\sqrt{n}}| \leqslant \frac{\gamma}{2\sqrt{n}}, \frac{d_j}{2} \leqslant d_{j,n} \leqslant 2d_j\ \forall\ j \leqslant m_n\right\} \tag{3.15}$$

there holds

$$\left(\sqrt{\sum_{j=k_n+1}^{m_n} d_j^2(\bar{Y}_n,u_j)^2}+\sqrt{\sum_{j=k_n+1}^{m_n} d_j^2(\bar{Y}_n-\hat{y},u_j)^2}\right)^{\frac{\nu'}{\nu'+1}}\chi_{\Omega_n}$$

$$\leqslant 2^{\frac{\nu'}{\nu'+1}}\left(\sqrt{\sum_{j=k_n+1}^{m_n} d_{j,n}^2(\bar{Y}_n,u_j)^2}+\sqrt{\sum_{j=k_n+1}^{m_n} d_{j,n}^2(\bar{Y}_n-\hat{y},u_j)^2}\right)^{\frac{\nu'}{\nu'+1}}\chi_{\Omega_n}$$

$$\leqslant 2^{\frac{\nu'}{\nu'+1}}\left(\delta_n^{\mathrm{est}'}+\frac{\delta_n^{\mathrm{est}'}}{2}\right)^{\frac{\nu'}{\nu'+1}}\chi_{\Omega_n}\leqslant\left(\frac{4\gamma}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}$$

by the definition of Ω_n (3.15) and k_n. Consequently, for the approximation error and the discretisation error there holds

$$\sqrt{\sum_{j=k_n+1}^{m_n}(\hat{x},u_j)^2\chi_{\Omega_n}}+\sqrt{\sum_{j=m_n+1}^{\infty}(\hat{x},u_j)^2}$$

$$\leqslant \rho'^{\frac{1}{\nu'+1}}\left(\frac{4\gamma}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}+\sqrt{\sum_{j=m_n+1}^{\infty}\sigma_j^{2\nu}(\xi,v_j)^2}\leqslant\left(d^{-\nu}\rho\right)^{\frac{1}{\nu'+1}}\left(\frac{4\gamma}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}+\sigma_{m_n}^{\nu}\rho$$

$$\leqslant\frac{L}{2}\max\left(\rho^{\frac{q+1+\varepsilon_2-p}{(\nu+1)q+1+\varepsilon_2-p}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu}{\nu+\frac{q+1+\varepsilon_2-p}{q}}},\rho\left(\frac{1}{\sqrt{n}}\right)^{(1-\varepsilon_1)q\nu}\right)\tag{3.16}$$

for $L=4\max\left((4\gamma)^{\frac{\nu'}{\nu'+1}}d^{-\frac{\nu}{\nu+1}},d^{-\frac{\nu}{\nu+1}},1\right)$ and we obtain

$$\mathbb{P}\left(\sqrt{\sum_{j=k_n+1}^{m_n}(\hat{x},v_j)^2}+\sqrt{\sum_{j=m_n+1}^{\infty}(\hat{x},v_j)^2}\right.$$

$$\left.\leqslant\frac{L}{2}\max\left(\rho^{\frac{q+1+\varepsilon_2-p}{(\nu+1)q+1+\varepsilon_2-p}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu}{\nu+\frac{q+1+\varepsilon_2-p}{q}}},\rho\left(\frac{1}{\sqrt{n}}\right)^{(1-\varepsilon_1)q\nu}\right)\right)$$

$$\geqslant\mathbb{P}\left(\Omega_n\right)\to 1\tag{3.17}$$

as $n\to\infty$, where we used (3.7), (3.8) and lemma 3.1 for Ω_n given in (3.15).

To finish the proof we need to verify a similar bound for the data propagation error. By definition of the discrepancy principle (algorithm 1) and Ω_n in (3.15) there holds

$$\delta_n^{\mathrm{est}'}\chi_{\Omega_n}<\sqrt{\sum_{j=k_n}^{m_n}d_{j,n}^2(\bar{Y}_n,u_j)^2\chi_{\Omega_n}}$$

$$\leqslant\sqrt{\sum_{j=k_n}^{m_n}d_{j,n}^2(\hat{y},u_j)^2\chi_{\Omega_n}}+\sqrt{\sum_{j=k_n}^{m_n}d_{j,n}^2(\bar{Y}_n-\hat{y},u_j)^2\chi_{\Omega_n}}$$

$$< 2\sqrt{\sum_{j=k_n}^{m_n} d_j^2(\hat{y},u_j)^2}\,\chi_{\Omega_n} + \frac{\delta_n^{\mathrm{est}\prime}}{2}\chi_{\Omega_n}$$

$$= 2\sqrt{\sum_{j=k_n}^{m_n}(d_j\sigma_j)^{2(1+\nu')}(\xi',v_j)^2}\chi_{\Omega_n} + \frac{\delta_n^{\mathrm{est}\prime}}{2}\chi_{\Omega_n}$$

$$\leqslant 2\rho'(d_{k_n}\sigma_{k_n})^{\nu'+1}\chi_{\Omega_n} + \frac{\delta_n^{\mathrm{est}\prime}}{2}\chi_{\Omega_n}$$

$$\Rightarrow \frac{1}{d_{k_n}\sigma_{k_n}}\chi_{\Omega_n}$$

$$< \left(\frac{4\rho'}{\delta_n^{\mathrm{est}\prime}}\right)^{\frac{1}{\nu'+1}}\chi_{\Omega_n} \leqslant \left(\frac{16\ d^{-\nu}\rho\sqrt{n}}{\gamma}\right)^{\frac{1}{\nu'+1}}\chi_{\Omega_n} =: b_n\chi_{\Omega_n},$$

where we used that $d_1\sigma_1 \geqslant d_2\sigma_2 \geqslant \ldots$ by definition of d_j. So,

$$\mathbb{P}\left(d_{k_n}\sigma_{k_n} > b_n^{-1}\right) \geqslant \mathbb{P}(\Omega_n) \to 1 \tag{3.18}$$

as $n\to\infty$, for Ω_n given in (3.15). Now we show, that for all $\varepsilon > 0$ it holds that

$$\mathbb{P}\left(\sqrt{\sum_{j=1}^{k_n}\frac{(\bar{Y}_n-\hat{y},u_j)^2}{\sigma_j^2}} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\right) \geqslant 1-3\varepsilon \tag{3.19}$$

for n large enough. Let j_ε be such that $\sum_{j\geqslant j_\varepsilon} j^{-(1+\varepsilon_2)} \leqslant \varepsilon\left(\frac{\gamma}{16d^{-\nu}}\right)^2$ and set $C_\varepsilon := \sqrt{\sum_{j=1}^{j_\varepsilon}\frac{\mathbb{E}(Y_1-\hat{y},u_j)^2}{\sigma_j^2}}$. Define

$$\Omega_{n,\varepsilon} = \left\{ d_{k_n}\sigma_{k_n} \geqslant b_n^{-1},\ \sqrt{\sum_{j=1}^{j_\varepsilon}\frac{(\bar{Y}_n-\hat{y},u_j)^2}{\sigma_j^2}} \leqslant C_\varepsilon\sqrt{n}^{\varepsilon'-1}, \right.$$
$$\left. \times\sqrt{\sum_{j=j_\varepsilon+1}^{\infty} d_j^2(\bar{Y}_n-\hat{y},u_j)^2} \leqslant \left(\frac{\gamma}{16d^{-\nu}}\right)^{\frac{1}{\nu'+1}}\sqrt{n}^{-1}\right\} \tag{3.20}$$

with $\varepsilon' = \frac{1}{2}\frac{1}{\nu'+1}$. Then,

$$\sqrt{\sum_{j=1}^{k_n}\frac{(\bar{Y}_n-\hat{y},u_j)^2}{\sigma_j^2}}\chi_{\Omega_{n,\varepsilon}} \leqslant \sqrt{\sum_{j=1}^{j_\varepsilon}\frac{(\bar{Y}_n-\hat{y},u_j)^2}{\sigma_j^2}}\chi_{\Omega_{n,\varepsilon}} + \sqrt{\sum_{j=j_\varepsilon+1}^{k_n}\frac{d_j^2(\bar{Y}_n-\hat{y},u_j)^2}{d_j^2\sigma_j^2}}\chi_{\Omega_{n,\varepsilon}}$$

$$\leqslant C_\varepsilon\sqrt{n}^{\varepsilon'-1} + \frac{1}{d_{k_n}\sigma_{k_n}}\sqrt{\sum_{j=j_\varepsilon+1}^{k_n} d_j^2(\bar{Y}_n-\hat{y},u_j)^2\chi_{\Omega_{n,\varepsilon}}}$$

$$\leqslant C_\varepsilon\sqrt{n}^{\varepsilon'-1} + b_n\chi_{\Omega_{n,\varepsilon}}\left(\frac{\gamma}{16C_2}\right)^{\frac{1}{\nu'+1}}\sqrt{n}^{-1}$$

$$\leqslant C_\varepsilon \sqrt{n}^{\varepsilon'-1} + \rho^{\frac{1}{\nu'+1}} \sqrt{n}^{\frac{1}{\nu'+1}-1} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}} \left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}$$

for n large enough, since $\varepsilon' < \frac{1}{\nu'+1}$. To prove (3.19) it remains to show that $\mathbb{P}\left(\Omega_{n,\varepsilon}\right) \geqslant 1 - 3\varepsilon$ for n large enough. We apply Markov's inequality and obtain

$$\mathbb{P}\left(\sqrt{\sum_{j=1}^{j_\varepsilon} \frac{\left(\bar{Y}_n - \hat{y}, u_j\right)^2}{\sigma_j^2}} > C_\varepsilon \sqrt{n}^{\varepsilon'-1}\right)$$

$$\leqslant C_\varepsilon^{-2} n^{1-\varepsilon'} \sum_{j=1}^{j_\varepsilon} \frac{\mathbb{E}(\bar{Y}_n - \hat{y}, u_j)^2}{\sigma_j^2} = C_\varepsilon^{-2} n^{-\varepsilon'} \sum_{j=1}^{j_\varepsilon} \frac{\mathbb{E}(Y_1 - \hat{y}, u_j)^2}{\sigma_j^2} = n^{-\varepsilon'} \leqslant \varepsilon \qquad (3.21)$$

for n large enough by definition of C_ε. Further, by the choice of j_ε,

$$\mathbb{P}\left(\sqrt{\sum_{j=j_\varepsilon+1}^{\infty} d_j^2 (\bar{Y}_n - \hat{y}, u_j)^2} > \left(\frac{\gamma}{16 d^{-\nu}}\right)^{\frac{1}{\nu'+1}} \sqrt{n}^{-1}\right)$$

$$\leqslant \left(\frac{16 d^{-\nu}}{\gamma}\right)^{\frac{2}{\nu'+1}} n \sum_{j=j_\varepsilon+1}^{\infty} d_j^2 \mathbb{E}(\bar{Y}_n - \hat{y}, u_j)^2$$

$$= \left(\frac{16 d^{-\nu}}{\gamma}\right)^{\frac{2}{\nu'+1}} \sum_{j=j_\varepsilon+1}^{\infty} d_j^2 \mathbb{E}(Y_1 - \hat{y}, u_j)^2 \qquad (3.22)$$

$$\leqslant \left(\frac{16 d^{-\nu}}{\gamma}\right)^{\frac{2}{\nu'+1}} \sum_{j=j_\varepsilon+1}^{\infty} j^{-(1+\varepsilon_2)} \leqslant \varepsilon. \qquad (3.23)$$

Therefore, by (3.18), (3.21) and (3.22) there holds $\mathbb{P}\left(\Omega_{n,\varepsilon}\right) \geqslant 1 - 3\varepsilon$ for n large enough and thus (3.19). Since $\varepsilon > 0$ was arbitrary it follows that

$$\mathbb{P}\left(\sqrt{\sum_{j=1}^{k_n} \frac{(\bar{Y}_n - \hat{y}, u_j)^2}{\sigma_j^2}} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}} \left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\right) \to 1 \qquad (3.24)$$

as $n \to \infty$. Finally, (3.17) and (3.24) together prove theorem 2.3 for the case, that for all $k \in \mathbb{N}$ there is $j_k \geqslant k$ with $(\hat{y}, u_{j_k}) \neq 0$.

3.2.2. Case 2. Now we assume that there exists $J \in \mathbb{N}$ such that $(\hat{y}, u_j) = 0$ for all $j \geqslant J$. We cannot expect a result similar to lemma 3.1 (since k_n will not converge to ∞ in probability), but the true solution $\hat{x}$ has arbitrarily large smoothness. Let $\varepsilon > 0$ be such that $(d_J \sigma_J)^{-\varepsilon} \leqslant 2$. We set $\nu'' = \nu' + \varepsilon$ and use the representation from (3.6)

$$\hat{x} = \sum_{j=1}^{\infty} \sigma_j^{\nu} (\xi, v_j) = \sum_{j=1}^{K} (d_j \sigma_j)^{\nu''} (d_j \sigma_j)^{-\varepsilon} \frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}} (\xi, v_j) = \sum_{j=1}^{\infty} (d_j \sigma_j)^{\nu''} (\xi'', v_j) v_j, \qquad (3.25)$$

with $\xi'' := \sum_{j=1}^{\infty}(d_j\sigma_j)^{-\varepsilon}\frac{\sigma_j^{\nu-\nu'}}{d_j^{\nu'}}(\xi,v_j)v_j$ and $\|\xi''\| \leqslant 2d^{-\nu}\rho =: \rho''$. We denote

$$\Omega_n := \left\{\frac{d_j}{2} \leqslant d_{j,n} \leqslant 2d_j,\ \forall\ l \leqslant m_n,\ \sqrt{\sum_{j=1}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} \leqslant n^{\frac{\varepsilon'-1}{2}},\ \delta_n^{\mathrm{est}\prime} \leqslant n^{\frac{\varepsilon'-1}{2}}\right\}, \tag{3.26}$$

with $\varepsilon' = \frac{1}{2}\left(\frac{\nu''}{\nu''+1} - \frac{\nu'}{\nu'+1}\right)$. It holds that

$$\mathbb{P}\left(\Omega_n\right) \to 1 \tag{3.27}$$

as $n \to \infty$ because of ((3.7) and (3.8)) and

$$\begin{aligned}\mathbb{P}\left(\sqrt{\sum_{j=1}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} > n^{\frac{\varepsilon'-1}{2}}\right) &\leqslant n^{-\varepsilon'}\sum_{j=1}^{m_n} d_j^2\mathbb{E}(Y_1 - \hat{y}, u_j)^2 \\ &\leqslant n^{-\varepsilon'}\sum_{j=1}^{\infty} j^{-(1+\varepsilon_2)} \to 0\end{aligned}$$

as $n \to \infty$. For n large enough (such that $m_n \geqslant J$) the approximation and discretisation error is

$$\begin{aligned}\sqrt{\sum_{j=k_n+1}^{\infty}(\hat{x}, v_j)^2}\chi_{\Omega_n} &= \sqrt{\sum_{j=k_n+1}^{J}(\hat{x}, v_j)^2}\chi_{\Omega_n} = \sqrt{\sum_{j=k_n+1}^{J}\frac{d_{j,n}^2(\hat{y}, u_j)^2}{d_{j,n}^2\sigma_j^2}}\chi_{\Omega_n} \\ &\leqslant \frac{1}{d_{J,n}\sigma_J}\left(\sqrt{\sum_{j=k_n+1}^{J} d_{j,n}^2(\bar{Y}_n, u_j)^2}\right. \\ &\quad \left.+ \sqrt{\sum_{j=k_n+1}^{J} d_{j,n}^2(\bar{Y}_n - \hat{y}, u_j)^2}\right)\chi_{\Omega_n} \\ &\leqslant \frac{1}{d_{J,n}\sigma_J}\left(\sqrt{\sum_{j=k_n+1}^{m_n} d_{j,n}^2(\bar{Y}_n, u_j)^2}\right. \\ &\quad \left.+ \sqrt{\sum_{j=1}^{m_n} d_{j,n}^2(\bar{Y}_n - \hat{y}, u_j)^2}\right)\chi_{\Omega_n} \\ &\leqslant \frac{2}{d_J\sigma_J}\left(\delta_n^{\mathrm{est}\prime} + 2\sqrt{\sum_{j=1}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2}\right)\chi_{\Omega_n} \leqslant 2^{\frac{1}{\varepsilon}+1} \\ &\quad \times\left(n^{\frac{\varepsilon'-1}{2}} + 2n^{\frac{\varepsilon'-1}{2}}\right) \\ &\leqslant 2^{\frac{1}{\varepsilon}+3}n^{\frac{\varepsilon'-1}{2}} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\end{aligned}$$

for n large enough, where we used $d_J\sigma_J \geqslant 2^{-\frac{1}{\varepsilon}}$ in the sixth step, the definition of the discrepancy principle in the fifth step and

$$\varepsilon' - 1 = -\frac{\nu'}{\nu'+1} + \frac{1}{2}\frac{\nu''}{\nu''+1} + \frac{1}{2}\frac{\nu'}{\nu'+1} - 1 > -\frac{\nu'}{\nu'+1} + \frac{1}{2} + \frac{1}{2} - 1 = -\frac{\nu'}{\nu'+1}$$

in the last step. We therefore obtain

$$\mathbb{P}\left(\sqrt{\sum_{j=k_n+1}^{\infty}(\hat{x},v_j)^2} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\right) \geqslant \mathbb{P}\left(\Omega_n\right) \to 1 \tag{3.28}$$

as $n\to\infty$. It remains to treat the data propagation error. We set $b_n := \left(\frac{\gamma}{8\rho''\sqrt{n}}\right)^{\frac{1}{\nu''+1}}$. Let the deterministic sequence $(q_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ be defined via

$$q_n := \min\left\{j \leqslant m_n \ : \ d_j\sigma_j \leqslant b_n\right\}. \tag{3.29}$$

If $b_n > d_j\sigma_j$ for all $j = 1,\ldots,m_n$ we set $q_n =: m_n$. Note that $q_n \to \infty$ as $n\to\infty$, since $b_n \to 0$. Define

$$\bar{\Omega}_n := \left\{\sqrt{\sum_{j=q_n}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} < \gamma/\sqrt{8n}\right\} \cap \Omega_n \tag{3.30}$$

with Ω_n from (3.26). We claim that

$$k_n\chi_{\bar{\Omega}_n} \leqslant q_n - 1 \tag{3.31}$$

as $n\to\infty$. Since (3.31) trivially holds for $q_n = m_n$, we may assume that $d_{q_n}\sigma_{q_n} \leqslant b_n$. Further,

$$\begin{aligned}
\sqrt{\sum_{j=q_n}^{m_n} d_{j,n}^2(\bar{Y}_n,u_j)^2}\chi_{\bar{\Omega}_n} &\leqslant \left(\sqrt{\sum_{j=q_n}^{m_n} d_{j,n}^2(\hat{y},u_j)^2} + \sqrt{\sum_{j=q_n}^{m_n} d_{j,n}^2(\bar{Y}_n-\hat{y},u_j)^2}\right)\chi_{\bar{\Omega}_n}\\
&\leqslant 2\left(\sqrt{\sum_{j=q_n}^{m_n} d_j^2(\hat{y},u_j)^2} + \sqrt{\sum_{j=q_n}^{m_n} d_j^2(\bar{Y}_n-\hat{y},u_j)^2}\right)\chi_{\bar{\Omega}_n}\\
&\leqslant 2\left(\sqrt{\sum_{j=q_n}^{m_n} (d_j\sigma_j)^{2(\nu''+1)}(\xi'',v_j)^2} + \frac{\gamma}{8\sqrt{n}}\right)\chi_{\bar{\Omega}_n}\\
&\leqslant 2\left((d_{q_n}\sigma_{q_n})^{\nu''+1}\rho'' + \frac{\gamma}{8\sqrt{n}}\right)\chi_{\bar{\Omega}_n}\\
&\leqslant \left(2b_n^{\nu''+1}\rho'' + \frac{\gamma}{4\sqrt{n}}\right)\chi_{\bar{\Omega}_n} \leqslant \left(\frac{\gamma}{4\sqrt{n}} + \frac{\gamma}{4\sqrt{n}}\right)\chi_{\bar{\Omega}_n} \leqslant \delta_n^{\mathrm{est}\prime}
\end{aligned}$$

and (3.31) follows by the definition of k_n in algorithm 1. It holds that

$$\mathbb{P}\left(\bar{\Omega}_n\right) \to 1 \tag{3.32}$$

as $n \to \infty$ because of (3.27) and

$$\mathbb{P}\left(\sqrt{\sum_{j=q_n}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2} > \frac{\gamma}{\sqrt{8n}}\right) \leqslant \frac{8n}{\gamma^2}\sum_{j=q_n}^{m_n} d_j^2 \mathbb{E}(\bar{Y}_n - \hat{y}, u_j)^2 \leqslant \frac{8}{\gamma^2}\sum_{j=q_n}^{m_n} j^{-(1+\varepsilon_2)} \to 0$$

as $n \to \infty$. Finally,

$$\begin{aligned}
\sqrt{\sum_{j=1}^{k_n} \frac{(\bar{Y}_n - \hat{y}, u_j)^2}{\sigma_j^2}}\chi_{\bar{\Omega}_n} &= \sqrt{\sum_{j=1}^{k_n} \frac{d_j^2(\bar{Y}_n - \hat{y}, u_j)^2}{d_j^2\sigma_j^2}}\chi_{\bar{\Omega}_n} \\
&\leqslant \frac{1}{d_{k_n}\sigma_{k_n}}\sqrt{\sum_{j=1}^{k_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2\chi_{\bar{\Omega}_n}} \\
&\leqslant \frac{1}{b_n}\sqrt{\sum_{j=1}^{m_n} d_j^2(\bar{Y}_n - \hat{y}, u_j)^2\chi_{\bar{\Omega}_n}} \\
&\leqslant \frac{1}{b_n} n^{\frac{\varepsilon'-1}{2}} \leqslant \left(\frac{8\rho''}{\gamma}\right)^{\frac{1}{\nu''+1}} \sqrt{n}^{\frac{1}{\nu''+1}+\varepsilon'-1} \\
&\leqslant \left(\frac{16 d^{-\nu}\rho}{\gamma}\right)^{\frac{1}{\nu''+1}} \sqrt{n}^{-\frac{1}{2}\left(\frac{\nu''}{\nu''+1}+\frac{\nu'}{\nu'+1}\right)} \\
&\leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}},
\end{aligned}$$

for n large enough, where we used $d_{k_n}\sigma_{k_n}\chi_{\bar{\Omega}_n} > b_n\chi_{\bar{\Omega}_n}$ (which follows from $k_n\chi_{\bar{\Omega}_n} \leqslant q_n - 1$ and (3.29)) in the third, the definition of Ω_n in the fourth, $\rho'' = 2d^{-\nu}\rho$ and $\varepsilon' = \frac{1}{2}\left(\frac{\nu''}{\nu''+1} - \frac{\nu'}{\nu'+1}\right)$ in the sixth and $\nu'' > \nu'$ in the last step. Thus

$$\mathbb{P}\left(\sqrt{\sum_{j=1}^{k_n} \frac{(\bar{Y}_n - \hat{y}, u_j)^2}{\sigma_j^2}} \leqslant \frac{L}{2}\rho^{\frac{1}{\nu'+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu'}{\nu'+1}}\right) \geqslant \mathbb{P}\left(\bar{\Omega}_n\right) \to 1 \tag{3.33}$$

as $n \to \infty$. Both (3.28) and (3.33) together yield the claim of theorem 2.3 for the case that there is a $J \in \mathbb{N}$ such that $(\hat{y}, u_j) = 0$ for all $j \geqslant J$.

4. Numerical demonstration

We now numerically test the modified discrepancy principle for the toy problem 'deriv2' from the open source MATLAB package Regutools [10]. This is a discretisation of a 1d-Fredholm integral equation by means of the Galerkin approximation with box functions. The resulting discrete problem reads $A\hat{x} = b$, with $\hat{x}, b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times m}$. We perturbed the right-hand side component wise according to

$$z_i = b_i + \frac{\|b\|}{\sqrt{m}}\delta_i,$$

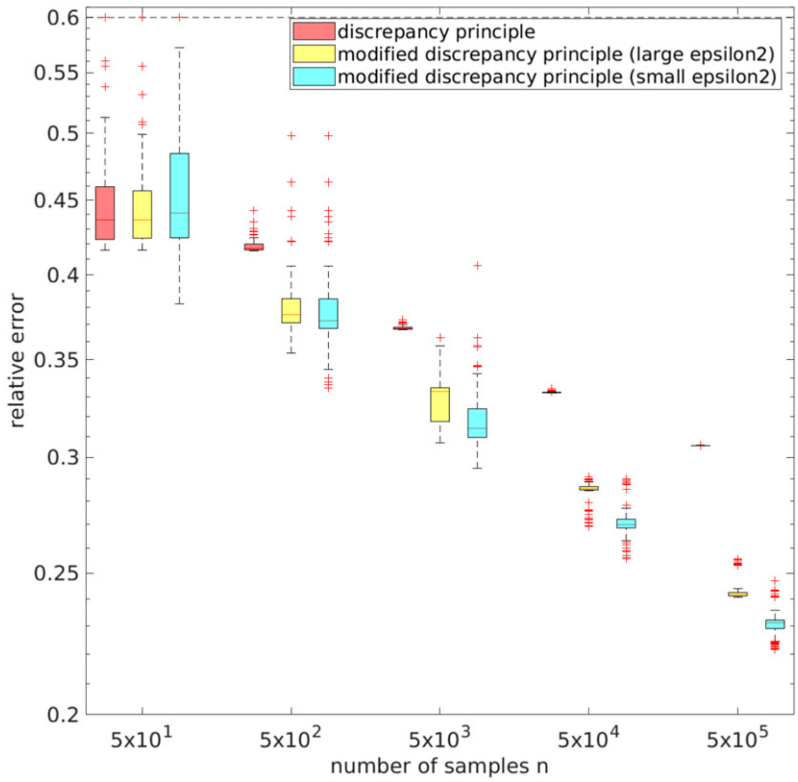

Figure 1. Relative errors for (modified) discrepancy principle. We see that the resulting errors decay faster for the modified discrepancy principle.

Table 1. Sampled median of the relative error of the plain and modified discrepancy principle (dp) and sampled relative minimax risk for different sample sizes.

Sample size n	Plain dp Median error	Modified dp, ε_2 large Median error	Modified dp, ε_2 small Median error	Oracle Square root of the minimax risk
5×10^1	4.36×10^{-1}	4.36×10^{-1}	4.4×10^{-1}	4.48×10^{-1}
5×10^2	4.17×10^{-1}	3.75×10^{-1}	3.72×10^{-1}	3.69×10^{-1}
5×10^3	3.67×10^{-1}	3.32×10^{-1}	3.14×10^{-1}	3.07×10^{-1}
5×10^4	3.32×10^{-1}	2.85×10^{-1}	2.69×10^{-1}	2.54×10^{-1}
5×10^5	3.05×10^{-1}	2.41×10^{-1}	2.32×10^{-1}	2.09×10^{-1}

where the δ_i are centralised i.i.d. random variables following a generalised Pareto-distribution with finite fourth moment, but infinite higher moments (function gprnd(K, σ, θ, m, n) with $K = 1/5$, $\sigma = \sqrt{(1-K)^2(1-2K)}$ and $\theta = 0$). We consider the symmetrised equation (as in example 2.1) and set $K := A^*A$ with i.i.d. measurements $Y_1, Y_2, \ldots$ distributed as

$$Y_1 \stackrel{d}{=} A^* \begin{pmatrix} z_1 \\ \ldots \\ z_m \end{pmatrix}.$$

We verify that the condition for the fourth moments in theorem 2.3 is satisfied. Indeed, it holds that

$$\sup_j \frac{\mathbb{E}[\left(Y_1 - \hat{y}, v_j\right)^4]}{(\mathbb{E}[(Y_1-\hat{y}, v_j)^2])^2}$$

$$= \sup_j \frac{\mathbb{E}\left[\left(A^*\begin{pmatrix}\delta_1\\ \delta_2\\ \dots\end{pmatrix}, v_j\right)^4\right]}{\left(\mathbb{E}\left[\left(A^*\begin{pmatrix}\delta_1\\ \delta_2\\ \dots\end{pmatrix}, v_j\right)^2\right]\right)^2} = \sup_j \frac{\mathbb{E}\left[\left(\left(\begin{pmatrix}\delta_1\\ \delta_2\\ \dots\end{pmatrix}, u_j\right)\right)^4\right]}{\left(\mathbb{E}\left[\left(\left(\begin{pmatrix}\delta_1\\ \delta_2\\ \dots\end{pmatrix}, u_j\right)\right)^2\right]\right)^2}$$

$$= \sup_j \frac{\mathbb{E}\left[\left(\sum_l \delta_l (e_l, u_j)\right)^4\right]}{\left(\mathbb{E}\left[\left(\sum_l \delta_l (e_l, u_j)\right)^2\right]\right)^2}$$

$$= \sup_j \frac{\mathbb{E}[\delta_1^4]\sum_l (e_l, u_j)^4 + 3\left(\mathbb{E}[\delta_1^2]\right)^2 \sum_{\substack{l,l'\\ l\neq l'}} (e_l, u_j)^2 (e_{l'}, u_j)^2}{\left(\mathbb{E}[\delta_1^2]\sum_l (e_l, u_j)^2\right)^2}$$

$$\leqslant \frac{\mathbb{E}[\delta_1^4]}{\left(\mathbb{E}[\delta_1^2]\right)^2} \sup_j \frac{\sum_l (e_l, u_j)^4 + 3\sum_{\substack{l,l'\\ l\neq l'}} (e_l, u_j)^2 (e_{l'}, u_j)^2}{\sum_l (e_l, u_j)^4 + \sum_{\substack{l,l'\\ l\neq l'}} (e_l, u_j)^2 (e_{l'}, u_j)^2} \leqslant 4\frac{\mathbb{E}[\delta_1^4]}{(\mathbb{E}[\delta_1^2])^2},$$

where $e_1, e_2, \dots$ is the (orthonormal) Galerkin basis and $\hat{y} = A^* b$. We set the discretization to $m = 1000$ and approximated the singular value decomposition (σ_j, u_j, v_j) of A with the function 'csvd'. We used $n = [50, 500, \dots, 500\,000]$ measurements and compared the classical discrepancy principle to the modified one implemented in algorithm 1 with $\varepsilon_1 = 0.5$ and $\varepsilon_2 = 0.5$ (large) and $\varepsilon_2 = 0.1$ (small). We calculated the relative errors for 100 independent runs and visualised the results as box plots in figure 1. We clearly see that the errors decay faster for the modified discrepancy principle. Moreover, in table 1 we compare the (relative) median error of the plain and modified discrepancy principle (this is the red bar in each of the boxes) to the square root of the minimax risk from theorem 2.1, where we sampled the latter from the same data. We see that the error of the modified discrepancy principle is comparable to the minimax risk for smaller sample sizes. For larger sample sizes the minimax risk is better, which is consistent with the loss of ε_2 in the exponent of (2.6).

5. Concluding remarks

In this work we have presented a modified discrepancy principle, which yields (almost) optimal convergence rates for arbitrary unknown error distributions, if one is able to repeat the measurements. This was achieved in estimating the variances of one measurement along the singular directions of the operator K, which was then used to rescale the measurements and the operator.

We restricted to linear mildly ill-posed problems and classical Hölder-type source conditions in Hilbert spaces, but the results probably can be extended to general degree of ill-posedness and general source conditions. A major drawback is that the singular value decomposition of the operator needs to be known. It would be interesting to investigate whether the approach could be adapted to settings, where the singular valued decomposition is not given.

Acknowledgments

The author would like to thank Dr Peter Mathé for kindly hosting him in Berlin, where we worked out the basic idea of the presented approach.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Tim Jahn https://orcid.org/0000-0002-7633-8265

References

[1] Angelova J A 2012 On moments of sample mean and variance *Int. J. Pure Appl. Math.* **79** 67–85
[2] Bakushinskii A B 1984 Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion *USSR Comput. Math. Math. Phys.* **24** 181–2
[3] Bauer F and Reiß M 2008 Regularization independent of the noise level: an analysis of quasi-optimality *Inverse Problems* **24** 055009
[4] Belitser E N and Levit B Y 1995 On minimax filtering over ellipsoids *Math. Methods Stat.* **4** 259–73
[5] Bevington P R and Robinson D K 2003 *Data Reduction and Error Analysis* (New York: McGraw-Hill)
[6] Bissantz N, Hohage T, Munk A and Ruymgaart F 2007 Convergence rates of general regularization methods for statistical inverse problems and applications *SIAM J. Numer. Anal.* **45** 2610–36
[7] Blanchard G, Hoffmann M and Reiß M 2018 Early stopping for statistical inverse problems via truncated SVD estimation *Electron. J. Stat.* **12** 3204–31
[8] Blanchard G and Mathé P 2012 Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration *Inverse Problems* **28** 115011
[9] Cavalier L 2011 Inverse problems in statistics *Inverse Problems and High-Dimensional Estimation* (Berlin: Springer) pp 3–96
[10] Hansen P C 2007 Regularization tools version 4.0 for Matlab 7.3 *Numer. Algorithms* **46** 189–94
[11] Harrach B, Jahn T and Potthast R 2020 Beyond the Bakushinkii veto: regularising linear inverse problems without knowing the noise distribution *Numer. Math.* **145** 581–603
[12] Harrach B, Jahn T and Potthast R 2020 Regularising linear inverse problems under unknown non-Gaussian white noise (arXiv:2010.04519)
[13] Jahn T 2021 Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise (arXiv:2104.06184)
[14] Jahn T 2021 Regularising linear inverse problems under unknown non-Gaussian noise *PhD Thesis* Universitätsbibliothek Johann Christian Senckenberg http://dx.doi.org/10.21248/gups.60931
[15] Jin Q and Mathé P 2013 Oracle inequality for a statistical Raus–Gfrerer-type rule *SIAM/ASA J. Uncertain. Quantification* **1** 386–407
[16] Li H and Werner F 2020 Empirical risk minimization as parameter choice rule for general linear regularization methods *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **56** 405–27

[17] Lu S and Mathé P 2014 Discrepancy based model selection in statistical inverse problems *J. Complexity* **30** 290–308

[18] Lyons R G 2004 *Understanding Digital Signal Processing* 3rd edn (Upper Saddle River, NJ: Pearson Education)

[19] Mathé P and Pereverzev S V 2003 Discretization strategy for linear ill-posed problems in variable Hilbert scales *Inverse Problems* **19** 1263

[20] Mika G and Szkutnik Z 2021 Towards adaptivity via a new discrepancy principle for Poisson inverse problems *Electron. J. Stat.* **15** 2029–59

[21] 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

[22] Pinsker M S 1980 Optimal filtration of square-integrable signals in Gaussian noise *Probl. Inf. Transm.* **16** 120–33

[23] Taylor J 1997 *Introduction to Error Analysis* (*The Study of Uncertainties in Physical Measurements*) (Sausalito, California: University Science Books)