Optimal Convergence of the Discrepancy Principle for Polynomially and Exponentially Ill-Posed Operators under White Noise

Tim Jahn

Institute for Numerical Simulation and Hausdorff Center for Mathematics, University of Bonn

ABSTRACT

We consider a linear ill-posed equation in the Hilbert space setting under white noise. Known convergence results for the discrepancy principle are either restricted to Hilbert-Schmidt operators (and they require a self-similarity condition for the unknown solution $\hat{x}$, additional to a classical source condition) or to polynomially ill-posed operators (excluding exponentially ill-posed problems). In this work, we show optimal convergence for a modified discrepancy principle for both polynomially and exponentially ill-posed operators (without further restrictions) solely under either Hölder-type or logarithmic source conditions. In particular, the method includes only a single simple hyper parameter, which does not need to be adapted to the type of ill-posedness.

ARTICLE HISTORY

Received 14 April 2021
Revised 30 November 2021
Accepted 30 November 2021

KEYWORDS

Statistical inverse problems; non-Bayesian approach; discrepancy principle; convergence; optimality

1. Introduction

Let $K : \mathcal{X} \to \mathcal{Y}$ be a compact operator with dense range between infinite-dimensional Hilbert spaces. We aim to solve the following equation

$$Kx = y^\delta$$

for $y^\delta = \hat{y} + \delta Z$ a noisy perturbation of the unknown true data $\hat{y} = K\hat{x}$, with $\hat{x} = K^+\hat{y}$ the minimum norm solution (thus $K^+$ denotes the Moore-Penrose inverse of $K$). Here, $Z$ is white noise, i.e. it holds that

$$\mathbb{E}[(Z,y)] = 0 \quad \text{and} \quad \mathbb{E}[(Z,y)(Z,y')] = (y,y')$$

and $(Z,y)$ and $(Z,y')$ have the same distribution for all $y, y' \in \mathcal{Y}$. The parameter $\delta > 0$ denotes the noise level and controls the variance. Note that the above is only a formal notation, since $y^\delta$ will not be an element of the Hilbert space with probability 1. In order to obtain a well-defined approximation to the unknown solution $\hat{x}$ we use spectral cutoff thus assume, that we know $(\sigma_j, u_j, v_j)$ the singular value decomposition of $K$, i.e. $\sigma_1 \geq \sigma_2 \geq ... > 0$ is a
decreasing sequence, \((u_j)_{j \in \mathbb{N}}, (v_j)_{j \in \mathbb{N}}\) are orthonormal bases of \(\mathcal{Y}\) and \(\mathcal{N}(K)^\perp \subset \mathcal{X}\) respectively, and there holds \(Kv_j = \sigma_j u_j, K^* u_j = \sigma_j v_j\). Consequently,

\[ x^\delta_k := \sum_{j=1}^k \frac{(y^\delta, u_j)}{\sigma_j} v_j \]

is our approximation to \(\hat{x}\), and the task is to determine a good choice of the truncation level \(k\), dependent on the measurement \(y^\delta\) and the noise level \(\delta\). Because the sum \(\sum_{j=1}^\infty (Z, u_j)^2\) is almost surely infinite, we cannot directly apply the discrepancy principle [1] to determine \(k\). We thus truncate the sum and discretise (as an additional regularization). Specifically, for \(m \in \mathbb{N}\) it holds that \(\mathbb{E} \left[ \sum_{j=1}^m (y^\delta - \hat{y}, u_j)^2 \right] = \sum_{j=1}^m \delta^2 \mathbb{E} (Z, u_j)^2 = m \delta^2\), so we define for fixed \(\tau > 1\) the classical discrepancy principle to the discretized measurements

\[ k^\delta_{dp}(m) := \min \left\{ 0 \leq k \leq m \ : \ \sqrt{\sum_{j=k+1}^m (y^\delta, u_j)^2} \leq \tau \sqrt{m \delta} \right\} \]  

(1.1)

In order to determine our final approximation, we have to choose the discretization level \(m\). The main results (Theorem 2.3 and 2.5) state, that for the adaptive choice

\[ k^\delta_{dp} := \max_{m \in \mathbb{N}} k^\delta_{dp}(m) \]  

(1.2)

we obtain the either order optimal or even asymptotically optimal rate of convergence (in probability) under the natural source conditions for the two main types of linear ill-posed problems. These are polynomial ill-posed and exponentially ill-posed problems (also called mildly and severely ill-posed, respectively). For the sake of simplicity we assume that, in the first case, the singular values fulfill \(\sigma_j^2 = j^{-q}\) for some \(q > 0\) and in the second case \(\sigma_j^2 = e^{-ag}\) for \(a > 0\). Note that for polynomial ill-posed problems we do not assume that \(q > 1/2\), so \(K\) might be a non-Hilbert-Schmidt operator. In practice, the singular values of the problem are not analytically given and usually not exactly of the above specific types. So for general decaying singular values we will at least prove in Theorem 2.1 that the method is convergent, without giving rates. An important practical problem of our approach (1.1) is, that one cannot perform the maximization over all \(m \in \mathbb{N}\) in (1.2). We will comment on that and the fact that \(k^\delta_{dp}\) is well-defined below.

As mentioned above, convergence rates are obtained only under additional source conditions. These are certain subspace of \(\mathcal{X}\), in which the true solution is supposed to reside. For polynomial ill-posed problems,
these are Holder-type conditions (see e.g. [2] and the references therein)

\[ \mathcal{X}_{\nu, \rho} := \left\{ (K^*K)\frac{\xi}{\rho} : \xi \in \mathcal{X} \right\} \tag{1.3} \]

for \( \nu, \rho > 0 \) (so \( \hat{x} \in \mathcal{X}_{\nu, \rho} \) has the following representation \( \hat{x} = \sum_{j=1}^{\infty} \sigma_j^* (\xi_j v_j) v_j \) with \( ||\xi|| \leq \rho \)). In the scenario of exponentially ill-posed problems it is well known, that Holder-type smoothness conditions are often too restrictive. For example, for the severely ill-posed problem of the inverse heat equation any Holder-type source condition for the true solution \( \hat{x} \) would imply that it is infinitely often differentiable. Here, a natural choice are so-called logarithmic source conditions [3]

\[ \mathcal{X}_{p, q} := \left\{ \frac{-\log (K^*K)^{\frac{p}{q}} \xi}{\rho} : \xi \in \mathcal{X}, \ ||\xi|| \leq \rho \right\} \tag{1.4} \]

for \( p, q > 0 \).

In the literature plenty of work has been done on linear ill-posed problems under white noise, see e.g. [4] and [5] for an overview. Among the first adaptive methods studied were, cross validation [6], unbiased or penalized empirical risk minimization [7, 8], Lepski’s balancing principle [9], and others. Here, optimal rates are usually obtained only up to a logarithmic factor. This is mostly due to the fact that the deduced rates are in \( L^2 \) (also called integrated mean squared error), whereas we prove rates in probability which is a weaker type of convergence. For example, in [10] bounds in probability are deduced for Lepski’s principle without a logarithmic factor in the context of inverse learning. It is notably that the hyperparameters for the methods have to be chosen differently for mildly and severely ill-posed problems. Our proposed modified discrepancy principle has only one free parameter \( \tau \), which can be chosen independently of the degree of ill-posedness. This might be beneficial in practice, since usually the singular values of the problem will not behave exactly as the both cases considered here, and thus might not be classified unambiguously. We want to stress here that for exponentially ill-posed problems we do not only obtain convergence rates of optimal order, but asymptotically the optimal convergence rate, i.e. the constant in the convergence rate will converge to 1. This is notable in light of the fact that for exponentially ill-posed problems under deterministic noise the classical discrepancy principle will only provide order-optimal rates (see [11]) and has to be adapted properly to obtain the optimal rate asymptotically (see [3]). Moreover, we do not loose a logarithmic factor in the convergence rates (which is mainly due to the fact that we consider a different type of convergence). One advantage of the discrepancy principle compared to the above mentioned other parameter choice rules is that it potentially allows a simple and low-cost implementation. For example, in Ref. [12], the superiority (in terms of computational complexity) of the discrepancy principle over Lepski’s balancing principle is presented in a particular framework. More recently
variants of the discrepancy principle were studied for statistical inverse problems. In Refs. [13, 14], a modified discrepancy principle was introduced. It is based on symmetrization and thus restricted to Hilbert-Schmidt operators. Also, the true solution \( \hat{x} \) has to fulfill a self-similarity condition. Relatively new approaches in Refs. [15, 16] also use Discretization to apply the discrepancy principle, as we do here. There the main goal was to minimize computational costs and an early stopping property of a modified discrepancy principle is elaborated. Optimal rates are achieved for polynomially ill-posed problems and the true solution must not be too smooth. Another modification of this approach trying to overcome the saturation effect through an additional smoothing of the residuals is presented in Ref. [17]. The method proposed in these three articles will not work directly for exponentially ill-posed problems, as explained e.g. in Remark 3.9 of Ref. [16]. However, the computational costs there are substantially smaller than in our case since we have to calculate all the singular vectors. It would be interesting to study whether one could combine the benefits of both methods. Also, an extension of our discrepancy principle to computationally effective methods as Landweber method and more practical Discretization spaces (cf. [12]) would be highly desirable. Finally, we want to mention that in the above papers usually bounds in \( L^2 \) are provided, under the assumption that the white noise is Gaussian (e.g., that \( \mathbb{E}(Z, y) \) is Gaussian for all \( y \in \mathcal{Y} \)). This is often the reason for the logarithmic correction term in the rates mentioned above. We assume solely a finite second moment, but provide only rates which hold with high probability. It would be interesting whether the approach could be adapted such that it provides \( L^2 \) rates under Gaussian noise. Also note, that we assume that the noise level \( \delta \) is ad hoc known. In case we have access to multiple measurements we may drop this assumption and use the average of those measurements as our data, and estimate the noise level in a natural way, see e.g. [18–20]. Hereby a careful choice of the repetitions allows to obtain any prescribed level of accuracy efficiently, see also [21].

**Remark 1.1.** First of all it is not directly clear, that \( k_{dp}^{\delta} \) is well-defined. However, since

\[
\sqrt{\sum_{j=1}^{m} (y^{\delta}, u_k)^2} \leq \sqrt{\sum_{j=1}^{m} (\hat{y}, u_k)^2} + \sqrt{\sum_{j=1}^{m} (\hat{y} - y^{\delta}, u_j)^2} \approx ||\hat{y}|| + \sqrt{m\delta} \leq \tau \sqrt{m\delta}
\]

for \( m \) large, we see that \( k_{dp}^{\delta}(m) \to 1 \) (a.s.) as \( m \to \infty \) (compare to the proof of Proposition 3.1). This assures that \( k_{dp}^{\delta} < \infty \) a.s. and moreover, that there exists (random) \( m(\delta, ||\hat{y}||) \) with

\[
\max_{m \in \mathbb{N}} k_{dp}^{\delta}(m) = \max_{m \leq m(\delta, ||\hat{y}||)} k_{dp}^{\delta}(m).
\]
It would be desirable to have a rough idea of how large \( m(\delta, \rho) \) will be, unfortunately we cannot give a satisfying solution for that. One natural idea to obtain an upper bound would be to balance the measurement error \( \sqrt{\sum_{j=1}^{m} (y^d - \hat{y}, u_j)^2} \) and the discretization error \( \sqrt{\sum_{j=m+1}^{\infty} (\hat{y}, u_j)^2} \). This would mean to determine \( m \) such that roughly

\[
\sqrt{m \delta} \approx \sigma_m |\hat{x}|,
\]

since \( (\hat{y}, u_j) = \sigma_j (\hat{x}, v_j) \). However, this can only be achieved if at least an upper bound for \( |\hat{x}| \) is available. Another (heuristic) possibility is based on the observation, that for \( m \) small we have that \( k^\delta_{dp}(m) \approx m \), whereas \( k^\delta_{dp}(m) \to 1 \) as \( m \to \infty \). Thus we could increase \( m \) gradually, until \( k^\delta_{dp}(m)/m \) is small.

2. Main results

We formulate the first Theorem, which states that our modified discrepancy principle resembles a convergent regularization method for arbitrary compact \( K \) with dense image.

**Theorem 2.1.** Assume that \( K \) is compact with dense range and let \( \hat{y} \in \mathcal{R}(K) \). Let \( \tau > 1 \) and let \( k^\delta_{dp} \) be the truncation level determined by the discrepancy principle as in (1.1) and (1.2). Then, for all \( \varepsilon > 0 \) there holds

\[
\mathbb{P}
( \|x^\delta_{k^\delta_{dp}} - \hat{x}\| \leq \varepsilon ) \to 1
\]

as \( \delta \to 0 \).

From now we restrict to polynomially and exponentially ill-posed operators. We first calculate the optimal a priori rate, to which we afterwards compare the rate of the discrepancy principle.

**Theorem 2.2.** Assume that the problem is polynomially ill-posed, i.e. \( \sigma_j^2 = j^{-q} \) for some \( q > 0 \). Then there holds

\[
\inf_{k \in \mathbb{N}} \sup_{\hat{x} \in \mathcal{X}_{\nu, \rho}} \sqrt{\mathbb{E}\|x^\delta_k - \hat{x}\|^2} \lesssim \rho^{\frac{q+1}{q+1+q}} \delta^{\frac{q+1}{q+1+q}},
\]

with \( \mathcal{X}_{\nu, \rho} \) given in (1.3).

**Proof of Theorem 2.2.** The proof is standard. We split the total error in a customary way into a data propagation error and an approximation error (also called variance and bias here, cf (3.1))
\[ \sup_{\hat{x} \in X_{\nu, \rho}} \mathbb{E}[\|x_\delta^\rho - \hat{x}\|^2] = \sup_{\hat{x} \in X_{\nu, \rho}} \sum_{j=1}^{k} \frac{\mathbb{E}(y_\delta^\rho - \hat{y}, u_j)^2}{\sigma_j^2} + \sup_{\hat{x} \in X_{\nu, \rho}} \sum_{j=k+1}^{\infty} (\hat{x}, v_j)^2 \]

\[ = \delta^2 \sum_{j=1}^{k} \mathbb{E}(Z, u_j)^2 + \sup_{\xi \in X} \sum_{j=k+1}^{\infty} \sigma^{2\nu}(\xi, v_j)^2 \]

\[ = \delta^2 \sum_{j=1}^{k} + \sigma^{2\nu}_{k+1} \rho^2 \approx \delta^2 k^{q+1} + k^{-q_2} \rho^2. \]

The right-hand side is minimized (up to a constant factor) by a choice fulfilling

\[ k \asymp \left( \frac{\rho}{\delta} \right)^{\frac{q+1}{q+1}}, \tag{2.1} \]

which yields the rate from Theorem 2.2.

The above optimal a priori choice (2.1) depends on the unknown smoothness parameter \( \nu \) and \( \rho \) and hence is not practical. The next Theorem assures optimal adaptivity of our modified discrepancy principle, in the sense that the optimal rate from Theorem 2.2 holds in probability up to a constant (order-optimal convergence).

**Theorem 2.3.** Assume that the problem is polynomially ill-posed, i.e. \( \sigma_j^2 = j^{-q} \) for \( q > 0 \). Let \( \tau > 1 \) and let \( k^\delta_{dp} \) be the truncation level determined by the discrepancy principle as in (1.1) and (1.2). Then there holds

\[ \sup_{\hat{x} \in X_{\nu, \rho}} \mathbb{P}\left( \|x_\delta^\rho - \hat{x}\| \leq L_{\nu, \rho} \right) \xrightarrow[\delta/\rho \to 0]{} 1, \]

as \( \delta/\rho \to 0 \), with \( L_{\nu, \rho} := \left( \frac{2}{\tau - 1} + 1 \right)^{\frac{q+1}{q+1}} + \left( \frac{\tau + 1}{\tau - 1} \right)^{\frac{q+1}{q+1}} + 1 \) and \( X_{\nu, \rho} \) given in (1.3).

The proof is deferred to Section 3.

We now discuss the case of exponentially ill-posed problems and again calculate the optimal possible rate first.

**Theorem 2.4.** Assume that the problem is exponentially ill-posed, i.e. \( \sigma_j^2 = e^{-a_j} \) for \( a > 0 \). Then there holds

\[ \inf_{k \in \mathbb{N}} \sup_{\hat{x} \in X_{\nu, \rho}} \sqrt{\mathbb{E}[\|x_\delta^\rho - \hat{x}\|^2]} = \rho \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-\frac{q}{2}} \]

as \( \frac{\delta}{\rho} \to 0 \), with \( X_{\nu, \rho} \) given in (1.4).

We assure adaptivity of the above modified discrepancy principle in this case. Here, we even obtain asymptotically optimal convergence, i.e. the optimal rate holds (asymptotically) up to a multiplicative constant of 1. This is important, because of the very slow convergence under logarithmic source conditions.
Theorem 2.5. Assume that the problem is exponentially ill-posed, i.e. 
\[ \sigma_j^2 = e^{-aj} \]. Let \( \tau > 1 \) and \( k_{dp}^\delta \) be the truncation level determined by the discrepancy principle as in (1.1) and (1.2). Then there holds

\[
\sup_{\hat{x} \in X_{p,\rho}} \mathbb{P}\left( \|x_{k_{dp}^\delta} - \hat{x}\| \leq \rho \left( -\log \left( \frac{\sigma_j^2}{\rho^2} \right) \right) \left( 1 + o(1) \right) \right) \to 1,
\]
as \( \delta / \rho \to 0 \), with \( X_{p,\rho} \) given in (1.4).

The proofs of Theorem 2.4 and 2.5 are presented in Section 3.

Remark 2.1. The conditions for the singular values can be weakened to 
\[ c_q j^{-q} \leq \sigma_j^2 \leq C_j q^{-q} \] for all \( j \) large enough (and \( c_q, C_q, q > 0 \)) in the case of polynomially ill-posed problems, and to 
\[ c_a e^{-aj} \leq \sigma_j^2 \leq C_a e^{-aj} \] for all \( j \) large enough (and \( c_a, C_a, a > 0 \)) in the case of exponentially ill-posed problems.

3. Proofs

For the proofs of Theorem 2.1, 2.3 and 2.5, the following proposition is central, which states that the measurement error is highly concentrated simultaneously for all \( m \) large enough. This will allow to control the measurement error in the following.

Proposition 3.1. For \( m_{opt} = m_{opt}(\delta / \rho) \) with \( m_{opt}(\delta / \rho) \to \infty \) as \( \delta / \rho \to 0 \) there holds

\[
\mathbb{P}\left( \sqrt{\frac{1}{m}} \sum_{j=1}^{m} (y^j - \hat{y}, u_j)^2 \leq \frac{\tau + 1}{2} \sqrt{m \delta}, \ \forall m \geq m_{opt} \right) \to 1
\]
as \( \delta / \rho \to 0 \).

Proof of Proposition 3.1. We have that

\[
\mathbb{P}\left( \sqrt{\frac{1}{m}} \sum_{j=1}^{m} (y^j - \hat{y}, u_j)^2 \leq \frac{\tau + 1}{2} \sqrt{m \delta}, \ \forall m \geq m_{opt} \right) = \mathbb{P}\left( \sum_{j=1}^{m} (y^j - \hat{y}, u_j)^2 \leq \frac{(\tau + 1)^2}{4} m \delta^2, \ \forall m \geq m_{opt} \right)
\]
\[
= \mathbb{P}\left( \frac{1}{m} \sum_{j=1}^{m} (y^j - \hat{y}, u_j)^2 - \delta^2 \right) \leq \left( \frac{(\tau + 1)^2}{4} - 1 \right) \delta^2, \ \forall m \geq m_{opt}
\]
\[
\geq \mathbb{P}\left( \sup_{m \geq m_{opt}} \left| \frac{1}{m} \sum_{j=1}^{m} (y^j - \hat{y}, u_j)^2 - \delta^2 \right| \leq \frac{\tau^2 - 1}{4} \delta^2 \right) = \mathbb{P}\left( \sup_{m \geq m_{opt}} \frac{1}{m} \sum_{j=1}^{m} (Z, u_j)^2 - 1 \leq \frac{\tau^2 - 1}{4} \right)
\]
\[
= \mathbb{P}\left( \sup_{m \geq m_{opt}} \left| \frac{1}{m} \sum_{j=1}^{m} X_j \right| \leq \frac{\tau^2 - 1}{4} \right).
\]
with \( X_j := (Z, u_j)^2 - 1 \). It is \((X_j)_{j \in \mathbb{N}}\) an i.i.d sequence with \( \mathbb{E}[X_j] = 0 \) and \( \mathbb{E}|X_j| \leq 2 \). Since the sample mean \((S_m)_{m \in \mathbb{N}} = \left( \frac{1}{m} \sum_{j=1}^{m} X_j \right)_{m \in \mathbb{N}} \) is a reverse martingale (16.1 in Ref. [22]), we can apply the Kolmogorov-Doob-inequality (Theorem 16.2 in Ref. [22]) and obtain
\[
\mathbb{P} \left( \sup_{m \geq m_{\text{opt}}} \left| \frac{1}{m} \sum_{j=1}^{m} X_j \right| > \frac{\tau^2 - 1}{4} \right) \leq \frac{4}{\tau^2 - 1} \mathbb{E} \left[ \frac{1}{m_{\text{opt}}} \sum_{j=1}^{m_{\text{opt}}} X_j \right] \to 0
\]
as \( \delta \to 0 \), where we have used in the last step, that \( \lim_{\delta \to \infty} m_{\text{opt}}(\delta/\rho) = \infty \) and that \( \mathbb{E}[X_j] = 0 \) and that the sample mean converges in \( L^1 \) to its expectation (Theorem 16.4 in Ref. [22]). Putting all together concludes the proof.

In the proofs of all theorems, we will split the total error into a data propagation error and an approximation error
\[
||x_{k}^{\delta} - \hat{x}|| = \left\| \sum_{j=1}^{k} \frac{(y^\delta, u_j)}{\sigma_j} v_j - \sum_{j=1}^{\infty} (\hat{x}, v_j) v_j \right\| = \left\| \sum_{j=1}^{k} \left( \frac{(y^\delta, u_j)}{\sigma_j} - (\hat{x}, v_j) \right) v_j - \sum_{j=k+1}^{\infty} (\hat{x}, v_j) v_j \right\|
\]
\[
= \left\| \sum_{j=1}^{k} \frac{(y^\delta - \hat{y}, u_j)}{\sigma_j} v_j - \sum_{j=k+1}^{\infty} (\hat{x}, v_j) v_j \right\| \leq \sqrt{\sum_{j=1}^{k} \frac{(y^\delta - \hat{y}, u_j)^2}{\sigma_j^2}} + \sqrt{\sum_{j=k+1}^{\infty} (\hat{x}, v_j)^2}
\]
and treat both terms individually.

### 3.1. Proof of Theorem 2.1

We start with an auxiliary proposition.

**Proposition 3.2.** Let \( k \in \mathbb{N} \) be such that \((\hat{y}, u_k) \neq 0\). Then there holds
\[
\mathbb{P} \left( k^{\delta}_{dp} \geq k \right) \to 1
\]
as \( \delta \to 0 \).

**Proof.** Let \( m_\delta = \lceil \delta^{-1} \rceil \). It is
\[
\sqrt{\sum_{j=k}^{m_\delta} (y^\delta, u_j)^2} \geq \sqrt{\sum_{j=k}^{m_\delta} (\hat{y}, u_j)^2} - \sqrt{\sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2}
\]
\[
\geq |(\hat{y}, u_k)| \sqrt{\sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2}.
\]
(3.2)

Moreover by Markov’s inequality
\[
\mathbb{P}\left( \sqrt{\sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2} \geq |(\hat{y}, u_k)| - \tau \sqrt{m_\delta} \delta \right) \leq \frac{\mathbb{E}\left[ \sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2 \right]}{\left( |(\hat{y}, u_k)| - \tau \sqrt{m_\delta} \delta \right)^2} \]

(3.3)

as \( \delta \to 0 \) (since \((\hat{y}, u_k) \neq 0\)). With the definition of the discrepancy principle and (3.2), (3.3) we deduce

\[
\mathbb{P}\left( k_{dp}^\delta \geq k \right) \geq \mathbb{P}\left( k_{dp}^\delta (m_\delta) \geq k \right) = \mathbb{P}\left( \sqrt{\sum_{j=k}^{m_\delta} (y^\delta, u_j)^2} > \tau \sqrt{m_\delta} \delta \right)
\]

\[
\geq \mathbb{P}\left( |(\hat{y}, u_k)| - \sqrt{\sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2} > \tau \sqrt{m_\delta} \delta \right)
\]

\[
= \mathbb{P}\left( \sum_{j=k}^{m_\delta} (y^\delta - \hat{y}, u_j)^2 < |(\hat{y}, u_k)| - \tau \sqrt{m_\delta} \delta \right) \to 1
\]

as \( \delta \to 0 \), and the assertion follows. \( \square \)

Now we set

\[
J := \sup\{ j \in \mathbb{N} : (\hat{y}, u_j) \neq 0 \}
\]

(3.4)

and distinguish the cases \( J < \infty \) and \( J = \infty \).

3.1.1. Case 1

We start with an easy corollary, which assures the existence of a deterministic lower bound, which holds with high probability.

**Corollary 3.1.** Assume that \( J = \infty \). Then there exists \( (q_\delta) \subset \mathbb{N} \) with \( q_\delta \nearrow \infty \) and

\[
\mathbb{P}\left( k_{dp}^\delta \geq q_\delta \right) \to 1
\]

as \( \delta \to 0 \).

**Proof.** This follows directly from Proposition 3.2. \( \square \)

For \( (q_\delta)_{\delta > 0} \subset \mathbb{R} \) with \( q_\delta \to \infty \) and \( \mathbb{P}(k_{dp}^\delta \geq q_\delta) \to 1 \) for \( \delta \to 0 \) we now define
\[\Omega_\delta := \left\{ \sqrt{\sum_{j=1}^{m} (\hat{y} - y^\delta, u_j)^2} \leq \frac{\tau + 1}{2} \sqrt{m \delta} \ orall m \geq \delta, 
 k^\delta_{dp} \geq q_\delta, \ |(y^\delta - \hat{y}, u_1)| \leq \delta^{-\frac{1}{2}} \right\}. \tag{3.5}\]

Note that \(\mathbb{P}(\Omega_\delta) \to 1\) because of Proposition 3.1 and
\[
\mathbb{P}\left(|(y^\delta - \hat{y}, u_1)|^2 > \sqrt{\delta}\right) \leq \frac{\mathbb{E}(y^\delta - \hat{y}, u_1)^2}{\delta} = \delta \mathbb{E}(Z, u_1) \to 0
\]
as \(\delta \to 0\). The following proposition controls the data propagation error.

**Proposition 3.3.** Assume that \(J = \infty\). Then there holds
\[
\sqrt{\sum_{j=1}^{\infty} k^\delta_{dp} (y^\delta - \hat{y}, u_j)^2} \chi_{\Omega_\delta} \leq \frac{\tau + 1}{\tau - 1} \sqrt{\sum_{j=k^\delta_{dp}}^{\infty} (\hat{x}, v_j)^2}.
\]

**Proof.** Since \(k^\delta_{dp} \chi_{\Omega_\delta} \geq q_\delta \chi_{\Omega_\delta}\) there exists (random) \(M\) with \(M \chi_{\Omega_\delta} \geq q_\delta \chi_{\Omega_\delta}\) and \(k^\delta_{dp}(M) \chi_{\Omega_\delta} = k^\delta_{dp} \chi_{\Omega_\delta}\). Since \(M = \sum_{m=q_\delta}^{\infty} m \chi_{\{M=m\}}\) it suffices to show that
\[
\sqrt{\sum_{j=1}^{\infty} k^\delta_{dp}(m) (y^\delta - \hat{y}, u_j)^2} \chi_{\Omega_\delta} \leq \frac{\tau + 1}{\tau - 1} \sqrt{\sum_{j=k^\delta_{dp}(m)}^{\infty} (\hat{x}, v_j)^2}
\]
for all (deterministic) \(m \geq q_\delta\). We can assume that \(k^\delta_{dp}(m) \geq 2\), since \(M = m\) implies that \(k^\delta_{dp}(m) \geq q_\delta \geq 2\) for \(\delta\) small enough. Then the defining relation of the discrepancy principle is not fulfilled for \(k = k^\delta_{dp}(m) - 1\). Therefore
\[
\tau \sqrt{m \delta} \chi_{\Omega_\delta} < \sqrt{\sum_{j=k^\delta_{dp}(m)}^{m} (y^\delta, u_j)^2} \chi_{\Omega_\delta} \leq \left[ \sqrt{\sum_{j=k^\delta_{dp}(m)}^{m} (y^\delta, u_j)^2} + \sqrt{\sum_{j=k^\delta_{dp}(m)}^{m} (y^\delta - \hat{y}, u_j)^2} \right] \chi_{\Omega_\delta}
\]
\[
\leq \sigma_{k^\delta_{dp}(m)} \sqrt{\sum_{j=k^\delta_{dp}(m)}^{m} (\hat{x}, v_j)^2} + \frac{\tau + 1}{2} \sqrt{m \delta}.
\]

Rearranging the expression yields
\[
\frac{\sqrt{m}}{\sigma_{k^\delta_{dp}(m)}} \delta \chi_{\Omega_\delta} \leq \frac{2}{\tau - 1} \sqrt{\sum_{j=k^\delta_{dp}(m)}^{m} (\hat{x}, v_j)^2}.
\]
Finally,
\[
\sqrt{\frac{\sum_{j=1}^{k_{dp}^\delta (m)} (y^\delta - \hat{y}, u_j)^2}{\sigma_{k_{dp}^\delta (m)}}} \chi_{\Omega_\delta} \leq \sqrt{\frac{\sum_{j=1}^{\infty} (\hat{x}, v_j)^2}{\sigma_{k_{dp}^\delta (m)}}} \chi_{\Omega_\delta} \leq \frac{\tau + 1}{2} \frac{\sqrt{m}}{\sigma_{k_{dp}^\delta (m)}} \delta \chi_{\Omega_\delta} \\
\leq \frac{\tau + 1}{\tau - 1} \sqrt{\sum_{j=k_{dp}^\delta}^{\infty} (\hat{x}, v_j)^2}.
\]

We are now ready to prove Theorem 2.1 under the assumption that \( J = \infty \). We split into a data propagation error and an approximation error
\[
||x_{k_{dp}^\delta}^\delta - \hat{x}|| \chi_{\Omega_\delta} \leq \sqrt{\sum_{j=1}^{k_{dp}^\delta} (y^\delta - \hat{y}, u_j)^2} \chi_{\Omega_\delta} + \sqrt{\sum_{j=k_{dp}^\delta+1}^{\infty} (\hat{x}, v_j)^2} \chi_{\Omega_\delta} \leq \frac{1}{\sigma_{k_{dp}^\delta}} \sqrt{\sum_{j=1}^{k_{dp}^\delta} (y^\delta - \hat{y}, u_j)^2} \chi_{\Omega_\delta} + \sqrt{\sum_{j=k_{dp}^\delta+1}^{\infty} (\hat{x}, v_j)^2} \\
\leq \frac{\tau + 1}{\tau - 1} \sqrt{\sum_{j=k_{dp}^\delta}^{\infty} (\hat{x}, v_j)^2},
\]
where we have multiple times used \( k_{dp}^\delta \chi_{\Omega_\delta} \geq q_\delta \) and Proposition 3.3. Since \( q_\delta \to \infty \) it holds that
\[
\frac{\tau + 1}{\tau - 1} \sqrt{\sum_{j=q_\delta}^{\infty} (\hat{x}, v_j)^2} \leq \frac{\epsilon}{2}
\]
for \( \delta \) small enough, so finally
\[
\mathbb{P}(||x_{k_{dp}^\delta}^\delta - \hat{x}|| \leq \epsilon) \geq \mathbb{P}(\Omega_\delta) \to 1
\]
as \( \delta \to 0 \).

3.1.2. Case 2
We now consider the case \( J < \infty \). Define
\[
q_\delta := \max \left\{ q \in \mathbb{N} : \frac{\sqrt{q}}{\sigma_q} \leq \delta^{-1} \right\}.
\]
(3.6)

Note that \( q_\delta \to \infty \) and \( \frac{\sqrt{q}}{\sigma_q} \delta \to 0 \) as \( \delta \to 0 \). Moreover, \( q_\delta \) is an upper bound for \( k_{dp}^\delta \) with high probability.

Proposition 3.4. Assume that \( J < \infty \). Then, for \( q_\delta \) given in (3.6) there holds
\[
\mathbb{P}(k_{dp}^\delta \leq q_\delta) \to 1
\]
as \( \delta \to 0 \).
Proof. For $\delta$ small enough it is $q_\delta > J$, and thus $(y^\delta, u_j) = (y^\delta - \bar{y}, u_j)$ for all $j \geq q_\delta$. Clearly, $k_{dp}(m) \leq m \leq q_\delta$ for all $m \leq q_\delta$. Therefore, for $\delta$ sufficiently small and $\tau' = 2\tau - 1$,

$$P\left(k_{dp}^\delta \leq q_\delta\right) = P\left(k_{dp}(m) \leq q_\delta, \forall m \geq q_\delta\right) = P\left(\sum_{j=q_\delta}^{m} (y^\delta - \bar{y}, u_j)^2 \leq \tau\sqrt{m\delta}, \forall m \geq q_\delta\right)$$

$$= P\left(\sum_{j=q_\delta}^{m} (y^\delta - \bar{y}, u_j)^2 \leq \tau\sqrt{m\delta}, \forall m \geq q_\delta\right)$$

$$= P\left(\sum_{j=q_\delta}^{m} (y^\delta - \bar{y}, u_j)^2 \leq \frac{\tau'+1}{2} \sqrt{m\delta}, \forall m \geq q_\delta\right) \to 1$$

as $\delta \to 0$, where we used $\delta$ sufficiently small in the third and Proposition 3.1 (with $m_{opt} = q_\delta$ and $\tau = \tau'$) in the last step. 

We come to the main proof and define

$$\Omega_\delta := \left\{ \sqrt{\sum_{j=1}^{q_\delta} (y^\delta - \bar{y}, u_j)^2} \leq \frac{\tau+1}{2} \sqrt{q_\delta}\delta, J \leq k_{dp}^\delta \leq q_\delta \right\}.$$ 

It holds that $P(\Omega_\delta) \to 1$ as $\delta \to 0$ because of Proposition 3.1, 3.2 and 3.4. We split as usual

$$\Vert x_{k_{dp}} - \hat{x} \Vert_{X_\Omega} \leq \sum_{j=1}^{k_{dp}} (y^\delta - \bar{y}, u_j)^2 \sigma_j^2 + \sum_{j=k_{dp}^\delta+1}^{\infty} (\hat{x}, v_j)^2 \sigma_j \leq \frac{1}{\sigma_{q_\delta}} \sum_{j=1}^{q_\delta} (y^\delta - \bar{y}, u_j)^2 \sigma_j + \sum_{j=1}^{\infty} (\hat{x}, v_j)^2 \sigma_j$$

$$\leq \frac{\tau+1}{2} \sqrt{q_\delta} \delta \to 0$$

as $\delta \to 0$ by definition of $J$ and $q_\delta$. This concludes the proof of Theorem 2.1.

### 3.2. Proof of Theorem 2.3

So let $\hat{x} \in X_{\nu, \rho}$. We fix an (order-) optimal a priori choice from (2.1) and set

$$m_{opt} := m_{opt}(\delta, \rho, \nu, q) = \left\lfloor \frac{\rho}{\delta} \sqrt{\tau+1} q + \tau \right\rfloor.$$ 

(3.7)

We define the event

$$\Omega_{\delta/\rho} := \left\{ \sqrt{\sum_{j=1}^{m} (y^\delta - \bar{y}, u_j)^2} \leq \frac{\tau+1}{2} \sqrt{m\delta}, \forall m \geq m_{opt} \right\},$$

(3.8)

where we have good control of the random error. It is $\lim_{\delta/\rho \to 0} P(\Omega_{\delta/\rho}) = 1$ by Proposition 3.1 above. We first show, that somewhat surprisingly, $k_{dp}^\delta(m)$
is bounded on \( \Omega_{\delta/\rho} \) by the optimal choice from (2.1) uniformly in \( m \). This implies in particular, that the same bound holds for \( k^\delta_{dp} \) from which we will later conclude that the data propagation error has the optimal order.

**Proposition 3.5.** It holds that

\[
k^\delta_{dp}(m) \mathcal{I}_{\Omega_{\delta/\rho}} \leq C_{\tau, \nu, q} m_{opt}(\delta, \nu, \rho) \quad \text{for all} \quad m \in \mathbb{N},
\]

where \( C_{\tau, \nu, q} := \left( \frac{2}{\tau - 1} \right)^{\frac{1}{2q + 1}} \) and \( m_{opt}(\delta, \nu, \rho) \) given in (3.7).

**Proof of Proposition 3.5.** By definition of \( k^\delta_{dp}(m) \), it clearly holds that \( k^\delta_{dp}(m) \leq m \), so we can assume that \( m \geq m_{opt} \). Moreover, we can assume that \( k^\delta_{dp}(m) \geq 2 \), that is the defining relation of the discrepancy principle is not fulfilled for \( k^\delta_{dp}(m) - 1 \). Thus there holds

\[
\tau \sqrt{m\delta} < \left[ \sum_{j=k^\delta_{dp}(m)}^m (\hat{y}, u_j)^2 \right]^{\frac{1}{2}} \leq \left[ \sum_{j=k^\delta_{dp}(m)}^m (\hat{y}, u_j)^2 \right]^{\frac{1}{2}} + \left[ \sum_{j=1}^m (y^\delta - \hat{y}, u_j)^2 \right]^{\frac{1}{2}}.
\]

On \( \Omega_{\delta/\rho} \) (cf (3.8)) we can further bound the right-hand side

\[
\tau \sqrt{m\delta} \mathcal{I}_{\Omega_{\delta/\rho}} \leq \left[ \sum_{j=k^\delta_{dp}(m)}^m (\hat{y}, u_j)^2 \right]^{\frac{1}{2}} + \left[ \sum_{j=1}^m (y^\delta - \hat{y}, u_j)^2 \right]^{\frac{1}{2}} \mathcal{I}_{\Omega_{\delta/\rho}} \leq \left[ \sum_{j=k^\delta_{dp}(m)}^m \sigma_j^{2(\nu + 1)}(\hat{\xi}, v_j)^2 + \frac{1}{2} \sqrt{m\delta} \right]^{\frac{1}{2}} \leq k^\delta_{dp}(m) \frac{1}{\nu + 1} \rho + \frac{1}{2} \sqrt{m\delta}.
\]

We solve for \( k^\delta_{dp}(m) \) and obtain the assertion

\[
k^\delta_{dp}(m) \mathcal{I}_{\Omega_{\delta/\rho}} \leq \left( \frac{2}{\tau - 1} \frac{\rho}{\sqrt{m\delta}} \right)^{\frac{1}{2q + 1}} = C_{\tau, \nu, q} \left( \frac{\rho}{\sqrt{m\delta}} \right)^{\frac{1}{2q + 1}} \leq C_{\tau, \nu, q} \left( \frac{\rho}{\delta} \right)^{\frac{1}{2q + 1} + 1} = C_{\tau, \nu, q} \left( \frac{\rho}{\delta} \right)^{\frac{1}{2q + 1} + 1},
\]

where we have used \( m \geq m_{opt} \) in the third step. \( \Box \)

Now we treat the approximation error. Since \( k^\delta_{dp} \geq k^\delta_{dp}(m) \) there holds that

\[
\sum_{j=k^\delta_{dp}(m)+1}^\infty (\hat{x}, v_j)^2 \leq \sum_{j=k^\delta_{dp}(m)+1}^\infty (\hat{x}, v_j)^2
\]

for all \( m \in \mathbb{N} \). Thus \( k^\delta_{dp} \) minimizes the approximation error. In order to finish the proof of Theorem 2.3 it remains to show, that there is a \( m \in \mathbb{N} \)
such that \( \sqrt{\sum_{j=k_{dp}(m)+1}^{\infty} (\hat{x}, v_j)^2} \) has the optimal rate. We show, that on \( \Omega_{\delta/\rho} \) this holds for our optimal a priori choice \( m_{opt}(\delta, \nu, \rho) \) from (3.7).

**Proposition 3.6.** There holds

\[
\left( \sum_{j=k_{dp}(m_{opt})+1}^{\infty} (\hat{x}, v_j)^2 \right)^{\frac{1}{2}} \leq C_{\tau, \nu} \rho^{\frac{(q+1)\nu}{\nu+1+q\delta}} \delta^{\frac{\nu}{\nu+1+q\delta}},
\]

with \( C_{\tau, \nu} := \left( \frac{3\tau+1}{2} \right)^{\nu+1} + 1 \) and \( m_{opt} \) from (3.7).

**Proof of Proposition 3.6.** We use \( k_{dp}(m) \leq m \) and split

\[
\left( \sum_{j=k_{dp}(m_{opt})+1}^{\infty} (\hat{x}, v_j)^2 \right)^{\frac{1}{2}} \leq \sqrt{\sum_{j=m_{opt}+1}^{m_{opt}} (\hat{x}, v_j)^2} + \sqrt{\sum_{j=m_{opt}+1}^{\infty} (\hat{x}, v_j)^2}.
\]

For the second term it holds that

\[
\left( \sum_{j=m_{opt}+1}^{\infty} (\hat{x}, v_j)^2 \right)^{\frac{1}{2}} \leq \sqrt{\sum_{j=m_{opt}+1}^{m_{opt}} \sigma_j^{2\nu}(\xi, v_j)^2} \leq \sigma_{m_{opt}}^{2\nu}(\xi, v_j)^{\frac{1}{2\nu}} \Xi_{\delta/\rho}^{\frac{1}{2\nu}}
\]

It remains to bound the first term of the right-hand side (on \( \Omega_{\delta/\rho} \)). For that we use a standard argumentation to bound the approximation error for the discrepancy principle and apply Hölder’s inequality (for \( \rho = \frac{\nu+1}{\nu} \), \( q = \nu + 1 \))

\[
\left( \sqrt{\sum_{j=k_{dp}(m_{opt})+1}^{m_{opt}} (\hat{x}, v_j)^2} \right)^{\frac{1}{2}} \leq \rho^{\frac{1}{\nu+1+q\delta}} \left( \sum_{j=k_{dp}(m_{opt})+1}^{m_{opt}} (\hat{y}, u_j)^2 \right)^{\frac{1}{\nu+1+q\delta}} \Xi_{\delta/\rho}^{\frac{1}{\nu+1+q\delta}}
\]

\[
\leq \rho^{\frac{1}{\nu+1+q\delta}} \left( \sqrt{\sum_{j=k_{dp}(m_{opt})+1}^{m_{opt}} (\hat{y}, u_j)^2} \right)^{\frac{1}{\nu+1+q\delta}} \Xi_{\delta/\rho}^{\frac{1}{\nu+1+q\delta}}
\]

\[
\leq \rho^{\frac{1}{\nu+1+q\delta}} \left( \sqrt{\sum_{j=k_{dp}(m_{opt})}^{m_{opt}} (\hat{y}, u_j)^2} + \sum_{j=k_{dp}(m_{opt})}^{m_{opt}} (\hat{y}, u_j)^2 \right)^{\frac{1}{\nu+1+q\delta}} \Xi_{\delta/\rho}^{\frac{1}{\nu+1+q\delta}}
\]

\[
\leq \rho^{\frac{1}{\nu+1+q\delta}} \left( \tau \sqrt{m_{opt} \delta + \frac{\tau+1}{2} \sqrt{m_{opt} \delta}} + \frac{3\tau+1}{2} \rho \left( \frac{1}{\nu+1+q\delta} \right)^{\frac{1}{\nu+1+q\delta}} \right)
\]

where we used the definition of \( k_{dp}(m_{opt}) \) for the first term in the fifth step. Together with the preceding bound this finishes the proof of Proposition 3.6. \( \square \)
We finish the proof of Theorem 2.3. For the decomposition (3.1) on \( \Omega_{\delta/\rho} \) it holds that

\[
\| x_{\delta y}^{\nu} - \bar{x} \|_{\Omega_{\delta/\rho}} \leq \sqrt{\sum_{j=1}^{k_{dp}} \frac{(y^j - \bar{y}, u_j)^2}{\sigma_j^2} \chi_{\Omega_{\delta/\rho}}} + \sqrt{\sum_{j=k_{dp}+1}^{\infty} (\bar{x}, v_j)^2 \chi_{\Omega_{\delta/\rho}}} \leq \frac{1}{\sigma_{k_{dp}}} \sum_{j=1}^{k_{dp}} (y^j - \bar{y}, u_j)^2 \chi_{\Omega_{\delta/\rho}} + C_{\tau, \nu} \rho^{(q+1)} \delta^{\frac{1}{q+1}} \delta^{\frac{1}{q}} \leq \left( C_{\tau, \nu, \mu} + 1 \right) \frac{1}{2} \frac{1}{\sigma_{m_{opt}}} m_{opt} \delta + C_{\tau, \nu} \rho^{(q+1)} \delta^{\frac{1}{q+1}} \delta^{\frac{1}{q}} \leq L_{\tau, \nu, \mu} \rho^{(q+1)} \delta^{\frac{1}{q+1}} \delta^{\frac{1}{q}},
\]

with \( L_{\tau, \nu, \mu} \) from Theorem 2.3, where we used Proposition 3.6 in the second, Proposition 3.5 in the third and the definition of \( \Omega_{\delta/\rho} \) (3.7) in the fourth step. Finally, Proposition 3.1 implies \( \mathbb{P}(\Omega_{\delta/\rho}) \to 1 \) as \( \delta/\rho \to 0 \) and hence finishes the proof of Theorem 2.3.

### 3.3. Proof of Theorem 2.4

We start with an auxiliary proposition. For \( a, b > 0 \) define

\[
f(x) := x^b e^{ax}.
\]

**Proposition 3.7.** For every \( y > 0 \) the equation \( f(x) = y \) has a unique solution \( x^* \) in \((0, \infty)\). Moreover, there holds

\[
x^* = \frac{1}{a} \log(y) - \frac{1}{a} \log \left( \frac{1}{a} \log(y) \right)^b + o(1)
\]

as \( y \to \infty \).

**Proof.** \( f \) is continuous on \([0, \infty)\) and strictly monotonically increasing, with \( f(0) = 0 \) and \( \lim_{x \to \infty} f(x) = \infty \), this guarantees the existence of a unique solution \( x^* = x^*(y) \). Set \( z = z(y) := \frac{1}{a} \log(y) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log(y) \right)^b \right) \). First of all,
\[ f(z) = \left( \frac{1}{a} \log (y) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log (y) \right)^b \right) \right)^b e^{a \left( \frac{1}{a} \log (y) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log (y) \right)^b \right) \right)} \]

\[ = \left( \frac{1}{a} \log (y) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log (y) \right)^b \right) \right)^b \frac{1}{\left( \frac{1}{a} \log (y) \right)^b} y \]

\[ = y \left( 1 - \frac{\log \left( \left( \frac{1}{a} \log (y) \right)^b \right)}{\log (y)} \right)^b. \]

Since \( y = f(x^*) \) and \( f \) is monotonically increasing, the above reasoning implies \( z \leq x^* \). Note that \( f'(x) = bx^{b-1}e^{ax} + ax^ke^{bx} \geq af(x) \). This together with the above calculation finally yields

\[
0 \leq x^* - z = f^{-1}(f(x^*)) - f^{-1}(f(z)) = \int_{f(z)}^{f(x^*)} \frac{1}{f'(f^{-1}(t))} dt = \int_{f(z)}^{f(x^*)} \frac{1}{af(f^{-1}(t))} dt \leq \int_{f(z)}^{f(x^*)} \frac{1}{af(f^{-1}(t))} dt \\
\leq \int_{f(z)}^{f(x^*)} \frac{1}{at} dt = \frac{1}{a} \log \left( \frac{f(x^*)}{f(z)} \right) = \frac{1}{a} \log \left( 1 - \frac{\log \left( \left( \frac{1}{a} \log (y) \right)^b \right)}{\log (y)} \right)^{-b} \to 0
\]

for \( y \to \infty \), thus \( x^* = z + o(1) \).

We start with the proof of Theorem 2.4. The usual split gives

\[
\sup_{\hat{x} \in X_{p,r}} \mathbb{E} ||x^\delta - \hat{x}||^2 = \sup_{\hat{x} \in X_{p,r}} \sum_{j=1}^{k} \frac{\mathbb{E}(y^j - \hat{y}, u_j)^2}{\sigma_j^2} + \sup_{\hat{x} \in X_{p,r}} \sum_{j=k+1}^{\infty} (\hat{x}, v_j)^2 \\
= \delta^2 \sum_{j=1}^{k} e^{\alpha j} + \sum_{\xi \in X_{p}} \sum_{j=k+1}^{\infty} (-\log (\sigma_j^2))^{-p} (\xi, v_j)^2 \\
= \delta^2 \frac{e^{\alpha(k+1)}}{e - 1} + (a(k + 1))^{-p} \rho^2.
\]

We solve the minimization problem by substituting real-valued \( x \) for \( k + 1 \) for a moment and solve by standard means (the second derivative is positive, so if suffices to set the first derivative of the right hand side to zero and solve for \( x \)). We obtain the equation

\[
a \delta^2 \frac{e^{\alpha x}}{e - 1} - pa^{-p} x^{-(p+1)} \rho^2 = 0 \quad \Rightarrow \quad x^{p+1} e^{ax} = \frac{p(e-1)}{a^{p+1}} \rho^2 \delta^2.
\]

By Proposition 3.7 (with \( b = p + 1 \) and \( y = \frac{p(e-1)\rho^2}{a^{p+1}} \)), the unique solution \( x_{opt} \) fulfills

\[
x_{opt} = \frac{1}{a} \log (y) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log (y) \right)^{p+1} \right) + o(1)
\]
for $y \to \infty$ (which corresponds to $\delta/\rho \to 0$). We handle the data propagation error (variance) and the approximation error (bias) separately. For the approximation error,

$$\rho^2(a_{x_{\text{opt}}})^{-p} = \rho^2 \left( \log(y) - \log \left( \left( \frac{1}{a} \log(y) \right)^{p+1} \right) \right)^{-p} (1 + o(1)) = \rho^2 \log(y)^{-p}(1 + o(1))$$

$$= \rho^2 \left( \log \left( \frac{p(e-1)}{a^{p+1}} \right) + \log \left( \frac{p^2}{\delta^2} \right) \right)^{-p} (1 + o(1)) = \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-p} (1 + o(1))$$

for $\frac{\delta}{\rho} \to 0$. For the data propagation error

$$\frac{\delta^2}{e-1} e^{ax_{\text{opt}}} = \frac{\delta^2}{e-1} \frac{y}{\log(y)} \left( 1 + o(1) \right) = p \rho^2 \log \left( \frac{p(e-1)}{a^{p+1}} \frac{\rho^2}{\delta^2} \right)^{(p+1)} (1 + o(1))$$

$$= p \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-(p+1)} (1 + o(1)) = \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-p} \rho \left( 1 + o(1) \right)$$

for $\frac{\delta}{\rho} \to 0$. The above rates stay the same, if we change $x_{\text{opt}}$ to any $x \in [x_{\text{opt}} - 1, x_{\text{opt}} + 1]$. Therefore,

$$\inf_{k \in \mathbb{N}} \sup_{x \in \mathcal{X}_{p,a}} \mathbb{E}[|x_k - \hat{x}|^2] = \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-p} \rho \left( 1 + o(1) \right)$$

$$+ \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-p} (1 + o(1))$$

$$= \rho^2 \left( - \log \left( \frac{\delta^2}{p^2} \right) \right)^{-p} (1 + o(1))$$

as $\frac{\delta}{\rho} \to \infty$.

### 3.4. Proof of Theorem 2.5

We argue along the lines of the proof of Theorem 2.3 and set

$$m_{\text{opt}} = m_{\text{opt}}(\delta, \rho, p, a) := \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log \left( \frac{p^2}{\delta^2} \right) \right)^{p+1} \right). \hspace{1cm} (3.10)$$

Again, the stopping index $k_{d\rho}^\delta(m)$ is in essence bounded by $m_{\text{opt}}$.

**Proposition 3.8.** It holds that

$$k_{d\rho}^\delta(m)_{\Omega_{s}/\rho} \leq \max(1, m_{\text{opt}}(\delta, \rho, p, a) + C_{a,p,\tau} + o(1)) \quad \text{for all } m \in \mathbb{N},$$

where $C_{a,p,\tau} = \frac{1}{a} \log \left( \frac{4}{(\tau-1)^2} \right) + 1$ and $m_{\text{opt}}$ given in (3.10).
Proof. Since \( k_{dp}^\delta(m) \leq m \) we can assume that \( m \geq m_{opt} \). We write \( k = k_{dp}^\delta(m) \) for short and argue as in the proof of Proposition 3.5. Assuming \( k \geq 2 \) we obtain

\[
\tau \sqrt{m \delta} \chi_{\Omega_{\delta/p}} \leq \left( \sum_{j=k}^{m} (\hat{\gamma}, u_j)^2 + \sum_{j=k}^{m} (\gamma^\delta - \hat{\gamma}, u_j)^2 \chi_{\Omega_{\delta/p}} \right) \leq \left( \sum_{j=k}^{m} \sigma_j^2 (- \log (\sigma_j^2)) - \rho (\xi, v_j)^2 + \frac{\tau + 1}{2} \sqrt{m \delta} \right) (3.11)
\]

\[
\leq (ak)^{-\frac{p}{2}} e^{-\frac{ak}{2}} + \frac{\tau + 1}{2} \sqrt{m \delta} (3.12)
\]

\[
\Rightarrow k^p e^{ak} \chi_{\Omega_{\delta/p}} \leq \frac{4}{a^p (\tau - 1)^2} \chi_{\Omega_{\delta/p}} \leq e^{C_{a,p,\tau} - 1} \rho^2 m_{opt}^2 \delta^2. (3.13)
\]

We write \( C = e^{C_{a,p,\tau} - 1} \) for short and apply Proposition 3.7. Because of the monotonicity of \( f \) there thus holds

\[
k \chi_{\Omega_{\delta/p}} \leq \frac{1}{a} \log \left( C \frac{\rho^2}{m_{opt}^2 \delta^2} \right) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log \left( C \frac{\rho^2}{m_{opt}^2 \delta^2} \right) \right)^p \right) + 1.
\]

The summand 1 is due to \( k \in \mathbb{N} \). We now use the fact that, for (positive) functions \( f, g \) with \( g = o(f) \) there holds

\[
\log (f + g) = \log (f) + \log (f + g) - \log (f) = \log (f) + \log (1 + \frac{g}{f})
\]

\[
= \log (f) + o(1),
\]

to obtain

\[
\frac{1}{a} \log \left( C \frac{\rho^2}{m_{opt}^2 \delta^2} \right) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log \left( C \frac{\rho^2}{m_{opt}^2 \delta^2} \right) \right)^p \right) + 1
\]

\[
\leq \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) + \frac{1}{a} \log (C) - \frac{1}{a} \log (m_{opt}) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) \right)^p \right) + 1
\]

\[
= \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) - \frac{1}{a} \log \left( \left( \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) \right)^p \right) + C_{a,p,\tau} + o(1)
\]

for \( \delta/\rho \to 0 \). This proves the claim. \( \square \)

As in the proof of Theorem 2.3, we will use this fact to bound the data propagation error.
Proposition 3.9. It holds that
\[
P\left( \sum_{j=1}^{k_{dp}} \frac{(y^d - \hat{y}, u_j)^2}{\sigma_j^2} \leq \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1) \right) \rightarrow 1
\]
for $\delta/\rho \to 0$.

Proof. Because of Proposition 3.8 it holds that
\[
P\left( \sum_{j=1}^{k_{dp}} \frac{(y^d - \hat{y}, u_j)^2}{\sigma_j^2} \leq \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1) \right)
\geq P\left( \sum_{j=1}^{m_{opt}+C_{a,p},+o(1)} \frac{(y^d - \hat{y}, u_j)^2}{\sigma_j^2} \leq \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1), \Omega_{\delta/\rho} \right)
\geq 1 - P\left( \sum_{j=1}^{m_{opt}+C_{a,p},+o(1)} \frac{(y^d - \hat{y}, u_j)^2}{\sigma_j^2} > \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1) \right) - P\left( \Omega_{\delta/\rho}^c \right).
\]

By Proposition 3.5 there holds $P(\Omega_{\delta/\rho}) \to 1$ as $\delta/\rho \to 0$. Now Markov’s inequality and the definition of $m_{opt}$ (3.10) imply
\[
P\left( \sum_{j=1}^{m_{opt}+C_{a,p},+o(1)} \frac{(y^d - \hat{y}, u_j)^2}{\sigma_j^2} > \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1) \right)
\leq \sum_{j=1}^{m_{opt}+C_{a,p},+o(1)} e^{ajE(Z, u_j)^2} \delta^2
\leq \frac{e^{am_{opt}+e^{a(1+C_{a,p},+o(1))}\delta^2}}{\rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1)}
= \frac{\rho^2 \left( \frac{1}{a} \log \left( \frac{\rho^2}{\delta^2} \right) \right)^{(p+1)} \delta^2}{\rho^2 \left( \log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} o(1)}
= \log \left( \frac{\rho^2}{\delta^2} \right)^{-1} a^{p+1} e^{a(1+C_{a,p},+o(1))} \rightarrow 0
\]
as $\delta/\rho \to 0$ (since the $o(1)$ in the nominator can be chosen such that it converges to 0 arbitrarily slowly). This proves the claim of the Proposition. 

Regarding the approximation error we again rely on $k_{dp}^d \geq k_{dp}^d(m)$ and show that for $m = m_{opt}$ the approximation error is asymptotically optimal.
Proposition 3.10. There holds
\[
\sqrt{\sum_{j=k}^{\infty} (\hat{x}, v_j)^2 \chi_{\Omega_{j/\rho}} \leq \rho \left(-\log \left(\frac{\delta^2}{\rho^2}\right)\right)^{-\frac{p}{2}}} (1 + o(1))
\]
as \delta/\rho \to 0, with \(m_{opt}\) given in (3.10).

Proof. Let \(\epsilon > 0\) be arbitrary. We write \(k = k^{\delta}_{\text{dp}}(m_{opt}) + 1\) and \(m = m_{opt}\). First of all, there exist \(\alpha = \alpha(\hat{x})\) and \(\beta = \beta(\hat{x})\) with \(\alpha, \beta \geq 0\) and \(\alpha + \beta \leq 1\) and
\[
\sum_{j=k}^{m} (\hat{x}, v_j)^2 \leq \alpha \rho^2 \quad \text{and} \quad \sum_{j=m+1}^{\infty} (\hat{x}, v_j)^2 \leq \beta \rho^2.
\]

We apply Proposition 2 of [3] (and use Equation (10) therein) to \(x^t = \sum_{j=k}^{m} (\hat{x}, v_j)^2\) and \(T = K\) and obtain
\[
\sum_{j=k}^{m} (\hat{x}, v_j)^2 \leq (a + \epsilon) \rho^2 \left(-\log \left(\frac{\sum_{j=k}^{m} (\hat{y}, u_j)^2}{(a + \epsilon) \rho^2}\right)\right)^{-p} (1 + o(1))
\]
for \(\delta/\rho \to 0\). By definition of \(k = k^{\delta}_{\text{dp}}(m_{opt})\) there holds
\[
\sum_{j=k}^{m} (\hat{y}, u_j)^2 \chi_{\Omega_{j/\rho}} \leq 2 \sum_{j=k}^{m} (y^\delta, u_j)^2 + 2 \sum_{j=k}^{m} (y^\delta - \hat{y}, u_j)^2 \chi_{\Omega_{j/\rho}} \leq 2 \tau^2 m \delta^2 + 2 \left(\frac{\tau + 1}{2}\right)^2 m \delta^2 \leq 4 \tau^2 m \delta^2.
\]

Therefore,
\[
\sum_{j=k}^{m} (\hat{x}, v_j)^2 \chi_{\Omega_{j/\rho}} \leq (a + \epsilon) \rho^2 \left(-\log \left(\frac{4 \tau^2 m \delta^2}{(a + \epsilon) \rho^2}\right)\right)^{-p} (1 + o(1))
\]
\[
\leq (a + \epsilon) \rho^2 \left(-\log \left(\frac{\delta^2}{(a + \epsilon) \rho^2}\right) - \log \left(\frac{4 \tau^2}{a} \left(\log \left(\frac{\rho^2}{\delta^2}\right) - \log \left(\frac{1}{a} \log \left(\frac{\rho^2}{\delta^2}\right)\right)^{p+1}\right)\right)\right) (1 + o(1))
\]
\[
\leq (a + \epsilon) \rho^2 \left(-\log \left(\frac{\delta^2}{(a + \epsilon) \rho^2}\right)\right)^{-p} (1 + o(1)).
\]

Moreover,
\[
\sum_{j=m+1}^{\infty} (\hat{x}, v_j)^2 \leq \left(-\log \left(\sigma_{m+1}^2\right)\right)^{-p} \sum_{j=m+1}^{\infty} (\xi, v_j)^2 \leq (a m)^{-p} \beta \rho^2
\]
\[
= \beta \rho^2 \left(-\log \left(\frac{\delta^2}{\rho^2}\right)\right)^{-p} (1 + o(1)).
\]
Adding the estimates we obtain
\[
\sum_{j=k}^{\infty} (\hat{x}, v_j)^2 I_{\Omega_{\delta/\rho}} = \sum_{j=k}^{m} (\hat{x}, v_j)^2 + \sum_{j=m+1}^{\infty} (\hat{x}, v_j)^2
\]
\[
\leq (\varepsilon + \varepsilon) \rho^2 \left( -\log \left( \frac{\delta^2}{(\varepsilon + \varepsilon) \rho^2} \right) \right)^{-p} \left( 1 + o(1) \right) + \beta \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} \left( 1 + o(1) \right)
\]
\[
\leq (1 + \varepsilon) \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} \left( 1 + o(1) \right)
\]
\[
+ (\varepsilon + \varepsilon) \rho^2 \left( -\log \left( \frac{\delta^2}{(\varepsilon + \varepsilon) \rho^2} \right) \right)^{-p} \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} \left( 1 + o(1) \right)
\]

We bound the last term. For \( \frac{\delta^2}{\rho^2} < \varepsilon \) there holds
\[
\sup_{x \in [0,1]} (\varepsilon + \varepsilon) \rho^2 \left( -\log \left( \frac{\delta^2}{(\varepsilon + \varepsilon) \rho^2} \right) \right)^{-p} \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p}
\]
\[
\leq (1 + \varepsilon) \rho^2 \sup_{x \in [0,1]} \left( \frac{1}{(\log (\varepsilon + \log \left( \frac{\delta^2}{\rho^2} \right))} \right)^{\frac{1}{p}} \left( \frac{1}{(\log \left( \frac{\delta^2}{\rho^2} \right))} \right)^{\frac{1}{p}}
\]
\[
\leq (1 + \varepsilon) \rho^2 \left( \frac{1}{(\log (\varepsilon + \log \left( \frac{\delta^2}{\rho^2} \right))} \right)^{\frac{1}{p}} \left( \frac{1}{(\log \left( \frac{\delta^2}{\rho^2} \right))} \right)^{\frac{1}{p}} = (1 + \varepsilon) \rho^2 x^{-p} \frac{x^p -(c + x)^p}{(c + x)^p}
\]

with \( x = \log \left( \frac{\delta^2}{\rho^2} \right) \) and \( c = \log (\varepsilon) \), where we used \( \varepsilon \leq 1 \) and \( \varepsilon > 0 \) in the first and second step respectively. Finally,
\[
\lim_{x \to \infty} \frac{x^p -(x + c)^p}{(x + c)^p} = \lim_{x \to \infty} \left( \frac{1}{1 + \frac{c}{x}} \right)^p -1 = 0
\]
and we obtain
\[
\sum_{j=k}^{\infty} (\hat{x}, v_j)^2 I_{\Omega_{\delta/\rho}} \leq (1 + \varepsilon) \rho^2 \left( -\log \left( \frac{\delta^2}{\rho^2} \right) \right)^{-p} \left( 1 + o(1) \right)
\]
for \( \delta/\rho \to 0 \). This finishes the proof, since \( \varepsilon \) was arbitrary.

Finally, combining Proposition 3.10 and 3.9 concludes the proof of Theorem 2.5.

**Acknowledgments**

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – GZ 2047/1, Projekt-ID 390685813.
References

[1] Morozov, V. A. (1968). The error principle in the solution of operational equations by the regularization method. Zh. Vychisl. Mat. Mat. Fiz. 8(2):295–309.

[2] Engl, H. W., Hanke, M., Neubauer, A. (1996). Regularization of Inverse Problems. Dordrecht: Kluwer Academic Publishers Group.

[3] Hohage, T. (2000). Regularization of exponentially ill-posed problems. Numer. Funct. Anal. Optim. 21(3-4):439–464. Taylor & Francis.

[4] Bissantz, N., Hohage, T., Munk, A., Ruymgaart, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. SIAM J. Numer. Anal. 45(6):2610–2636. DOI: 10.1137/060651884.

[5] Cavalier, L. (2011). Inverse problems in statistics. In Inverse Problems and High-Dimensional Estimation. Berlin, Heidelberg: Springer, pp. 3–96.

[6] Wahba, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. Numer. Anal. 14(4):651–667. DOI: 10.1137/0714044.

[7] Cavalier, L., Golubev, G. K., Picard, D., Tsybakov, A. B. (2002). Oracle inequalities for inverse problems. Ann. Statist. 30(3):843–874. DOI: 10.1214/aoas/1028674843.

[8] Cavalier, L., Golubev, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. Ann. Statist. 34(4):1653–1677. DOI: 10.1214/009053606000000542.

[9] Mathé, P., Pereverzev, S. (2006). Regularization of some linear ill-posed problems with discretized random noisy data. Math. Comp. 75(256):1913–1929. DOI: 10.1090/S0025-5718-06-01873-4.

[10] De Vito, E., Pereverzyev, S., Rosasco, L. (2010). Adaptive kernel methods using the balancing principle. Found. Comput. 10(4):455–479. DOI: 10.1007/s10208-010-0964-2.

[11] Pereverzyev, S., Schock, E. (2000). Morozov’s discrepancy principle for tikhonov. Numer. Funct. Anal. Optim. 21(7-8):901–916. DOI: 10.1080/01630560008816993.

[12] Mathé, P., Pereverzyev, S. V. (2006). The discretized discrepancy principle under general source conditions. J. Complex. 22(3):371–381. DOI: 10.1016/j.jco.2005.06.008.

[13] Blanchard, G., Mathé, P. (2012). Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. Inverse Prob. 28(11):115011. DOI: 10.1088/0266-5611/28/11/115011.

[14] Lu, S., Mathé, P. (2014). Discrepancy based model selection in statistical inverse problems. J. Complex. 30(3):290–308. DOI: 10.1016/j.jco.2014.02.002.

[15] Blanchard, G., Hoffmann, M., Reiβ, M. (2018). Optimal adaptation for early stopping in statistical inverse problems. SIAM/ASA J. Uncertain. Quant. 6(3):1043–1075. DOI: 10.1137/17M1154096.

[16] Blanchard, G., Hoffmann, M., Reiβ, M. (2018). Early stopping for statistical inverse problems via truncated svd estimation. Electron. J. Statist. 12(2):3204–3231. DOI: 10.1214/18-EJS1482.

[17] Stankewitz, B. (2020). Smoothed residual stopping for statistical inverse problems via truncated svd estimation. Electron. J. Statist. 14(2):3396–3428. DOI: 10.1214/20-EJS1747.

[18] Harrach, B., Jahn, T., Potthast, R. (2020). Beyond the Bakushinskii veto: regularising linear inverse problems without knowing the noise distribution. Numer. Math. 145(3):581–603. DOI: 10.1007/s00211-020-01122-2.
[19] Harrach, B., Jahn, T., Potthast, R. (2020). Regularising linear inverse problems under unknown non-gaussian white noise. arXiv preprint arXiv:2010.04519.

[20] Jahn, T. (2021). A modified discrepancy principle to attain optimal convergence rates under unknown noise. Inverse Problems 37(9):095008.

[21] Mathé, P., Pereverzev, S. V. (2017). Complexity of linear ill-posed problems in hilbert space. J. Complex. 38:50–67. DOI: 10.1016/j.jco.2016.10.003.

[22] Gut, A. (2013). Probability: A Graduate Course, Vol. 75. New York, NY: Springer Science & Business Media. https://link.springer.com/book/10.1007/978-1-4614-4708-5