Beyond the Bakushinskii veto: Regularising linear inverse problems without knowing the noise distribution

Preprint · November 2018

3 authors:


Bastian Harrach
Goethe-Universität Frankfurt am Main


Tim Jahn
Goethe-Universität Frankfurt am Main


Roland Potthast
University of Reading

Beyond the Bakushinkii veto: Regularising linear inverse problems without knowing the noise distribution

Bastian Harrach · **Tim Jahn** · **Roland Potthast**

Received: date / Accepted: date

Abstract This article deals with the solution of an ill-posed equation $K\hat{x} = \hat{y}$ for a given compact linear operator K on separable Hilbert spaces. Often, one only has a corrupted version y^δ of $\hat{y}$ at hand and the Bakushinskii veto tells us, that we are not able to solve the equation if we do not know the noise level $\|\hat{y} - y^\delta\|$. But in applications it is ad hoc unrealistic to know the error of a measurement. In practice, the error of a measurement may often be estimated through averaging of multiple measurements. In this paper, we integrate a natural approach to that in our analysis, ending up with a scheme allowing to solve the ill-posed equation without any specific assumption for the error distribution of the measurement.
More precisely, we consider noisy but multiple measurements $Y_1, ..., Y_n$ of the true value $\hat{y}$. Furthermore, assuming that the noisy measurements are unbiased and independently and identically distributed according to an unknown distribution, the natural approach would be to use $(Y_1 + .. + Y_n)/n$ as an approximation to $\hat{y}$ with the

Bastian Harrach
Institute for Mathematics
Goethe-Universität Frankfurt
Frankfurt am Main
E-mail: harrach@math.uni-frankfurt.de

Tim Jahn
Institute for Mathematics
Goethe-Universität Frankfurt
Frankfurt am Main
Tel.: +49-69-79822715
E-mail: jahn@math.uni-frankfurt.de

Roland Potthast
Data Assimilation Unit
Deutscher Wetterdienst
Offenbach am Main
E-mail: Roland.Potthast@dwd.de

estimated error $s_n/\sqrt{n}$, where s_n is an estimation of the standard deviation of one measurement. We study whether and in what sense this natural approach converges. In particular, we show that using the discrepancy principle yields, in a certain sense, optimal convergence rates.

Keywords linear inverse problems · filter based regularisation · stochastic noise · discrepancy principle · optimality

1 Introduction

The goal is to solve the ill-posed equation $K\hat{x} = \hat{y}$, where $K : \mathscr{X} \to \mathscr{Y}$ is a compact operator between seperable infinite dimensional Hilbert spaces.We do not know the right hand side $\hat{y}$ exactly, but we are given several measurements $Y_1, Y_2, ...$ of it, which are independent, identically distributed and unbiased ($\mathbb{E}Y_i = \hat{y}$) random variables. Thus we assume, that we are able to measure the right hand side multiple times, and a crucial requirement is that the solution does not change at least on small time scales. Let us stress that using multiple measurements to decrease the data error is a standard engineering practice under the name 'signal averaging', see, e.g., [30] for an introducing monograph or [21] for a survey article. Examples with low or moderate numbers of measurements (up to a hundred) can be found in [9] or [31] on image averaging or [14] on satellite radar measurements. For the recent first image of a black hole, even up to 10^9 samples were averaged, cf. [1].
The given multiple measurements naturally lead to an estimator of $\hat{y}$, namely the sample mean

$$\bar{Y}_n := \frac{\sum_{i\le n} Y_i}{n}.$$

But, in general $K^+\bar{Y}_n \not\to K^+\hat{y}$ for $n \to \infty$, because the generalised inverse (Definition 2.2 of [13]) of a compact operator is not continuous. So the inverse is replaced with a family of continuous approximations $(R_\alpha)_{\alpha>0}$, called regularisation, e.g. the Tikhonov regularisation $R_\alpha := (K^*K + \alpha Id)^{-1} K^*$, where $Id : \mathscr{X} \to \mathscr{X}$ is the identity. The regularisation parameter α has to be chosen accordingly to the data $\bar{Y}_n$ and the true data error

$$\delta_n^{true} := \|\bar{Y}_n - \hat{y}\|,$$

which is also a random variable. Since $\hat{y}$ is unknown, δ_n^{true} is also unkown and has to be guessed. Natural guesses are

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

One first natural approach is now to use a (deterministic) regularisation method together with $\bar{Y}_n$ and δ_n^{est}. We are in particular interested in the discrepancy principle [33], wich is known to provide optimal convergence rates in the classical deterministic setting. The following main result states, that in a certain sense, the natural approach converges and yields the optimal deterministic rates asymptotically.

Corollary 1 (to Theorem 3 and 4) *Assume that $K : \mathscr{X} \to \mathscr{Y}$ is a compact operator between separable Hilbert spaces and that $Y_1, Y_2, ...$ are i.i.d. $\mathscr{Y}-$valued random variables which fullfill $\mathbb{E}[Y_1] = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. Define the Tikhonov regularisation $R_\alpha := (K^*K + \alpha Id)^{-1} K^*$ (or the truncated singular value regularisation, or Landweber iteration). Determine $(\alpha_n)_n$ through the discrepancy principle using δ_n^{est} (see Algorithm 1). Then $R_{\alpha_n}\bar{Y}_n$ converges to $K^+\hat{y}$ in probability, that is*

$$\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \varepsilon\right) \to 1, \quad n \to \infty, \quad \forall \varepsilon > 0.$$

Moreover, if $K^+\hat{y} = (K^*K)^{\nu/2} w$ with $w \in \mathscr{X}$ and $\|w\| \leq \rho$ for $\rho > 0$ and $0 < \nu < \nu_0 - 1$ (where ν_0 is the qualification of the chosen method, see Assumptions 1), then for all $\varepsilon > 0$,

$$\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \rho^{\frac{1}{\nu+1}} \left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu}{\nu+1} - \varepsilon}\right) \to 1, \quad n \to \infty.$$

Moreover it is shown, that the approach in general does not yield L^2 convergence [1] for a naive use of the discrepancy principle, but it does for a priori regularisation. We also discuss quickly, how one has to estimate the error to obtain almost sure convergence.

To solve an inverse problem, as already mentioned, typically some a priori information about the noise is required. This may be, in the classical deterministic case, the knowledge of an upper bound of the noise level, or, in the stochastic case, some knowledge of the error distribution or the restriction to certain classes of distributions, for example to Gaussian distributions. Thus the main difference of this article to existing results is, that beside the assumption that the measurements can be repeated, we have no other assumptions for the error distribution. Thus the approach can be easily used by everyone, who can measure multiple times.

Stochastic (or statistical) inverse problems are an active field of research with close ties to high dimensional statistics ([18],[17],[34]). In general, there are two approaches to tackle an ill-posed problem with stochastic noise. The Bayesian setting considers the solution of the problem itself as a random quantity, on which one has some a priori knowledge (see [24]). This opposes the frequentist setting, where the inverse problem is assumed to have a deterministic, exact solution ([10],[6]). We are working in the frequentist setting, but we stay close to the classic deterministic theory of linear inverse problems ([13],[35],[36]). For stastistical inverse problems, typical methods to determine the regularisation parameter are cross validation [38], Lepski's balancing principle [32] or penalised empirical risk minimisation [11]. Modifications of the discrepancy principle were studied recently ([8],[28],[7],[29]). In [8], it was first shown how to obtain optimal convergence in L^2 under Gaussian white noise with a modified version of the discrepancy principle.

Another approach is to transfer results from the classical deterministic theory using the Ky-Fan metric, which metrises convergence in probability. In ([22],[16]) it

[1] also called convergence of the integrated mean square error or root mean square error

is shown, how to obtain convergence if one knows the Ky-Fan distance between the measurements and the true data. Aspects of the Bakushinskii veto [3] for stochastic inverse problems are discussed in ([4],[5],[39]) under assumptions for the noise distribution. In particular, [5] gives an explicit non trivial example for a convergent regularisation, without knowing the exact error level, under Gaussian white noise. We extent this to arbitrary distributions here, if one has multiple measurements.
In the articles mentioned above, the error is usually modelled as a Hilbert space process (such as white noise), which makes it impossible to determine the regularisation parameter directly through the discrepancy principle. This is in contrast to our, more classic error model, where the measurement is an element of the Hilbert space itself. Under the popular assumption that the operator K is Hilbert-Schmidt, one could in principle extend our results to a general Hilbert space process error model (considering the symmetrised equation $K^*K\hat{x} = K^*\hat{y}$ instead of $K\hat{x} = \hat{y}$, as it is done for example in [8]). But we will postpone the discussion of the white noise case to a follow up paper.
To summarise the connection to the Bakushinskii veto let us state the following. The Bakushinskii veto states that the inverse problem can only be solved with a deterministic regularisation, if the noise level of the data is known. In this article we have shown, that if one has access to multiple i.i.d. measurements of an unkown distribution, one may use as data the average together with the estimated noise level and one obtains optimal convergence in probability, as the number of measurements tends to infinity. That is one can estimate the error from the data. Finally, the measurements potentially contain more information, which is not used here. For example one could estimate the third moment also, eventually quantifying how fast the rescaled average converges to the Gaussian, or one could directly regularise the non averaged measurements.
In the following section we apply our approach to a priori regularisations and in the main part we consider the widely used discrepancy principle, which is known to work optimal in the classic deterministic theory. After that we quickly show how to choose δ_n^{est} to obtain almost sure convergence and we compare the methods numerically.

2 A priori regularisation

We have in mind, that one is able to measure the true value $\hat{y}$ multiple times and that the measurements are correct in expectation. This can be formally modelled as an independent and identically distributed sequence $Y_1, Y_2, \ldots : \Omega \to \mathscr{Y}$ of random variables with values in $\mathscr{Y}$, such that $\mathbb{E}Y_1 = \hat{y} \in \mathscr{D}(K^+)$. In order to use the central limit theorem we require that $0 < \mathbb{E}\|Y_1\|^2 < \infty$. Note that some popular statistical error models, for example Gaussian white noise, are not included, since we require the measurement to be an element of the Hilbert space.
Here we apply the above approach to a priori parameter choice strategies $\alpha(y^\delta, \delta) = \alpha(\delta)$. The deterministic theory suggests that one should choose $\delta_n^{est} = 1/\sqrt{n}$, that is not to estimate the variance. Also otherwise it would not be an a priori regularisation method anymore since the sample variance depends, of course, on the data. This choice has the advantage, that δ_n^{est} and hence $\alpha(\delta_n^{est})$ are deterministic. Since

$\mathbb{E}\delta_n^{true\,2} = \mathbb{E}\|Y_1 - \hat{y}\|^2/n$ because of the i.i.d assumption, it is natural to try to prove convergence of $\mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2$. It turns out, that the convergence proof goes through without any problems. We use the usual definition that $R_\alpha : \mathscr{Y} \to \mathscr{X}$ is called a regularisation, if R_α is a bounded linear operator for all α_0 and if $R_\alpha y \to K^+Y$ for $\alpha \to 0$ for all $y \in \mathscr{D}(K^+)$. A regularisation method is a combination of a regularisation and a parameter choice strategy $\alpha : \mathbb{R}^+ \times \mathscr{Y} \to \mathbb{R}^+$, such that $R_{\alpha(\delta,y^\delta)}y^\delta \to K^+y$ for $\delta \to 0$, for all $y \in \mathscr{D}(K^+)$ and for all $(y^\delta)_{\delta>0} \subset \mathscr{Y}$ with $\|y^\delta - y\| \le \delta$. The method is called a priori, if the parameter choice does not depend on the data, that is if $\alpha(\delta, y) = \alpha(\delta)$.

Theorem 1 (A priori regularisation) *Assume that $K : \mathscr{X} \to \mathscr{Y}$ is a compact operator between separable Hilbert spaces and that $Y_1, Y_2, \ldots$ are i.i.d. $\mathscr{Y}-$valued random variables which fullfill $\mathbb{E}[Y_1] = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1\|^2 < \infty$. Take an a priori regularisation scheme, with $\alpha(\delta) \overset{\delta\to 0}{\longrightarrow} 0$ and $\|R_{\alpha(\delta)}\|\delta \overset{\delta\to 0}{\longrightarrow} 0$. Set $\bar{Y}_n := \sum_{i\le n} Y_i/n$ and $\delta_n^{est} := n^{-1/2}$. Then $\lim_{n\to\infty} \mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2 = 0$.*

Proof Note that infact it suffice to assume K to be bounded and linear. Remember that R_α is linear. Thus $\mathbb{E}\left[R_\alpha Y_1\right] = R_\alpha \mathbb{E}\left[Y_1\right] = R_\alpha \hat{y}$ and

$$
\begin{aligned}
\mathbb{E}\|R_\alpha \bar{Y}_n - R_\alpha \hat{y}\|^2 &= \mathbb{E}\|R_\alpha \frac{\sum_{i\le n} Y_i}{n} - R_\alpha \hat{y}\|^2 = \mathbb{E}\|\frac{\sum_{i\le n} R_\alpha Y_i}{n} - R_\alpha \hat{y}\|^2 \\
&= \mathbb{E}\|\frac{\sum_{i\le n} R_\alpha Y_i - R_\alpha \hat{y}}{n}\|^2 = \frac{\sum_{i\le n} \mathbb{E}\left[\|R_\alpha Y_i - R_\alpha \hat{y}\|^2\right]}{n^2} \\
&= \frac{\mathbb{E}\|R_\alpha Y_1 - R_\alpha \hat{y}\|^2}{n}.
\end{aligned}
$$

where we used unbiasedness and independency for the fourth equality. Therefore,

$$
\begin{aligned}
\mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - R_{\alpha(\delta_n^{est})}\hat{y} + R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - R_{\alpha(\delta_n^{est})}\hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \frac{\mathbb{E}\|R_{\alpha(\delta_n^{est})}Y_1 - R_{\alpha(\delta_n^{est})}\hat{y}\|^2}{n} + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&\le \frac{\|R_{\alpha(\delta_n^{est})}\|^2}{n}\mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \|R_{\alpha(\delta_n^{est})}\|^2 \delta_n^{est\,2}\mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&\to 0 \qquad \text{for} \quad n \to \infty.
\end{aligned}
$$

□

As in the deterministic case, under additional source conditions we can prove convergence rates. We restrict to regularisations induced by a regularising filter $R_\alpha := F_\alpha\left(K^*K\right)K^*$, see the assumptions below

Assumption 1 *$(F_\alpha)_{\alpha>0}$ is a regularising filter, i.e. a family of bounded real valued functions on $(0, \|K\|^2]$ with $\lim_{\alpha\to 0} F_\alpha(\lambda) = \frac{1}{\lambda}$ and $\lambda F_\alpha(\lambda) < C_R$ for all $\alpha > 0$ and all*

$\lambda \in \left(0, \|K\|^2\right]$, where $C_R > 0$ is some constant. Moreover, it has qualification $\nu_0 > 0$, i.e. ν_0 is maximal such that for all $\nu \in (0, \nu_0]$ there exist a constant $C_\nu > 0$ with

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

Finally, there is a constant $C_F > 0$ such that $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all $0 < \lambda \leq \|K\|^2$.

Remark 1 The generating filter of the following regulariation methods fullfill the Assumption 1:

1. Tikhonov regularisation (qualification 2)
2. n-times iterated Tikhonov regularisation (qualification $2n$),
3. truncated singular value regularisation (infinite qualification),
4. Landweber iteration (infinite qualification).

Theorem 2 (order of convergence for a-prioi regularisation) *Assume that $K : \mathscr{X} \to \mathscr{Y}$ is a compact operator between separable Hilbert spaces and that $Y_1, Y_2, \ldots$ are i.i.d. $\mathscr{Y}-$valued random variables which fullfill $\mathbb{E}[Y_1] = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1\|^2 < \infty$. Let R_α be induced by a filter fullfilling Assumption 1. Set $\bar{Y}_n := \sum_{i \leq n} Y_i/n$ and $\delta_n^{est} = n^{-1/2}$. Assume that for $0 < \nu \leq \nu_0$ and $\rho > 0$ we have that $K^+\hat{y} = (K^*K)^{\nu/2} w$ for some $w \in \mathscr{X}$ with $\|w\| \leq \rho$. Then for*

$$c \left(\frac{\delta_n^{est}}{\rho} \right)^{\frac{2}{\nu+1}} \leq \alpha(\delta_n^{est}) \leq C \left(\frac{\delta_n^{est}}{\rho} \right)^{\frac{2}{\nu+1}},$$

we have that $\sqrt{\mathbb{E}\|R_{\alpha(\delta_n^{est})} \bar{Y}_n - K^+\hat{y}\|^2} \leq C' {\delta_n^{est}}^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}} = \mathscr{O}(n^{-\frac{\nu}{2(\nu+1)}})$.

Proof We proceed similiary to the proof of Theorem 1, using additionally Proposition 1 and 2 of section 4.

$$\begin{aligned}
\mathbb{E}\|R_{\alpha(\delta_n^{est})} \bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\|R_{\alpha(\delta_n^{est})} \bar{Y}_n - R_{\alpha(\delta_n^{est})} \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})} \hat{y} - K^+\hat{y}\|^2 \\
&\leq \|R_{\alpha(\delta_n^{est})}\|^2 {\delta_n^{est}}^2 \mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})} \hat{y} - K^+\hat{y}\|^2 \\
&\leq C_R C_F \mathbb{E}\|Y_1 - \hat{y}\|^2 \frac{{\delta_n^{est}}^2}{\alpha(\delta_n^{est})} + C_\nu^2 \rho^2 \alpha(\delta_n^{est})^\nu \\
&\leq \frac{C_R C_F \mathbb{E}\|Y_1 - \hat{y}\|^2}{c} {\delta_n^{est}}^{\frac{-2}{\nu+1}} \rho^{\frac{2}{\nu+1}} {\delta_n^{est}}^2 \\
&\qquad + C_\nu^2 C^\nu {\delta_n^{est}}^{\frac{2\nu}{\nu+1}} \rho^{\frac{-2\nu}{\nu+1}} \rho^2 \\
&\leq C' {\delta_n^{est}}^{\frac{2\nu}{\nu+1}} \rho^{\frac{2}{\nu+1}}.
\end{aligned}$$

□

Remark 2 In case of Theorem 1 one could alternatively argue as follows: The spaces $\mathscr{X}' := L^2(\Omega, \mathscr{X}) = \{X : \Omega \to \mathscr{X} : \mathbb{E}\|X\|^2 < \infty\}$ and $\mathscr{Y}' := L^2(\Omega, \mathscr{Y})$ are also Hilbert spaces, with scalar products $(X, \tilde{X})_{\mathscr{X}'} := \sqrt{\mathbb{E}(X, \tilde{X})_{\mathscr{X}}}$ and $(\cdot, \cdot)_{\mathscr{Y}'}$ defined similary. The compact operator $K : \mathscr{X} \to \mathscr{Y}$ induces naturally a linear operator $K' : \mathscr{X}' \to \mathscr{Y}', X \mapsto KX$. Clearly we have that $\hat{y} \in \mathscr{Y}'$, and $(\bar{Y}_n)_n$ is a sequence in $\mathscr{Y}'$ which fullfills

$$\|\bar{Y}_n - \hat{y}\|_{\mathscr{Y}'} := \sqrt{(\bar{Y}_n - \hat{y}, \bar{Y}_n - \hat{y})_{\mathscr{Y}'}} = \sqrt{\frac{\mathbb{E}\|Y_1 - \hat{y}\|^2}{n}} = \sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2}\delta_n^{est}$$

and we can use the classic deterministic results for $K' : \mathscr{X}' \to \mathscr{Y}'$ and $\bar{Y}_n$ and δ_n^{est}. The same argumentation does not work exactly for Theorem 2, since the induced operator K' is not compact.

3 The discrepancy principle

In practice the above parameter choice strategies are of limited interest, since they require the knowledge of the abstract smoothness parameters ν and ρ. The classical discrepancy principle would be to choose α_n such that

$$\|(KR_{\alpha_n} - Id)\bar{Y}_n\| \approx \delta_n^{true} = \|\bar{Y}_n - \hat{y}\|, \tag{1}$$

which is not possible, because of the unknown δ_n^{true}. So we replace it with our estimator δ_n^{est} and implement the discrepancy principle via Algorithm 1 with or without the optional emergency stop.

Algorithm 1 Discrepancy principle with estimated data error (optional: with emergency stop)

Given measurements $Y_1, ..., Y_n$;
Set $\bar{Y}_n := \sum_{i \le n} Y_i/n$ and $\delta_n^{est} = 1/\sqrt{n}$ or $\delta_n^{est} = \sqrt{\sum_{i \le n} \|Y_i - \bar{Y}_n\|^2/(n-1)}/\sqrt{n}$.
Choose a $q \in (0,1)$.
$k = 0$;
while $\|(KR_{q^k} - Id)\bar{Y}_n\| > \delta_n^{est}$ (optional: and $q^k > 1/n$) **do**
    $k = k+1$;
end while
$\alpha_n = q^k$;

Remark 3 To our knowledge, the idea of an emergency stop first appeared in [8]. It provides a deterministic lower bound for the regularisation parameter, which may avoid overfitting. We use a more naive form of an emergency stop here, which does not require the knowledge of the singular value decomposition of K. It would be interesting to see, how more sophisticated versions of the emergency stop worked here, which is not clear to us since in our general setting we cannot rely on the concentration properties of Gaussian noise.

Algorithm 1 converges under two assumptions. The first one is also necessary in the classical deterministic setting. Equation (1) has a solution for all $\delta_n^{true} > 0$ and $\hat{y} \in \mathscr{Y}$ if K is injective. Since we replace δ_n^{true} with δ_n^{est}, we have to assure that $\delta_n^{est} \neq 0$.

This may not be the case if we choose to use the sample variance, it may happen that $Y_1 = ... = Y_n$. The assumption $\mathbb{E}\|Y_1 - \hat{y}\|^2 > 0$ guarantees that this happens with probability 1 only finitely many times. Anyway, infact the distribution of Y_1 is usually absolutely continuous wich implies that $\mathbb{P}(Y_1 = ... = Y_n) = 0$ for all $n \in \mathbb{N}$.
Unlike in the previous section, here the L^2 error will not converge in general. The regularisation parameter α_n is now random, since it depends on the potentially bad random data. With a diminishing probability p we are underestimating the data error significantly, and thus the discrepancy principle gives a too small α and we still have $p\|R_\alpha\| \gg 1$ in such a case.
In the following we will need the singular value decomposition of the compact operator K: there exists a monotone sequence $\|K\| = \sigma_1 \geq \sigma_2 \geq ... > 0$ with either $\sigma_l \to 0$ for $l \to \infty$ or there exists a $N \in \mathbb{N}$ with $\sigma_l = \sigma_N$ for all $l \geq N$. Moreover there are families of orthonormal vectors $(u_l)_{l\in\mathbb{N}}$ and $(v_l)_{l\in\mathbb{N}}$ with $span(u_l : l \in \mathbb{N}) = \overline{\mathscr{R}(K)}$, $span(v_l : l \in \mathbb{N}) = \mathscr{N}(K)^\perp$ such that $Kv_l = \sigma_l v_l$ and $K^* u_l = \sigma_l v_l$.

3.1 A counter example for convergence

We now show that the naive use of the discrepancy principle, as implemented in Algorithm 1 (without emergency stop), will yield nonconvergence in L^2. To simplify calculations we pick Gaussian noise and the truncated singular value regularisation and we set $\delta_n^{est} = 1/\sqrt{n}$. We choose $\mathscr{X} := l^2(\mathbb{N})$ with the standard basis $\{u_k := (0,...,0,1,0,...)\}$ and consider the diagonal operator

$$K : l^2(\mathbb{N}) \to l^2(\mathbb{N}), \quad u_l \mapsto \left(\frac{1}{100}\right)^{\frac{l}{2}} u_l.$$

Hence the $\sigma_l = (1/100)^{\frac{l}{2}}$ are the Eigenvalues of K and

$$R_\alpha : l^2(\mathbb{N}) \to l^2(\mathbb{N}), \quad y \mapsto \sum_{l:\sigma_l^2 \geq \alpha} \sigma_l^{-1}(y, u_l) u_l.$$

We assume that the noise is distributed along $y := \sum_{l\geq 2} 1/\sqrt{l(l-1)}u_l$, so we have that $\sum_{l>n}(y, u_l)^2 = 1/n$ and thus $y \in l^2(\mathbb{N})$. That is we set $\bar{Y}_n := \sum_{i\leq n} Y_i = \sum_{i\leq n} Z_i y$, where Z_i are i.i.d. standard Gaussians. We define $\Omega_n := \{Z_i \geq 1, i = 1...n\}$, a (very unlikely) event on which we significantly underestimate the true data error. We get that $\mathbb{P}(\Omega_n) := \mathbb{P}(Z_1 \geq 1)^n \geq 1/10^n$. Moreover, by the definition of the discrepancy principle

$$
\begin{aligned}
\frac{1}{n}\chi_{\Omega_n} = \delta_n^{est\,2}\chi_{\Omega_n} &\geq \|(KR_{\alpha_n} - Id)\bar{Y}_n\|^2\chi_{\Omega_n} = |\bar{Z}_n|^2\|(KR_{\alpha_n} - Id)y\|^2\chi_{\Omega_n} \\
&\geq \|(KR_{\alpha_n} - Id)y\|^2\chi_{\Omega_n} \\
&= \sum_{l:\sigma_l^2<\alpha_n} (y,u_l)^2\chi_{\Omega_n} = \sum_{l:(1/100)^i<\alpha_n} (y,u_l)^2\chi_{\Omega_n} \\
&= \sum_{l>\frac{\log(\alpha_n)}{\log(1/100)}} (y,u_l)^2\chi_{\Omega_n} \geq \frac{\log(1/100)}{\log(\alpha_n)}\chi_{\Omega_n} \\
\Longrightarrow \alpha_n\chi_{\Omega_n} &< \frac{1}{100^n}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\|R_{\alpha_n}\bar{Y}_n\|^2 \geq \mathbb{E}\|R_{\alpha_n}\bar{Y}_n\chi_{\Omega_n}\|^2 \\
&= \bar{Z}_n^2\mathbb{E}\|R_{\alpha_n}y\chi_{\Omega_n}\|^2 \geq \mathbb{E}\|R_{1/100^n}y\chi_{\Omega_n}\|^2 \\
&\geq \sum_{l:\sigma_i^2\geq 1/100^n} \sigma_l^{-2}(y,u_l)^2\mathbb{P}(\Omega_n) \geq \frac{1}{10^n}\sum_{l\leq n}\sigma_l^{-2}(y,u_l)^2 \\
&\geq \frac{1}{10^n}100^n(y,u_n)^2 = \frac{10^n}{n(n-1)} \to \infty.
\end{aligned}
$$

That is the probability of the events Ω_n is not small enough to compensate the huge error we have on these events, so in the end $\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \to \infty$ for $n \to \infty$.

3.2 Convergence in probability of the discrepancy principle

In this section we show, that the discrepancy principle yields convergence in probability and that this convergence is asymptotically optimal in some sense. The proofs of the Theorems 3 and 4 and of Corollary 3 can be found in the next section. For convergence in probability it does not matter how large the error is on sets with diminishing probability. We again consider regularisations induced by a filter, compared to Assumptions 1 we need an additional monotinicity property.

Assumption 2 *Assume that* $(F_\alpha)_{\alpha>0}$ *fullfills Assumption 1 and that additionally it is monotone, i.e.* $F_\alpha(\lambda) \geq F_\beta(\lambda)$ *for all* $0 < \lambda \leq \|K\|^2$ *and* $\alpha \leq \beta$.

Still, the additional assumption is compatible with the prominent regularisation methods.

Remark 4 The generating filter of the following regulariation methods fullfill the Assumption 2:

1. Tikhonov regularisation,
2. generalised Tikhonov regularisation,

3. truncated singular value regularisation,
4. Landweber iteration.

Theorem 3 *Assume that K is a compact and injective operator between separable Hilbert spaces $\mathscr{X}$ and $\mathscr{Y}$ and that $Y_1, Y_2, \ldots$ are i.i.d. $\mathscr{Y}-$valued random variables with $\mathbb{E}Y_1 = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. Let R_α be induced by a filter fullfilling Assumption 2.. Applying Algorithm 1 with or without the emergency stop yields a sequence $(\alpha_n)_n$. Then we have that for all $\varepsilon > 0$*

$$\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \varepsilon\right) \stackrel{n\to\infty}{\longrightarrow} 1,$$

i.e. $R_{\alpha_n}\bar{Y}_n \stackrel{\mathbb{P}}{\longrightarrow} K^+\hat{y}$.

Remark 5 If one tried to argue as in Remark 1 to show L^2 convergence one would have to determine the regularisation parameter not as given by equation (1), but such that $\mathbb{E}\|(KR_\alpha - Id)\bar{Y}_n\|^2 \approx \delta_n^{est}$, which is not practicable since we cannot calculate the expectation on the left hand side.

The popularity of the discrepancy principles is a result of the fact that it guarantees optimal convergence rates, if the true solution fullfills some abstract smoothness conditions. More precisely, the general classic result is the following: Assuming that the true solution fullfills $\|\hat{y}\|_{\mathscr{X}_\nu} \leq \rho$, then there is a constant $C > 0$ such that

$$\sup_{y^\delta : \|y^\delta - \hat{y}\| \leq \delta} \|R_{\alpha(y^\delta,\delta)}y^\delta - K^+\hat{y}\| \leq C\rho^{\frac{1}{\nu+1}}\delta^{\frac{\nu}{\nu+1}}. \tag{2}$$

The next theorem shows the analogous result for the natural approach: A similiar bound to (2) holds with increasing probability, where $\delta^{\frac{\nu}{\nu+1}}$ is replaced with the maximum of $\delta_n^{est\,\frac{\nu}{\nu+1}}$ and $\delta_n^{true\,\frac{\nu}{\nu+1}}(\delta_n^{true}/\delta_n^{est})^{\frac{1}{\nu+1}}$. That is, with a probability tending to 1, if $\delta_n^{true} \leq \delta_n^{est}$ the determinstic bound (2) holds with δ replaced by δ_n^{est}. This is consistent, it is no problem to overestimate the true data error. On the other hand, if one underestimates the data error, that is if $\delta_n^{true} > \delta_n^{est}$, the optimal bound (2) holds only modulo a fine $(\delta_n^{true}/\delta_n^{est})^{\frac{1}{\nu+1}} \approx Z^{\frac{1}{\nu+1}}$ for a Gaussian Z. Note that the bound is optimal if $\delta_n^{est} = \delta_n^{true}$, which gives a reason to estimate the sample variance. Note that in the deterministic setting, determining the regularisation parameter with some $\delta' < \delta$ would yield non convergence in general.

Theorem 4 (Discrepancy principle) *Assume that K is a compact and injective operator between separable Hilbert spaces $\mathscr{X}$ and $\mathscr{Y}$. Moreover, $Y_1, Y_2, \ldots$ are i.i.d. $\mathscr{Y}-$valued random variables with $\mathbb{E}Y_1 = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. Let R_α be induced by a filter fullfilling Assumption 2. Moreover, assume that there is a $0 < \nu \leq \nu_0 - 1$ and a $\rho > 0$ such that $K^+\hat{y} = (K^*K)^{\nu/2}w$ for some $w \in \mathscr{X}$ with $\|w\| \leq \rho$. Applying Algorithm 1 with or without the emergency stop yields a sequence $(\alpha_n)_n$. Then there is a constant L, such that*

$$\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq L\rho^{\frac{1}{\nu+1}}\max\left\{\delta_n^{est\,\frac{\nu}{\nu+1}}, \delta_n^{true\,\frac{\nu}{\nu+1}}\left(\delta_n^{true}/\delta_n^{est}\right)^{\frac{1}{\nu+1}}\right\}\right) \stackrel{n\to\infty}{\longrightarrow} 1.$$

While we see here the similiarities to the classic deterministic case, one may wish to have a deterministic bound on $\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|$ (for n large). Because of the central limit theorem, we expect that the error behaves like $1/\sqrt{n}$. Indeed, we will see this rate asymptotically.

Corollary 2 *Under the assumptions of Theorem 4, for all $\varepsilon > 0$ it holds that*

$$\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \rho^{\frac{1}{\nu+1}}\left(\frac{1}{\sqrt{n}}\right)^{\frac{\nu}{\nu+1}-\varepsilon}\right) \stackrel{n\to\infty}{\longrightarrow} 1.$$

Proof (Corollary 2) This follows from

$$\lim_{n\to\infty}\mathbb{P}\left(\delta_n^{true} \leq n^{\varepsilon'}\delta_n^{est} \leq n^{-\frac{1}{2}+\varepsilon}\right) = 1$$

for $\varepsilon > \varepsilon' > 0$. □

The ad hoc emergency stop $\alpha_n > 1/n$, additionally assures, that the L^2 error will not explode (unlike in the counter example of the previous subsection). Under the assumption that $\mathbb{E}\|Y_1 - \hat{y}\|^4 < \infty$, one can guarantee, that the L^2 error will converge.

Corollary 3 *Under the assumptions of Theorem 3, consider the sequence α_n determined by Algorithm 1 with emergency stop. Then there is a constant C such that $\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \leq C$ for all $n \in \mathbb{N}$. If additionally $\mathbb{E}\|Y_1 - \hat{y}\|^4 < \infty$, then it holds that $\mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2 \to 0$ for $n \to \infty$.*

The speed of the convergence is usually not optimal and will depend on how fast $\mathbb{P}(\Omega_{n,c})$ converges to 1.

3.3 Almost sure convergence

The results so far delievered either convergence in probability or convergence in L^2. We give a short remark how one can obtain almost sure convergence. Roughly speaking, one has to multiply a $\sqrt{\log\log n}$ term to δ_n^{est}. This is a simple consequence of the following theorem

Theorem 5 (Law of the iterated logarithm) *Assume that $Y_1, Y_2, \ldots$ is an i.i.d sequence with values in some seperable Hilbert space $\mathscr{Y}$. Moreover, assume that $\mathbb{E}Y_1 = 0$ and $\mathbb{E}\|Y_1\|^2 < \infty$. Then we have that*

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{\|\sum_{i\leq n} Y_i\|}{\sqrt{2\mathbb{E}\|Y_1\|^2 n\log\log n}} \leq 1\right) = 1.$$

Proof This is a simple consequence of Corollary 8.8 in [27].

So if $\mathbb{E}Y_1 = \hat{y} \in \mathscr{Y}$ we have for $\delta_n^{true} = \|\bar{Y}_n - \hat{y}\|$

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{\sqrt{n}\delta_n^{true}}{\sqrt{2\mathbb{E}\|Y_1-\hat{y}\|^2 \log\log n}} \leq 1\right) = 1,$$

that is, with probability 1 it holds that $\delta_n^{true} \leq \sqrt{\frac{2\mathbb{E}\|Y_1-\hat{y}\|^2 \log\log n}{n}}$ for n large enough. Consequently, for some $\tau > 1$ the estimator should be

$$\delta_n^{est} := \tau s_n \sqrt{\frac{2\log\log n}{n}},$$

where s_n is the square root of the sample variance. Since $\mathbb{P}(\lim_{n\to\infty} s_n^2 = \mathbb{E}\|Y_1 - \hat{y}\|^2) = 1$ and $\tau > 1$ it holds that $\sqrt{\mathbb{E}\|Y_1-\hat{y}\|} \leq \tau s_n$ for n large enough with probability 1 and thus $\delta_n^{true} \leq \delta_n^{est}$ for n large enough with probability 1. In other words, there is an event $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that for any $\omega \in \Omega_0$ there is a $N(\omega) \in \mathbb{N}$ with $\delta_n^{true}(\omega) \leq \delta_n^{est}(\omega)$ for all $n \geq N(\omega)$. So we can use $\bar{Y}_n$ and δ_n^{est} together with any deterministic regularisation method to get almost sure convergence.

4 Proofs of theorem 3 and 4

The central problem is, that $\sqrt{n}\delta_n^{est}$ converges almost surely to a constant (either 1 or $\sqrt{\mathbb{E}\|Y_1-\hat{y}\|^2}$), while $\sqrt{n}\delta_n^{true}$ converges in distribution to a Gaussian random variable (wich has the same covariance structure as Y_1). That is one will inherently for arbitrarily large n from time to time underestimate the data error (related to this, see ([12],[37], where the estimator is chosen such that the error is overestimated).
In the following we will focus on the case, where α_n is determined by Algorithm 1 without the emergency stop. The emergency stop will be treated in the subsection 4.1. We have to check that the inevitable underestimating of δ_n^{true} does not yield a too small regularisation parameter. In the classic proof one uses some bound f given by the true data $\|(KR_\alpha - Id)\hat{y}\| \leq f(\alpha)$ to control the size of α. Following these arguments we get

$$\begin{aligned} &\|(KR_{\alpha_n} - Id)\bar{Y}_n\| \leq \|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| + \|(KR_{\alpha_n} - Id)\hat{y}\| \\ \Rightarrow\quad &\delta_n^{est} - \|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq f(\alpha_n). \end{aligned}$$

by the definition of the discrepancy principle. The classical worst case error bound (which is for example $(KR_\alpha - Id)\| \leq 1$ for the Tikhonov regularisation) would give $f(\alpha) \geq \delta_n^{est} - \delta_n^{true}$, that is no control of α_n if $\delta_n^{true} > \delta_n^{est}$. But in our setting, although the noise is random, it is not coming from arbitrarily bad directions - it stabilises via the central limit theorem (This can be compared to deterministic convergence results of heuristic parameter choice rules, see for example [25]). Roghly speaking, $\sqrt{n}(\bar{Y}_n - \hat{y}) \approx Z$ for a Gaussian variable Z and therefore $\sqrt{n}\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \approx \|(KR_{\alpha_n} - Id)Z\| \stackrel{n\to\infty}{\longrightarrow} 0$, since $\alpha_n \to 0$. Thus for large n we have that

$$\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \approx \sqrt{n}\|(KR_{\alpha_n} - Id)Z\| \leq c\sqrt{n} \sim c\delta_n^{est} \tag{3}$$

for any constant $c > 0$. To make this rigorous, we have to carefully decouple the envolved two limites. This can be done by a montonicity assumption on the filter, which is fullfilled by all the standard filters which are used in practice. To prove convergence we introduce events $\Omega_{n,c}$ on which the equation (3) hold. For $0 < c < 1$ we define

$$\begin{aligned}\Omega_{n,c} :=& \{\|(KR_1 - Id)\bar{Y}_n\| > \delta_n^{est}\} \cap \{\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta_n^{est}\} \\ &\cap \{\|(KR_{\alpha_n/q} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta_n^{est}\} \cap \{\lim_{k\to\infty} \alpha_k = 0\}.\end{aligned}$$

Here we need $\|(KR_1 - Id)\bar{Y}_n\| > \delta_n^{est}$ to guarantee, that Algorithm 1 terminates with a $k > 0$, which implies that $\|(KR_{\alpha_n/q} - Id)\bar{Y}_n\| > \delta_n^{est}$. We first have to treat the special case where $K^+\hat{y} = R_\alpha \hat{y}$ for α small enough and then we show that $\mathbb{P}(\Omega_{n,c}) \to 1$ for $n \to \infty$ otherwise. First note, that the monotonicity implies that

$$\|(KR_\alpha - Id)y\|^2 = \sum_l \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 (y, u_l)^2 \tag{4}$$

$$\leq \sum_l \left(F_\beta(\sigma_l^2)\sigma_l^2 - 1\right)^2 (y, u_l)^2 = \|(KR_\beta - Id)y\|^2 \tag{5}$$

for all $\alpha \leq \beta$ and $y \in \mathscr{Y}$, since

$$\frac{1}{\sigma_l^2} \geq F_\alpha(\sigma_l^2) \geq F_\beta(\sigma_l^2) \tag{6}$$

$$\Longleftrightarrow |1 - F_\alpha(\sigma_l^2)\sigma_l^2| = 1 - F_\alpha(\sigma_l^2)\sigma_l^2 \leq 1 - F_\beta(\sigma_l^2)\sigma_l^2 = |1 - F_\beta(\sigma_l^2)\sigma_l^2|. \tag{7}$$

Moreover, from (6) it follows that if there is a $a_0 > 0$ with $R_{a_0}\hat{y} = K^+\hat{y}$ then it will hold that infact $R_\alpha \hat{y} = K^+\hat{y}$ for all $\alpha \leq a_0$, since

$$\begin{aligned}R_{a_0}\hat{y} = K^+\hat{y} &\Longleftrightarrow \|R_{a_0}\hat{y} - K^+\hat{y}\| = 0 \\ &\Longleftrightarrow \sum_l \left(F_{a_0}(\sigma_l^2)\sigma_l - \frac{1}{\sigma_l}\right)^2 (\hat{y}, u_l)^2 = 0 \\ &\Longleftrightarrow F_{a_0}(\sigma_l^2) = \frac{1}{\sigma_l^2} \quad \text{for all } l \text{ with } (\hat{y}, u_l) \neq 0.\end{aligned}$$

Thus the following Lemmas 1 and 2 cover all possible cases. We begin with the special case, where the problem is infact well-posed. Here one may see already the key ideas of the proof of the main Lemma 2.

Lemma 1 *Assume that it holds that there is an a_0 such that $R_\alpha \hat{y} = K^+\hat{y}$ for all $\alpha \leq a_0$ (this may happen if $\hat{y}$ has a finite expression in terms of the $\{u_l\}_{l\in\mathbb{N}}$ and if R_α is the truncated singular value regularisation). Then, for any sequence $(x_n)_n$ converging monotonically to 0, it holds that*

$$\mathbb{P}(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \delta_n^{true}/x_n) \longrightarrow 1,$$

in particular, $R_{\alpha_n}\bar{Y}_n \xrightarrow{\mathbb{P}} K^+\hat{y}$.

Proof We need to control $(\alpha_n)_n$, so the first step is to show that $\mathbb{P}(\alpha_n \geq qx_n) \to 1$, where q is defined in Algorithm 1.
For $x_n \leq a_0$ it holds that $(KR_{x_n} - Id)\hat{y} = 0$. So, for n large enough and $m \leq n$

$$\mathbb{P}(\|(KR_{x_n} - Id)\bar{Y}_n\| > \delta_n^{est}) = \mathbb{P}(\|(KR_{x_n} - Id)(\bar{Y}_n - \hat{y})\| > \delta_n^{est}) \tag{8}$$

$$= \mathbb{P}\left(\left\|(KR_{x_n} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| > \sqrt{n}\delta_n^{est}\right) \tag{9}$$

$$\leq \mathbb{P}\left(\left\|(KR_{x_m} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| > \sqrt{n}\delta_n^{est}\right) \tag{10}$$

because of equation (4) and (5). Using the central limit theorem gives $\sum_{i\leq n}(Y_i - \hat{y})\sqrt{n} \xrightarrow{\mathrm{w}} Z$ for $n \to \infty$ for a Gaussian Z. Since $KR_{x_m} - Id : \mathscr{Y} \to \mathscr{Y}$ and $\|.\| : \mathscr{Y} \to \mathbb{R}$ are continuous it follows that $\|(KR_{x_m} - Id)\sum_{i\leq n}(Y_i - \hat{y})/\sqrt{n}\| \xrightarrow{\mathrm{w}} \|(KR_{x_m} - Id)Z\|$ for $n \to \infty$ by the continuous mapping theorem ([26], Theorem 13.25). By the (weak) law of large numbers, $\sqrt{n}\delta_n^{est} \xrightarrow{\mathbb{P}} \gamma$ for $n \to \infty$, where γ is either 1 or $\sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2}$. Slutsky's theorem ([26], Theorem 13.18) implies that

$$\left\|(KR_{x_m} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| - \sqrt{n}\delta_n^{est} \xrightarrow{\mathrm{w}} \|(KR_{x_m} - Id)Z\| - \gamma \quad \text{for } n \to \infty.$$

Finally, by Portemanteau's lemma ([26], Lemma 13.1)

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\left\|(KR_{x_m} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| > \sqrt{n}\delta_n^{est}\right) \\
\leq &\limsup_{n\to\infty} \mathbb{P}\left(\left\|(KR_{x_m} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| - \sqrt{n}\delta_n^{est} \geq 0\right) \\
\leq &\mathbb{P}\left(\|(KR_{x_m} - Id)Z\| - \gamma \geq 0\right) = \mathbb{P}\left(\|(KR_{x_m} - Id)Z\| \geq \gamma\right)
\end{aligned}$$

for all $m \in \mathbb{N}$. By pointwise convergence we have that

$$(KR_{x_m} - Id)Z \overset{m\to\infty}{\longrightarrow} 0 \text{ a.s.},$$

so in particular, for all $\varepsilon > 0$,

$$\lim_{m\to\infty} \mathbb{P}\left(\|(KR_{\alpha_{x_m}} - Id)Z\| \geq \varepsilon\right) = 0.$$

Thus finally

$$\limsup_{n\to\infty} \mathbb{P}(\|(KR_{x_n} - Id)\bar{Y}_n\| > \delta_n^{est}) \leq \mathbb{P}(\|(KR_{x_m} - Id)Z\| \geq \gamma), \ \forall\, m \in \mathbb{N} \tag{11}$$

$$\overset{m\to\infty}{\longrightarrow} 0. \tag{12}$$

Now we set

$$\Omega_n := \|\alpha_n \geq qx_n, \delta_n^{est} \leq \delta_n^{true}/x_n\}.$$

Since $\lim_{n\to\infty} x_n = 0$, clearly $\lim_{n\to\infty} \mathbb{P}(\delta_n^{est} \leq \delta_n^{true}/x_n) = 1$ and $x_n \leq \min(1, a_0)$ for n large enough. Thus assuming $\alpha_n < qx_n$ implies that α_n/q does not fullfill the discrepancy principle and therefore

$$\delta_n^{est} < \|(KR_{\alpha_n/q} - Id)\bar{Y}_n\| \leq \|(KR_{x_n} - Id)\bar{Y}_n\| = \|(KR_{x_n} - Id)(\bar{Y}_n - \hat{y})\|.$$

So equation (11) and (12) yield (for n large enough)

$$\mathbb{P}(\alpha_n/q \geq x_n) = 1 - \mathbb{P}(\alpha_n/q < x_n) \geq 1 - \mathbb{P}(\|(KR_{x_n} - Id)\bar{Y}_n\| > \delta_n^{est}) \to 1$$

for $n \to \infty$ and hence $\mathbb{P}(\Omega_n) \to 1$ for $n \to \infty$. Moreover,

$$R_\alpha \hat{y} = K^+\hat{y} \Longleftrightarrow \sum_{l\in\mathbb{N}} \left(F_\alpha(\sigma_l)^2\sigma_l - 1/\sigma_l\right)^2 (\hat{y}, u_l)^2 = 0$$

Because the above holds by assumption for all $\alpha \leq a_0$, the boundedness of F_α implies that there is a $L \in \mathbb{N}$, such that $(\hat{y}, u_l) = 0$ for all $l \geq L$. So

$$\begin{aligned}
\|R_{\alpha_n}\hat{y} - K^+\hat{y}\| &= \sqrt{\sum_{l\leq L} \left(F_{\alpha_n}(\sigma_l)^2\sigma_l - 1/\sigma_l\right)^2 (\hat{y}, u_l)^2} \\
&\leq \sqrt{1/\sigma_L \sum_{l\leq L} \left(F_\alpha(\sigma_l)^2\sigma_l^2 - 1\right)^2 (\hat{y}, u_l)^2} \\
&= 1/\sqrt{\sigma_L}\|KR_{\alpha_n}\hat{y} - \hat{y}\| \\
&\leq 1/\sqrt{\sigma_L}\left(\|(KR_{\alpha_n} - Id)\bar{Y}_n\| + \|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\|\right) \\
&\leq 1/\sqrt{\sigma_L}\left(\delta_n^{est} + C\delta_n^{true}\right).
\end{aligned}$$

We deduce that

$$\begin{aligned}
\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|\chi_{\Omega_n} &\leq \|R_{\alpha_n}(\bar{Y}_n - \hat{y})\|\chi_{\Omega_n} + \|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \\
&\leq \|R_{\alpha_n}\|\|\bar{Y}_n - \hat{y}\|\chi_{\Omega_n} + (\delta_n^{est} + C\delta_n^{true})/\sqrt{\sigma_L}\chi_{\Omega_n} \\
&\leq \sqrt{C_R C_F/\alpha_n}\delta_n^{true}\chi_{\Omega_n} + (1/x_n + C)\delta_n^{true}/\sqrt{\sigma_L} \\
&\leq \left(\sqrt{C_R C_F/qx_n} + (1/x_n + C)/\sqrt{\sigma_L}\right)\delta_n^{true} \leq \delta_n^{true}/x_n'.
\end{aligned}$$

for some monotonically to 0 converging sequence $(x_n')_{n\in\mathbb{N}}$. After redefining $x_n := x_n'$,

$$\lim_{n\to\infty} \mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \delta_n^{true}/x_n\right) \geq \lim_{n\to\infty} \mathbb{P}(\chi_{\Omega_n}) = 1.$$

□

Now we can formulate the central lemma.

Lemma 2 *Assume that we have that $R_\alpha\hat{y} \neq K^+\hat{y}$ for all $\alpha > 0$. Then it holds that $\mathbb{P}(\Omega_{n,c}) \stackrel{n\to\infty}{\longrightarrow} 1$ for all $0 < c < 1$.*

Proof As in the special case, we need a deterministic bound for α_n (with high probability) to separate the two limites. By the strong law of large numbers we have that $\mathbb{P}\left(\lim_{n\to\infty}\bar{Y}_n = \hat{y}\right) = 1$. In particular, for all $k \in \mathbb{N}$ we have that

$$\mathbb{P}\left((\bar{Y}_n,u_1)^2 \geq \frac{1}{2}(\hat{y},u_1)^2,\dots,(\bar{Y}_n,u_k)^2 \geq \frac{1}{2}(\hat{y},u_k)^2\right) \stackrel{n\to\infty}{\longrightarrow} 1. \tag{13}$$

So we define

$$N_k := \min\left\{n\in\mathbb{N} : \mathbb{P}\left((\bar{Y}_n,u_1)^2 \geq \frac{1}{2}(\hat{y},u_1)^2,\dots,(\bar{Y}_n,u_k)^2 \geq \frac{1}{2}(\hat{y},u_k)^2\right) \geq 1-1/k\right\}.$$

By equation (13) N_k is well defined with $1 = N_1 \leq N_2 \leq N_3 \leq \dots$. For all $n\in\mathbb{N}$ we set

$$K_n := \sup\{k\in\mathbb{N} : N_k \leq n\} \stackrel{n\to\infty}{\longrightarrow} \infty.$$

Thus we have that $1 \leq K_1 \leq K_2 \leq \dots$ (note that $K_n = \infty$ is possible). It holds that

$$\mathbb{P}\left((\bar{Y}_n,u_1)^2 \geq \frac{1}{2}(\hat{y},u_1)^2,\dots,(\bar{Y}_n,u_{K_n})^2 \geq \frac{1}{2}(\hat{y},u_{K_n})^2\right) \geq 1-1/K_n \stackrel{n\to\infty}{\longrightarrow} 1.$$

Moreover, again by the strong law of large numbers, we have that

$$\mathbb{P}\left(\sqrt{n}\delta_n^{est} \leq 2\gamma\right) \to 1,$$

Where γ is either 1 or $\sqrt{\mathbb{E}\|Y_1-\hat{y}\|^2}$, depending on if we use the estimated sample variance or not.

$$\Omega_{n,c}^1 := \left\{\sqrt{n}\delta_n^{est} \leq 2\gamma, (\bar{Y}_n,u_1)^2 \geq \frac{1}{2}(\hat{y},u_1)^2,\dots,(\bar{Y}_n,u_{K_n})^2 \geq \frac{1}{2}(\hat{y},u_{K_n})^2\right\}.$$

also fullfills $\lim_{n\to\infty}\mathbb{P}\left(\Omega_{n,c}^1\right) = 1$. Then,

$$\begin{aligned}
\frac{4\gamma^2}{n}\chi_{\Omega_{n,c}^1} &\geq \delta_n^{est\,2}\chi_{\Omega_{n,c}^1} \geq \|(KR_{\alpha_n} - Id)\bar{Y}_n\|^2\chi_{\Omega_{n,c}^1} &(14)\\
&= \sum_{l\geq 1}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2(\bar{Y}_n,u_l)^2\chi_{\Omega_{n,c}^1} &(15)\\
&\geq \sum_{l\leq K_n}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2(\bar{Y}_n,u_l)^2\chi_{\Omega_{n,c}^1} &(16)\\
&\geq \frac{1}{2}\sum_{l\leq K_n}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2(\hat{y},u_l)^2\chi_{\Omega_{n,c}^1} &(17)
\end{aligned}$$

We define

$$g(n) := \sup\left\{\alpha > 0 : \sum_{l\leq K_n}\left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2(\hat{y},u_l)^2 \leq 4\gamma^2/n\right\}$$

The assumption $R_\alpha\hat{y} \neq K^+\hat{y}$ implies that there are arbitrarily large $l\in\mathbb{N}$ with $(\hat{y},u_l) \neq 0$. Because F_α is bounded and $4\gamma^2/n \to 0$ for $n\to\infty$, this gives that $g(n) \searrow 0$ for $n\to\infty$. Moreover, by equations (14) to (17)

$$\mathbb{P}(\alpha_n \leq g(n)) \geq \mathbb{P}(\Omega^1_{n,c}) \to 1. \tag{18}$$

So we see, that for n large enough

$$\mathbb{P}(\|(KR_{a_0} - Id)\bar{Y}_n\| > \delta^{est}_n, \lim_{k\to\infty} \alpha_k = 0) \geq \mathbb{P}(\Omega^1_{n,c}), \tag{19}$$

since then $\alpha_n \chi_{\Omega^1_{n,c}} \leq g(n) < a_0$. We define

1. $\Omega^2_{n,c} := \{\omega : \|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta^{est}_n\}$,
2. $\Omega^3_{n,c} := \{\omega : \|(KR_{\alpha_n/q} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta^{est}_n\}$.

Now we follow the ideas of the proof of Lemma 1(equation (8) to (12)). For $m \leq n$

$$\mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta^{est}_n\right) \tag{20}$$

$$\leq \mathbb{P}\left(\|(KR_{g(m)} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\| \geq c\sqrt{n}\delta^{est}_n\right) \tag{21}$$

$$\stackrel{n\to\infty}{\longrightarrow} \mathbb{P}\left(\|(KR_{g(m)} - Id)Z\| \geq c\gamma\right) \stackrel{m\to\infty}{\longrightarrow} 1. \tag{22}$$

We conclude that

$$\mathbb{P}\left(\Omega^2_{n,c}\right) = \mathbb{P}\left(\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta^{est}_n\right) \tag{23}$$

$$\geq \mathbb{P}\left(\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta^{est}_n, \alpha_n \leq g(n)\right) \tag{24}$$

$$\geq \mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta^{est}_n, \alpha_n \leq g(n)\right) \tag{25}$$

$$= 1 - \mathbb{P}\left(\{\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta^{est}_n\} \cup \{\alpha_n > g(n)\}\right) \tag{26}$$

$$\geq 1 - \mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta^{est}_n\right) - \mathbb{P}\left(\alpha_n > g(n)\right) \tag{27}$$

$$\stackrel{n\to\infty}{\longrightarrow} 1, \tag{28}$$

where we used the monotonicity of $\mathbb{P}$ in line(24), the monotinicity of F_α in line (25), the subadditivity of $\mathbb{P}$ in line (27) and equation (18) and (20) to (22) in line (28). The same argumentation shows that $\lim_{n\to\infty} \mathbb{P}(\Omega^3_{n,c}) = 1$ too. From the inequality (19) it follows that for n large enough

$$\mathbb{P}(\Omega_{n,c}) \geq \mathbb{P}(\Omega^1_{n,c} \cap \Omega^2_{n,c} \cap \Omega^3_{n,c}).$$

The proof is concluded by (for n large enough)

$$\begin{aligned}
\mathbb{P}({\Omega_{n,c}}^C) &\leq \mathbb{P}\left(\left(\Omega^1_{n,c} \cap \Omega^2_{n,c} \cap \Omega^3_{n,c}\right)^C\right) \\
&= \mathbb{P}\left({\Omega^1_{n,c}}^C \cup {\Omega^2_{n,c}}^C \cup {\Omega^3_{n,c}}^C\right) \\
&\leq \mathbb{P}\left({\Omega^1_{n,c}}^C\right) + \mathbb{P}\left({\Omega^2_{n,c}}^C\right) + \mathbb{P}\left({\Omega^3_{n,c}}^C\right) \\
&\stackrel{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

□

Now we come to the proofs of the main theorems. We first give some well known properties of regularisations defined by filters which fullfill Assumption 1 or 2. We omit the proofs here.

Proposition 1 *Assume that $(R_\alpha)_{\alpha>0}$ is induced by a regularising filter fullfilling $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all $0<\lambda\leq\|K\|^2$. Then*

$$\|R_\alpha\| \leq \sqrt{C_R C_F}/\sqrt{\alpha}$$

Proposition 2 *Assume that $(R_\alpha)_{\alpha>0}$ is induced by a regularising filter of qualification $\nu_0>0$ and assume that $K^+\hat{y}=\hat{x}\in\mathscr{X}_\nu$ with $\|\hat{x}\|_\nu\leq\rho$ for $\nu\leq\nu_0$. Then*

$$\begin{aligned}
&\|R_\alpha\hat{y}-K^+\hat{y}\| \leq C_\nu\rho\alpha^{\nu/2}\\
&\|R_\alpha\hat{y}-K^+\hat{y}\| \leq \|KR_\alpha\hat{y}-KK^+\hat{y}\|^{\frac{\nu}{\nu+1}}C^{\frac{1}{\nu+1}}\rho^{\frac{1}{\nu+1}}
\end{aligned}$$

If $\nu_0>1$, then for all $\nu\leq\nu_0-1$

$$\|KR_\alpha\hat{y}-KK^+\hat{y}\| \leq C_{\nu+1}\rho\alpha^{\frac{\nu+1}{2}}.$$

Proof (Theorem 3) By Lemma 1, it suffices to consider the case where $R_\alpha\hat{y}\neq K^+\hat{y}$ for all $\alpha>0$. Then, by Lemma 2, $\lim_{n\to\infty}\mathbb{P}(\Omega_{n,c})=1$. So

$$\begin{aligned}
\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n-K^+\hat{y}\|\leq\varepsilon\right) \geq&\mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n-K^+\hat{y}\|\leq\varepsilon,\Omega_{n,c}\right)\\
\geq&\mathbb{P}\left(\|R_{\alpha_n}(\bar{Y}_n-\hat{y})\|+\|R_{\alpha_n}\hat{y}-K^+\hat{y}\|\leq\varepsilon,\Omega_{n,c}\right)\\
\geq&\mathbb{P}\left(\|R_{\alpha_n}\|\sqrt{n}\left\|\frac{\sum_{i\leq n}Y_i-\hat{y}}{\sqrt{n}}\right\|+\|R_{\alpha_n}\hat{y}-K^+\hat{y}\|\leq\varepsilon,\Omega_{n,c}\right).
\end{aligned}$$

From $\lim_{n\to\infty}\alpha_n\chi_{\Omega_{n,c}}=0$ for $n\to\infty$ it follows that

$$\|R_{\alpha_n}\hat{y}-K^+\hat{y}\|\chi_{\Omega_{n,c}}\to 0 \tag{29}$$

Since $\left\|\frac{\sum_{i\leq n}Y_i-\hat{y}}{\sqrt{n}}\right\| \xrightarrow{\text{w}} \|Z\|$ for $n\to\infty$, with a Gaussian Z, it suffices to show that $\|R_{\alpha_n}\|\sqrt{n}\chi_{\Omega_{n,c}}\to 0$ for $n\to\infty$, because of Slutsky's theorem ([26], Theorem 13.18). Since $\mathbb{P}(\sqrt{n}\delta_n^{est}=\gamma)=1$ with $\gamma=1$ or $\gamma=\sqrt{\mathbb{E}\|Y_1-\hat{y}\|^2}$, it suffices infact to show $\|R_{\alpha_n}\|\delta_n^{est}\chi_{\Omega_{n,c}}\to 0$. By definition of $\Omega_{n,c}$ we have that $\alpha_n\chi_{\Omega_{n,c}}<1$ for n large enough and therefore

$$\begin{aligned}
\delta_n^{est}\chi_{\Omega_{n,c}} &< \|(KR_{\alpha_n/q}-Id)\bar{Y}_n\|\chi_{\Omega_{n,c}}\\
&\leq \|(KR_{\alpha_n/q}-Id)(\bar{Y}_n-\hat{y})\|\chi_{\Omega_{n,c}}+\|(KR_{\alpha_n/q}-Id)\hat{y}\|\chi_{\Omega_{n,c}}\\
&\leq c\delta_n^{est}\chi_{\Omega_{n,c}}+\|(KR_{\alpha_n/q}-Id)\hat{y}\|\chi_{\Omega_{n,c}},\\
\Longrightarrow \delta_n^{est}\chi_{\Omega_{n,c}} &< \frac{1}{1-c}\|(KR_{\alpha_n/q}-Id)\hat{y}\|\chi_{\Omega_{n,c}}.
\end{aligned}$$

So by Proposition 1,

$$
\begin{aligned}
\|R_{\alpha_n}\|^2 \delta_n^{est\,2} \chi_{\Omega_{n,c}} &\leq \frac{C_R C_F}{\alpha_n} \frac{1}{(1-c)^2} \|(KR_{\alpha_n/q} - Id)\hat{y}\|^2 \chi_{\Omega_{n,c}} \\
&= \frac{C_R C_F}{\alpha_n (1-c)^2} \sum_l \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\hat{y}, u_l)^2 \chi_{\Omega_{n,c}} \\
&= \frac{C_R C_F}{\alpha_n (1-c)^2} \sum_l \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (K\hat{x}, u_l)^2 \chi_{\Omega_{n,c}} \\
&= \frac{C_R C_F}{q (1-c)^2} \sum_l \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q} (\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} \\
&= \frac{C_R C_F}{q (1-c)^2} \sum_l \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q} (\hat{x}, v_l)^2 \chi_{\Omega_{n,c}}
\end{aligned}
$$

Note that the qualification ν_0 of $(F_\alpha)_\alpha$ is greater than 1. That is, there exist constants C_1, C_ν, such that

$$
\begin{aligned}
\sup_{1\geq\alpha>0} \sup_{\sigma_l<\|K\|} \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 \sigma_l^2/\alpha &= \sup_{1\geq\alpha>0} \sup_{\sigma_l<\|K\|} \left(\sigma_l |F_\alpha(\sigma_l^2)\sigma_l^2 - 1|\right)^2 1/\alpha \\
&\leq \sup_{1\geq\alpha>0} \left(C_1 \alpha^{\frac{1}{2}}\right)^2 /\alpha \leq C_1^2.
\end{aligned}
$$

and

$$
\begin{aligned}
\left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 \sigma_l^2/\alpha &= \left(\sigma_l^{\nu_0} |F_\alpha(\sigma_l^2)\sigma_l^2 - 1|\right)^2 \sigma_l^{2-2\nu_0}/\alpha \\
&\leq \left(C_{\nu_0} \alpha^{\frac{\nu_0}{2}}\right)^2 \sigma_l^{2-2\nu_0}/\alpha \leq C_{\nu_0}^2 \alpha^{\nu_0 - 1} \sigma_l^{2-2\nu_0}.
\end{aligned}
$$

Let $\varepsilon > 0$ be arbitrary. There is an $N \in \mathbb{N}$ such that $\sum_{l>N} (\hat{x}, v_l)^2 < \varepsilon/2C_1^2$, so for $\alpha_n/q = \alpha$ small enough (that is n large enough)

$$
\begin{aligned}
&\sum_{l} \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q}(\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} \\
&= \sum_{l \le N} \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q}(\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} \\
&\quad + \sum_{l > N} \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q}(\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} \\
&\le \sum_{l \le N} \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2 - 1\right)^2 \frac{\sigma_l^2}{\alpha_n/q}(\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} + C_1^2 \sum_{l > N} (\hat{x}, v_l)^2 \\
&\le \sum_{l \le N} C_{v_0}^2 \sigma_l^{2-2v_0} (\alpha_n/q)^{v_0 - 1} (\hat{x}, v_l)^2 \chi_{\Omega_{n,c}} + \varepsilon/2 \\
&\le C_{v_0}^2 \sigma_N^{2-2v_0} (\alpha_n/q)^{v_0 - 1} \chi_{\Omega_{n,c}} + \varepsilon/2 \le \varepsilon \quad \text{for } n \text{ large enough.}
\end{aligned}
$$

Thus $\lim_{n\to\infty} \|R_{\alpha_n}\| \delta_n^{est} \chi_{\Omega_{n,c}} = 0$ also. Together with (29) it follows that

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \le \varepsilon\right) \\
&\ge \lim_{n\to\infty} \mathbb{P}\left(\|R_{\alpha_n}\|\sqrt{n}\left\|\frac{\sum_{i\le n} Y_i - \hat{y}}{\sqrt{n}}\right\| + \|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \le \varepsilon, \Omega_{n,c}\right) \\
&= \lim_{n\to\infty} \mathbb{P}(\Omega_{n,c}) = 1.
\end{aligned}
$$

□

Proof (Theorem 4) We split

$$
\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \le \|R_{\alpha_n}\hat{y} - K^+\hat{y}\| + \|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|.
$$

Because K is injective it holds that $KK^+\hat{y} = \hat{y}$, so by Proposition 2

$$
\begin{aligned}
\|R_{\alpha_n}\hat{y} - K^+\hat{y}\| &\le \|KR_\alpha\hat{y} - KK^+\hat{y}\|^{\frac{\nu}{\nu+1}} C^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}} = \|KR_\alpha\hat{y} - \hat{y}\|^{\frac{\nu}{\nu+1}} C^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}} \\
&\le \left(\|(KR_{\alpha_n} - Id)\bar{Y}_n\| + \|(KR_{\alpha_n} - Id)(\hat{y} - \bar{Y}_n)\|\right)^{\frac{\nu}{\nu+1}} C^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}} \\
&\le \left(\delta_n^{est} + \|(KR_{\alpha_n} - Id)(\hat{y} - \bar{Y}_n)\|\right)^{\frac{\nu}{\nu+1}} C^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}}
\end{aligned}
$$

Therefore by definition of $\Omega_{n,c}$ we get

$$
\|K^+\hat{y} - R_{\alpha_n}\hat{y}\|\chi_{\Omega_{n,c}} \le (1+c)^{\frac{\nu}{\nu+1}} C^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}} \delta_n^{est\,\frac{\nu}{\nu+1}}.
$$

Now we treat the second term. Proposition 1 yields

$$
\|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\| \le \|R_{\alpha_n}\|\|\bar{Y}_n - \hat{y}\| \le \sqrt{C_R C_F}\frac{\delta_n^{true}}{\sqrt{\alpha_n}}.
$$

By Proposition 2 we have that for $\alpha_n < 1$

$$\begin{aligned}\delta_n^{est} &\leq \|(KR_{\alpha_n/q} - Id)\bar{Y}_n\| \\ &\leq \|(KR_{\alpha_n/q} - Id)\hat{y}\| + \|(K(R_{\alpha_n/q} - Id)(\bar{Y}_n - \hat{y})\| \\ &\leq C_{\nu+1}\,(\alpha_n/q)^{(\nu+1)/2}\,\rho + \|(K(R_{\alpha_n/q} - Id)(\bar{Y}_n - \hat{y})\|.\end{aligned}$$

By the definition of $\Omega_{n,c}$,

$$\delta_n^{est}\Omega_{n,c} \leq \frac{C_{\nu+1}}{q^{\frac{\nu+1}{2}}}\alpha_n^{(\nu+1)/2}\rho + c\delta_n^{est}.$$

It follows that

$$\sqrt{\alpha_n}\chi_{\Omega_{n,c}} \geq q\left(\frac{1-c}{C_{\nu+1}\rho}\delta_n^{est}\right)^{\frac{1}{\nu+1}}\chi_{\Omega_{n,c}}. \tag{30}$$

Putting it all together

$$\begin{aligned}&\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|\chi_{\Omega_{n,c}} \\ &\leq \|R_{\alpha_n}\hat{y} - K^+\hat{y}\|\chi_{\Omega_{n,c}} + \|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|\chi_{\Omega_{n,c}} \\ &\leq (1+c)^{\frac{\nu}{\nu+1}}\,C^{\frac{1}{\nu+1}}\,\rho^{\frac{1}{\nu+1}}\,\delta_n^{est\,\frac{\nu}{\nu+1}} + \sqrt{C_RC_F}\frac{\delta_n^{true}}{\sqrt{\alpha_n}}\chi_{\Omega_{n,c}} \\ &\leq (1+c)^{\frac{\nu+1}{\nu}}\,C^{\frac{1}{\nu+1}}\,\rho^{\frac{1}{\nu+1}}\,\delta_n^{est\,\frac{\nu}{\nu+1}} + \sqrt{C_RC_F}\frac{\delta_n^{true}}{q}\left(\frac{C_{\nu+1}\rho}{(1-c)\delta_n^{est}}\right)^{\frac{1}{\nu+1}}\chi_{\Omega_{n,c}} \\ &\leq \left((1+c)^{\frac{\nu}{\nu+1}}C^{\frac{1}{\nu+1}} + \frac{\sqrt{C_RC_F}}{q}\left(\frac{C_{\nu+1}}{1-c}\right)^{\frac{1}{\nu+1}}\right) \\ &\qquad * \rho^{\frac{1}{\nu+1}}\max\left\{\delta_n^{est\,\frac{\nu}{\nu+1}}, \delta_n^{true\,\frac{\nu}{\nu+1}}\left(\delta_n^{true}/\delta_n^{est}\right)^{\frac{1}{\nu+1}}\right\}.\end{aligned}$$

□

4.1 Proofs for the emergency stop case

Again, denote by α_n the output of Algorithm 1 without the emergency stop. For the emergency stop we have to consider the behaviour of $\|R_{\max\{\alpha_n,1/n\}}\bar{Y}_n - K^+\hat{y}\|$.

4.1.1 Proof of Lemma 1

Assume that there is a $a_0 > 0$ with $R_\alpha\hat{y} = K^+\hat{y}$ for all $\alpha \leq a_0$. In the proof of Lemma 1 we showed that $\mathbb{P}(\alpha_n \geq x_n) \to 1$ for $n \to \infty$ for arbitrary sequences $(x_n)_n$ converging to 0. Since $1/n \to 0$ for $n \to \infty$, it follows that $\mathbb{P}(\max\{\alpha_n, 1/n\} = \alpha_n) \to 1$ for $n \to \infty$, thus the emergency stop does not trigger with probability converging to 1.

4.1.2 Proof of Theorem 3

In the proof of Theorem 3 we have shown that

$$\|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \Omega_{n,c} \to 1$$

for $n \to \infty$. Since $1/n \to 0$ and $\mathbb{P}(\Omega_{n,c}) \to 1$ for $n \to \infty$ (where $\Omega_{n,c}$ is defined via the sequence α_n (without emergency stop)), we deduce

$$\mathbb{P}\left(\|R_{\max\{\alpha_n,1/n\}}\hat{y} - K^+\hat{y}\| \leq \varepsilon\right) \to 1$$

for any $\varepsilon > 0$. Moreover, we have shown that

$$\delta_n^{est}/\sqrt{\alpha_n}\chi_{\Omega_{n,c}} \to 0$$

for $n \to \infty$. Since $\delta_n^{true}/\delta_n^{est}$ converges weakly to a Gaussian, it follows that

$$\mathbb{P}\left(\delta_n^{true}/\sqrt{\alpha_n} \leq \varepsilon\right) \to 1$$

for $n \to \infty$. The proof is finished by decomposing

$$\begin{aligned}
\|R_{\max\{\alpha_n,1/n\}}\bar{Y}_n - K^+\hat{y}\| &\leq \|R_{\max\{\alpha_n,1/n\}}(\bar{Y}_n - \hat{y})\| + \|R_{\max\{\alpha_n,1/n\}}\hat{y} - K^+\hat{y}\| \\
&\leq \frac{\delta_n^{true}}{\sqrt{\max\{\alpha_n, 1/n\}}} + \|R_{\max\{\alpha_n,1/n\}}\hat{y} - K^+\hat{y}\| \\
&\leq \frac{\delta_n^{true}}{\sqrt{\alpha_n}} + \|R_{\max\{\alpha_n,1/n\}}\hat{y} - K^+\hat{y}\|.
\end{aligned}$$

4.1.3 Proof of Theorem 4

From equation 30 and since $\sqrt{n}\delta_n^{est}$ converges almost surely to a positive constant, it follows that

$$\mathbb{P}\left(\sqrt{\alpha_n} \geq c' n^{-\frac{1}{2(\nu+1)}}\right) \to 1$$

for $n \to \infty$ for some positive constant c'. Thus because of $\nu > 1$ we again deduce

$$\mathbb{P}(\max\{\alpha_n, 1/n\} = \alpha_n) \to 1$$

for $n \to \infty$.

4.2 Proof of Corollary 3

Proof (Corollary 3) Denote by α_n the output of the discrepancy principle with emergency stop. First it holds that

$$\begin{aligned}\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 &\leq 2\left(\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|^2 + \mathbb{E}\|R_{\alpha_n}\hat{y} - K^+\hat{y}\|^2\right)\\ &\leq 2\left(\mathbb{E}\|R_{\alpha_n}\|^2\delta_n^{true2} + \|R_1\hat{y} - K^+\hat{y}\|^2\right)\\ &\leq 2\left(C\mathbb{E}\frac{\delta_n^{true2}}{\alpha_n} + \|R_1\hat{y} - K^+\hat{y}\|^2\right)\\ &\leq 2\left(nC\mathbb{E}\delta_n^{true2} + \|R_1\hat{y} - K^+\hat{y}\|^2\right)\\ &= 2\left(\mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_1\hat{y} - K^+\hat{y}\|^2\right) \leq C',\end{aligned}$$

where C' does not depend on n and where we used $\alpha_n \leq 1$ and the monotonicity property of the regularising filter for the second inequality and $\alpha_n > 1/n$ for the fourth inequality. Now assume that $\mathbb{E}\|Y_1\|^4 < \infty$. We will show that then $\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \to 0$ for $n \to \infty$, where α_n is determined by Algorithm 1 with emergency stop $\alpha_n > 1/n$. First

$$\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 = \mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2\chi_{\Omega_{n,c}}\right] + \mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2\chi_{\Omega_{n,c}^C}\right]$$

and because of the Theorems 3 and 4, the first term converges to 0 for $n \to \infty$. For the second term we apply Cauchy-Schwartz

$$\begin{aligned}\mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2\chi_{\Omega_{n,c}^C}\right] &\leq \sqrt{\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^4\mathbb{E}\chi^2_{\Omega_{n,c}^C}}\\ &= \sqrt{\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^4\ \mathbb{P}\left(\Omega_{n,c}^C\right)}\end{aligned}$$

and it remains to show that there is a constant A with $\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^4 \leq A$ for all n large enough.

$$\begin{aligned}&\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^4\\ \leq&\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|^4 + 2\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|^2\|R_{\alpha_n}\hat{y} - K^+\hat{y}\|^2 + \mathbb{E}\|R_{\alpha_n}\hat{y} - K^+\hat{y}\|^4\\ \leq&\mathbb{E}\|R_{\alpha_n}\|^4\delta_n^{true4} + 2\|R_1\hat{y} - K^+\hat{y}\|^2\mathbb{E}\|R_{\alpha_n}\|^2\delta_n^{true2} + \|R_1\hat{y} - K^+\hat{y}\|^4\\ \leq&B\left(\mathbb{E}\delta_n^{true4}/\alpha_n^2 + \mathbb{E}\delta_n^{true2}/\alpha_n + 1\right)\end{aligned}$$

for some constant B, where we used the monotonicity of the generating filter for the second inequality. Finally,

$$\mathbb{E}\delta_n^{true4}/\alpha_n^2 \leq n^2 \mathbb{E}\delta_n^{true4} \leq n^2 \left(\frac{1}{n^4} \sum_{i \leq n} \mathbb{E}\|Y_i - \hat{y}\|^4 + \frac{3}{n^4} \sum_{i \neq j} \mathbb{E}\|Y_i - \hat{y}\|^2 \mathbb{E}\|Y_j - \hat{y}\|^2 \right)$$
$$\leq \frac{\mathbb{E}\|Y_1 - \hat{y}\|^4}{n} + 3\left(\mathbb{E}\|Y_1 - \hat{y}\|^2\right)^2$$

and $\mathbb{E}\delta_n^{true2}/\alpha_n \leq n\mathbb{E}\delta_n^{true2} \leq \mathbb{E}\|Y_1 - \hat{y}\|^2$. □

5 Numerical demonstration

We conclude with some numerical results.

5.1 Differentiation of binary option prices

A natural example is given if the data is acquired by a Monte-Carlo simulation, here we consider an example from mathematical finance. The buyer of a binary call option receives after T days a payoff Q, if then a certain stock price S_T is higher then the strike value K. Otherwise he gets nothing. Thus the value V of the binary option depends on the expected evolution of the stock price. We denote by r the riskfree rate, for which we could have invested the buying price of the option until the expiry rate T. If we already knew today for sure, that the stock price will hit the strike (insider information), we would pay $V = e^{-rT}Q$ for the binary option (e^{-rT} is called discount factor). Otherwise, if we believed that the stock price will hit the strike with probability p, we would pay $V = e^{-rT}Qp$. In the Black Scholes model one assumes, that the relative change of the stock price in a short time intervall is normally distributed, that is

$$S_{t+\delta t} - S_t \sim \mathcal{N}(\mu\delta t, \sigma^2 \delta t).$$

Under this assumption one can show that (see [23])

$$S_T = S_0 e^{sT},$$

where S_0 is the initial stock price and $s \sim \mathcal{N}\left(\mu - \sigma^2/2, \sigma^2/T\right)$. Under this assumptions one has $V = e^{-rT}Q\Phi(d)$, with

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} d\xi, \qquad d = \frac{\log\frac{S_0}{K} + T\left(\mu - \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}}.$$

Ultimatively we are interested in the sensitivity of V with respect to the starting stock price S_0, that is $\partial V(S_0)/\partial S_0$. We formulate this as the inverse problem of differentiation. Set $\mathcal{X} = \mathcal{Y} = L^2([0,1] =$ and define

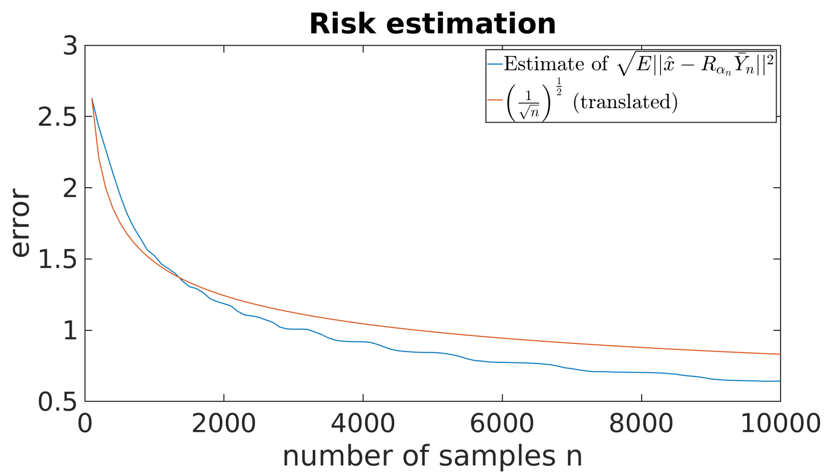

Fig. 1 Estimated Risk of a binary option.

$$
\begin{aligned}
K :& L^2([0,1]) \to L^2([0,1]) \\
& f \mapsto Af = g : x \mapsto \int_0^x f(y)dy.
\end{aligned}
$$

Then our true data is $\hat{y} = V = e^{-rT} Q\Phi(d)$. To demonstrate our results we now approximate $V : S_0 \mapsto e^{-rT} Qp(S_0)$ through a Monte-Carlo approach. That is we generate independent gaussian random variables $Z_1, Z_2, \ldots$ identically distributed to s and set $Y_i := e^{-rT} Q\chi_{\{S_0 e^{TZ_i} \geq K\}}$. Then we have $\mathbb{E}Y_i = e^{-rT} Q\mathbb{P}(S_0 e^{TZ_i}) = e^{-rT} Qp(S_0) = V(S_0)$ and $\mathbb{E}\|Y_i\|^2 \leq e^{-rT} Q < \infty$. We replace $L^2([0,1])$ with piecewise continuous linear splines on a homogeneous grid with $m = 50000$ elements (we can calculate Kg exactly for such a spline g). We use in total $n = 10000$ random variables for each simulation. As parameters we chose $r = 0.0001, T = 30, K = 0.5, Q = 1, \mu = 0.01, \sigma = 0.1$. It is easy to see that $\hat{x} = K^+\hat{y} \in \mathscr{X}_\nu$ for all $\nu > 0$ using the transformation $z(\xi) = 0,5e^{\sqrt{0,3\xi}-0,15}$. Since the qualification of the Tikhonov regularisation is 2, Corollary 2 gives an error bound which is asymptotically proportional to $(1/\sqrt{n})^{\frac{1}{2}}$. In Figure 1 we plot the L^2 average of 100 simulations of the discrepancy principle together with the (translated) optimal error bound. In this case the emergency stop did not trigger once - this is plausible, since the true solution is very smooth, which yields comparably higher values of the regularisation parameter and also, the error distribution is Gaussian and the problem is only mildly ill-posed.

Let us stress that this is only an academic example to demonstrate the possibility of using our new generic approach in the context of Monte Carlo simulations. Explicit solution formulas for standard binary options are well-known, and for more complex financial derivatives with discontinuous payoff profiles (such as autocallables or Coco-bonds) one would rather resort to stably differentiable Monte Carlo methods ([2] or [15]) or use specific regularization methods for numerical differentiation [19].

Fig. 2 Estimated L^2 error for 'heat'

	dp	dp+es	a priori
$n = 10^3$	1410	1.62	2.05
$n = 10^4$	196	1.21	1.86
$n = 10^5$	264	0.76	1.69

5.2 Inverse heat equation

We consider the toy problem 'heat' from [20]. We chose the discretisation level $m = 100$ and set $\sigma = 0.7$. Under this choice, the last seven singular values (calculated with the function 'csvd') fall below the machine precision of 10^{-16}. The discretised large systems of linear equations are solved iteratively using the conjugate gradient method ('pcg' from MATLAB) with a tolerance of 10^{-8}. As a regularisation method we chose Tikhonov regularisation and we compared the a priori choice $\alpha_n = 1/\sqrt{n}$, the discrepancy principle (dp) and the discrepancy principle with emergency stop (dp+es), as implemented in Algorithm 1 with $q = 0.7$ and estimated sample variance. The unbiased i.i.d measurements fullfill $\sqrt{\mathbb{E}\|Y_i - \hat{y}\|^2} \approx 1.16$ and $\mathbb{E}\|Y_i - \mathbb{E}Y_i\|^k = \infty$ for $k \geq 3$. Concretely, we chose $Y_i := \hat{y} + E_i$ with $E_i := U_i * Z_i * v$, where the U_i are independent and uniformly on $[-1/2, 1/2]$ distributed, the Z_i are independent Pareto distributed (MATLAB function 'gprnd' with parameters 1/3, 1/2 and 3/2), and v is a uniform permutation of $1, 1/2^{\frac{3}{4}}, ..., 1/m^{\frac{3}{4}}$. Thus we chose a rather ill-posed problem together with a heavy-tailed error distribution. We considered three different sample sizes $n = 10^3, 10^4, 10^5$ with 200 simulations for each one. The results are presented as boxplots in Figure 3. It is visible, that the results are much more concentrated for a priori regularisation and discrepancy prinicple with emergency stop, indicating the L^2 convergence (strictly speaking we do not know if the discrepancy principle with emergency stop converges in L^2, since the additional assumption of Corollary 3 is violated here). Moreover the statistics of the discrepancy principle with and without emergency stop become more similiar with increasing sample size - with the crucial difference, that the outliers (this are the red crosses above the blue box, thus the cases where the mehod performed badly) are only present in case of the discrepancy principle without emergency stop, causing non-convergence in L^2, see Figure 2. Thus here the discrepancy principle with emergency stop is superior to the discrepancy principle without emergency stop, in particular for large sample sizes. Beside that, the error is falling slower in case of the a priori parameter choice. The number of outliers falls with increasing sample size from 37 for $n = 10^3$ to 19 for $n = 10^5$, indicating the (slow) convergence in probability of the discrepancy principle. Note that for the outliers in average $\delta_n^{true}/\delta_n^{est} \approx 1.9$, while for the non outliers in average $\delta_n^{true}/\delta_n^{est} \approx 0.6$. This illustrates, that the reason for non-convergence in L^2 is the occasional underestimation of the data error.

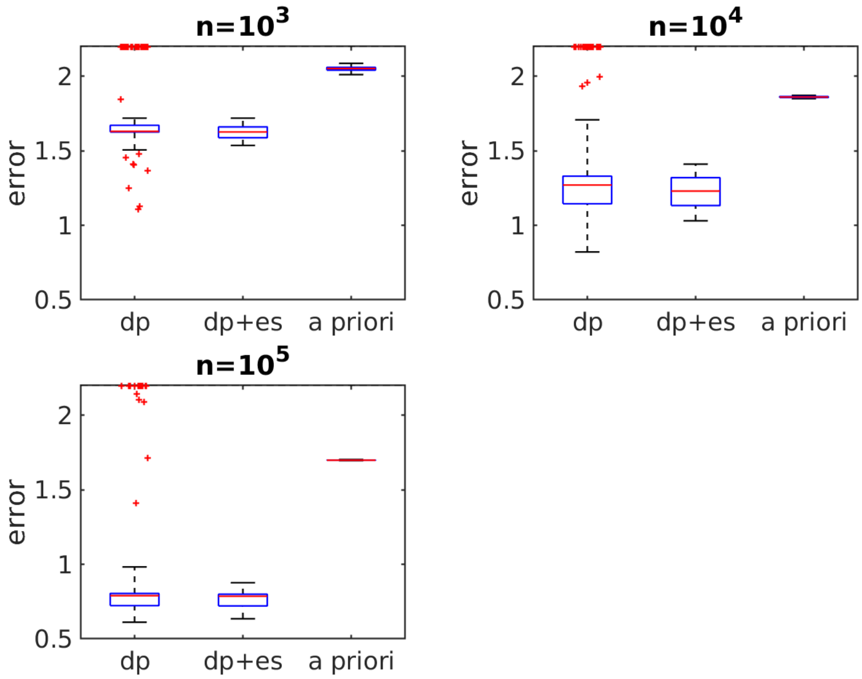

Fig. 3 Comparison of Tikhonov regularisation with discrepancy principle (dp, Algorithm 1), discrepancy principle with emergency stop (dp+es, Algorithm 1 (optional)) and a priori choice for 'heat'. Boxplots of the errors $\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|$ for 200 simulations with three different sample sizes.

References

1. Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczko, A.K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al.: First m87 event horizon telescope results. iii. data processing and calibration. The Astrophysical Journal Letters **875**(1), L3 (2019)
2. Alm, T., Harrach, B., Harrach, D., Keller, M.: A monte carlo pricing algorithm for autocallables that allows for stable differentiation. Journal of Computational Finance **17**(1) (2013)
3. Bakushinskiı, A.: Remarks on the choice of regularization parameter from quasioptimality and relation tests. Zh. Vychisl. Mat. i Mat. Fiz. **24**(8), 1258–1259 (1984)
4. Bauer, F., Reiß, M.: Regularization independent of the noise level: an analysis of quasi-optimality. Inverse Problems **24**(5), 055009 (2008)
5. Becker, S.: Regularization of statistical inverse problems and the bakushinskiı̆ veto. Inverse Problems **27**(11), 115010 (2011)
6. Bissantz, N., Hohage, T., Munk, A., Ruymgaart, F.: Convergence rates of general regularization methods for statistical inverse problems and applications. SIAM Journal on Numerical Analysis **45**(6), 2610–2636 (2007)
7. Blanchard, G., Hoffmann, M., Reiß, M.: Optimal adaptation for early stopping in statistical inverse problems. SIAM/ASA Journal on Uncertainty Quantification **6**(3), 1043–1075 (2018)
8. Blanchard, G., Mathé, P.: Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. Inverse problems **28**(11), 115011 (2012)
9. Buades, T., Lou, Y., Morel, J.M., Tang, Z.: A note on multi-image denoising. In: 2009 International Workshop on Local and Non-Local Approximation in Image Processing, pp. 1–15. IEEE (2009)
10. Cavalier, L.: Inverse problems in statistics. In: Inverse problems and high-dimensional estimation, pp. 3–96. Springer (2011)
11. Cavalier, L., Golubev, Y., et al.: Risk hull method and regularization by projections of ill-posed inverse problems. The Annals of Statistics **34**(4), 1653–1677 (2006)

12. De Vito, E., Caponnetto, A., Rosasco, L.: Model selection for regularized least-squares algorithm in learning theory. Foundations of Computational Mathematics **5**(1), 59–85 (2005)
13. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
14. Garcia, E.S., Sandwell, D.T., Smith, W.H.: Retracking cryosat-2, envisat and jason-1 radar altimetry waveforms for improved gravity field recovery. Geophysical Journal International **196**(3), 1402–1422 (2014)
15. Gerstner, T., Harrach, B., Roth, D.: Monte carlo pathwise sensitivities for barrier options. arXiv preprint arXiv:1804.03975 (2018)
16. Gerth, D., Hofinger, A., Ramlau, R.: On the lifting of deterministic convergence rates for inverse problems with stochastic noise. Inverse Problems & Imaging **11**(4), 663–687 (2017)
17. Ghosal, S., Van der Vaart, A.: Fundamentals of nonparametric Bayesian inference, vol. 44. Cambridge University Press (2017)
18. Giné, E., Nickl, R.: Mathematical foundations of infinite-dimensional statistical models, vol. 40. Cambridge University Press (2016)
19. Hanke, M., Scherzer, O.: Inverse problems light: numerical differentiation. The American Mathematical Monthly **108**(6), 512–521 (2001)
20. Hansen, P.C.: Discrete inverse problems: insight and algorithms, vol. 7. Siam (2010)
21. Hassan, U., Anwar, M.S.: Reducing noise by repetition: introduction to signal averaging. European Journal of Physics **31**(3), 453 (2010)
22. Hofinger, A.: Ill-posed problems: Extending the deterministic theory to a stochastic setup. Trauner (2006)
23. Hull, J.C., Basu, S.: Options, futures, and other derivatives. Pearson Education India (2016)
24. Kaipio, J., Somersalo, E.: Statistical and computational inverse problems, vol. 160. Springer Science & Business Media (2006)
25. Kindermann, S.: Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. Electron. Trans. Numer. Anal **38**, 233–257 (2011)
26. Klenke, A.: Probability theory: a comprehensive course. Springer Science & Business Media (2013)
27. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: isoperimetry and processes. Springer Science & Business Media (2013)
28. Lu, S., Mathé, P.: Discrepancy based model selection in statistical inverse problems. Journal of Complexity **30**(3), 290–308 (2014)
29. Lucka, F., Proksch, K., Brune, C., Bissantz, N., Burger, M., Dette, H., Wübbeling, F.: Risk estimators for choosing regularization parameters in ill-posed problems-properties and limitations. Inverse Problems & Imaging **12**(5), 1121–1155 (2018)
30. Lyons, R.G.: Understanding digital signal processing, 3/E. Pearson Education India (2004)
31. Mackay, C.D., Baldwin, J., Law, N., Warner, P.: High-resolution imaging in the visible from the ground without adaptive optics: new techniques and results. In: Ground-based Instrumentation for Astronomy, vol. 5492, pp. 128–136. International Society for Optics and Photonics (2004)
32. Mathé, P., Pereverzev, S.V.: Geometry of linear ill-posed problems in variable hilbert scales. Inverse problems **19**(3), 789 (2003)
33. Morozov, V.A.: The error principle in the solution of operational equations by the regularization method. Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki **8**(2), 295–309 (1968)
34. Nakamura, G., Potthast, R.: Inverse Modeling. 2053-2563. IOP Publishing (2015). DOI 10.1088/978-0-7503-1218-9. URL http://dx.doi.org/10.1088/978-0-7503-1218-9
35. Rieder, A.: Keine Probleme mit inversen Problemen: eine Einführung in ihre stabile Lösung. Springer-Verlag (2013)
36. Tikhonov, A., Arsenin, V.Y.: Methods for solving ill-posed problems. John Wiley and Sons, Inc (1977)
37. Vito, E.D., Rosasco, L., Caponnetto, A., Giovannini, U.D., Odone, F.: Learning from examples as an inverse problem. Journal of Machine Learning Research **6**(May), 883–904 (2005)
38. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. SIAM Journal on Numerical Analysis **14**(4), 651–667 (1977)
39. Werner, F.: Adaptivity and oracle inequalities in linear statistical inverse problems: A (numerical) survey. In: New Trends in Parameter Identification for Mathematical Models, pp. 291–316. Springer (2018)