
Beyond the Bakushinskii 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 linear ill-posed equations in Hilbert spaces. Often, one only has a corrupted measurement of the right hand side 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. 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. We integrated that in our analysis and obtained convergence to the true solution, with the only assumption that the measurements are unbiased, independent and identically distributed according to an unknown distribution.

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

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

1 Introduction

The goal is to solve the ill-posed equation $K\hat{x} = \hat{y}$, where $\hat{x} \in \mathcal{X}$ and $\hat{y} \in \mathcal{Y}$ are elements of infinite dimensional Hilbert spaces and K is either linear and bounded with non-closed range, or more specifically compact. We do not know the right hand side \hat{y} exactly, but we are given several measurements Y_1, Y_2, \dots 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., [27] for an introducing monograph or [20] for a survey article. Examples with low or moderate numbers of measurements (up to a hundred) can be found in [9] or [28] on image averaging or [13] 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

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

But, in general $K^+\bar{Y}_n \not\rightarrow K^+\hat{y}$ for $n \rightarrow \infty$, because the generalised inverse (Definition 2.2 of [12]) of K 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 $Id : \mathcal{X} \rightarrow \mathcal{X}$ 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 unknown and has to be guessed. Natural guesses are

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

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 [30], which is known to provide optimal convergence rates (for some \hat{y}) 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 $K : \mathcal{X} \rightarrow \mathcal{Y}$ is a compact operator with dense range between Hilbert spaces and that Y_1, Y_2, \dots are i.i.d. \mathcal{Y} -valued random variables which fulfill $\mathbb{E}[Y_1] = \hat{y} \in \mathcal{R}(K)$ 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*

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

Moreover, if $K^+\hat{y} = (K^*K)^{v/2}w$ with $w \in \mathcal{X}$ and $\|w\| \leq \rho$ for $\rho > 0$ and $0 < v < v_0 - 1$ (where v_0 is the qualification of the chosen method, see Assumptions 1), then for all $\varepsilon > 0$,

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

Moreover it is shown, that the approach in general does not yield L^2 convergence¹ 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. Here we present the first rigorous convergence theory for noisy measurements without any knowledge of the error distribution. 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 ([17],[16],[31]). 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 [23]). 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 ([12],[32],[33]). For statistical inverse problems, typical methods to determine the regularisation parameter are cross validation [34], Lepski's balancing principle [29] or penalised empirical risk minimisation [11]. Modifications of the discrepancy principle were studied recently ([8],[25],[7],[26]). 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 ([21],[15]) it 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],[35]) 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 extend 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), thus it is impossible to determine the regularisation parameter directly through the discrepancy principle. This is in contrast to our, more classic

¹ also called convergence of the integrated mean squared error or root mean squared error

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 show, that if one has access to multiple i.i.d. measurements of an unknown distribution, one may use as data the average together with the estimated noise level and one obtains the optimal deterministic rate with high 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 whole covariance structure of one measurement and use this to rescale the measurements and the operator, eventually increasing the relative smoothness of the data. Also 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 use the usual definition that $R_\alpha : \mathcal{Y} \rightarrow \mathcal{X}$ is called a linear regularisation, if R_α is a bounded linear operator for all $\alpha > 0$ and if $R_\alpha y \rightarrow K^+Y$ for $\alpha \rightarrow 0$ for all $y \in \mathcal{D}(K^+)$. A regularisation method is a combination of a regularisation and a parameter choice strategy $\alpha : \mathbb{R}^+ \times \mathcal{Y} \rightarrow \mathbb{R}^+$, such that $R_{\alpha(\delta, y)}y^\delta \rightarrow K^+y$ for $\delta \rightarrow 0$, for all $y \in \mathcal{D}(K^+)$ and for all $(y^\delta)_{\delta > 0} \subset \mathcal{Y}$ with $\|y^\delta - y\| \leq \delta$. The method is called a priori, if the parameter choice does not depend on the data, that is if $\alpha(\delta, y) = \alpha(\delta)$. The measurements can be formally modelled as realisations of an independent and identically distributed sequence $Y_1, Y_2, \dots : \Omega \rightarrow \mathcal{Y}$ of random variables with values in \mathcal{Y} , such that $\mathbb{E}Y_1 = \hat{y} \in \mathcal{D}(K^+)$. Moreover, we require that $0 < \mathbb{E}\|Y_1\|^2 < \infty$, that is the measurements are (almost surely) in the Hilbert space.

In the following we apply the above approach to a priori parameter choice strategies $\alpha(y^\delta, \delta) = \alpha(\delta)$. We restrict to $\delta_n^{est} = 1/\sqrt{n}$ here, that is we do not estimate the variance here (otherwise the parameter choice would depend on the data). Since then δ_n^{est} and hence $\alpha(\delta_n^{est})$ are deterministic, the situation is very easy here and the results are not surprising (see Remark 2).

Theorem 1 (Convergence of a priori regularisation) *Assume that $K : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator with non-closed range between Hilbert spaces and that Y_1, Y_2, \dots are i.i.d. \mathcal{Y} -valued random variables which fulfill $\mathbb{E}[Y_1] = \hat{y} \in \mathcal{D}(K^+)$ and $0 < \mathbb{E}\|Y_1\|^2 < \infty$. Take an a priori regularisation scheme, with $\alpha(\delta) \xrightarrow{\delta \rightarrow 0} 0$ and*

$\|R_{\alpha(\delta)}\| \xrightarrow{\delta \rightarrow 0} 0$. Set $\bar{Y}_n := \sum_{i \leq n} Y_i / n$ and $\delta_n^{est} := n^{-1/2}$. Then $\lim_{n \rightarrow \infty} \mathbb{E} \|R_{\alpha(\delta_n^{est})} \bar{Y}_n - K^+ \hat{y}\|^2 = 0$.

Proof Because of linearity, $\mathbb{E}[R_{\alpha} Y_1] = R_{\alpha} \mathbb{E}[Y_1] = R_{\alpha} \hat{y}$ and thus by (3)

$$\mathbb{E} \|R_{\alpha} \bar{Y}_n - R_{\alpha} \hat{y}\|^2 = \frac{1}{n^2} \mathbb{E} \left\| \sum_{i=1}^n R_{\alpha} (Y_i - \hat{y}) \right\|^2 = \frac{\mathbb{E} \|R_{\alpha} Y_1 - R_{\alpha} \hat{y}\|^2}{n},$$

since $R_{\alpha} Y_i \in \mathcal{R}(K^*)$ where the latter is separable. Therefore, by the bias-variance-decomposition,

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

□

As in the deterministic case, under additional source conditions we can prove convergence rates. We restrict to regularisations $R_{\alpha} := F_{\alpha} (K^* K) K^*$ defined via the spectral decomposition (see [12]) with the following assumptions for the generating filter.

Assumption 1 $(F_{\alpha})_{\alpha > 0}$ is a regularising filter, i.e. a family of bounded real valued functions on $(0, \|K\|^2]$ with $\lim_{\alpha \rightarrow 0} F_{\alpha}(\lambda) = \frac{1}{\lambda}$ and $\lambda F_{\alpha}(\lambda) \leq C_R$ for all $\alpha > 0$ and all $\lambda \in (0, \|K\|^2]$, 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 regularisation 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 (Rate of convergence of apriori regularisation) Assume that $K : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator with non-closed range between Hilbert spaces and that Y_1, Y_2, \dots are i.i.d. \mathcal{Y} -valued random variables which fulfill $\mathbb{E}[Y_1] = \hat{y} \in \mathcal{D}(K^+)$ and $0 < \mathbb{E}\|Y_1\|^2 < \infty$. Let R_α be induced by a filter fulfilling Assumption 1. Set $\bar{Y}_n := \sum_{i \leq n} Y_i/n$ and $\delta_n^{est} = n^{-1/2}$. Assume that for $0 < v \leq v_0$ and $\rho > 0$ we have that $K^+\hat{y} = (K^*K)^{v/2}w$ for some $w \in \mathcal{X}$ with $\|w\| \leq \rho$. Then if for constants $0 < c < C$,

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

we have that $\sqrt{\mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2} \leq C' \delta_n^{est \frac{v}{v+1}} \rho^{\frac{1}{v+1}} = \mathcal{O}(n^{-\frac{v}{2(v+1)}})$ for some constant $C' > 0$.

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

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

□

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

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

and we can use the classic deterministic results for $K' : \mathcal{X}' \rightarrow \mathcal{Y}'$ and \bar{Y}_n and δ_n^{est} .

3 The discrepancy principle

In this section we restrict to compact operators with dense range. Note that then $\mathcal{Y} = \mathcal{R}(K)$ will be automatically separable. In practice the above parameter choice

strategies are of limited interest, since they require the knowledge of the abstract smoothness parameters \mathbf{v} and ρ . The classical discrepancy principle would be to choose α_n such that

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

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)

```

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

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 an elementary 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 will terminate, if we use the emergency stop. Otherwise, we can guarantee that Algorithm 1 terminates, if K has dense image (or equivalently, if K^* is injective) and if $\delta_n^{est} > 0$. This is because then $\lim_{\alpha \rightarrow 0} KR_\alpha = P_{\overline{\mathcal{R}(K)}} = Id$ pointwise, so $\|(KR_{q^k} - Id)\bar{Y}_n\| < \delta_n^{est}$ for k large enough. If we decided to use the sample variance, it may happen that $\delta_n^{est} = 0$. But assuming $\mathbb{E}\|Y_1 - \hat{y}\|^2 > 0$, it follows that $\mathbb{P}(\delta_n^{est} = 0) = \mathbb{P}(Y_1 = \dots = Y_n) \rightarrow 0$ for $n \rightarrow \infty$ (with exponential rate). If the distribution of Y_1 possesses a density (with respect to the Gaussian measure for example), then actually $\mathbb{P}(Y_1 = \dots = Y_n) = 0$ for all $n \in \mathbb{N}$.

Unlike in the previous section, here the L^2 error will not converge in general, even if Y_1 has a density. 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 with dense range (see [10]): there exists a monotone sequence $\|K\| = \sigma_1 \geq$

$\sigma_2 \geq \dots > 0$ with $\sigma_l \rightarrow 0$ for $l \rightarrow \infty$. Moreover there are families of orthonormal vectors $(u_l)_{l \in \mathbb{N}}$ and $(v_l)_{l \in \mathbb{N}}$ with $\text{span}(u_l : l \in \mathbb{N}) = \mathcal{Y}$, $\text{span}(v_l : l \in \mathbb{N}) = \mathcal{N}(K)^\perp$ 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 a naive use of the discrepancy principle, as implemented in Algorithm 1 without emergency stop, may fail to converge 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 $\mathcal{X} := l^2(\mathbb{N})$ with the standard basis $\{u_k := (0, \dots, 0, 1, 0, \dots)\}$ and consider the diagonal operator

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

with $\hat{x} = 0 = \hat{y} = K\hat{x}$. Hence the $\sigma_l = (1/100)^{\frac{l}{2}}$ are the Eigenvalues of K and

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

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 $\Omega_n := \{Z_i \geq 1, i = 1 \dots n\}$, 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

$$\begin{aligned} \frac{1}{n} \chi_{\Omega_n} &= \delta_n^{est2} \chi_{\Omega_n} \geq \|(KR_{\alpha_n} - Id)\bar{Y}_n\|^2 \chi_{\Omega_n} = |\bar{Z}_n|^2 \|(KR_{\alpha_n} - Id)y\|^2 \chi_{\Omega_n} \\ &\geq \|(KR_{\alpha_n} - Id)y\|^2 \chi_{\Omega_n} \\ &= \sum_{l: \sigma_l^2 < \alpha_n} (y, u_l)^2 \chi_{\Omega_n} = \sum_{l: (1/100)^l < \alpha_n} (y, u_l)^2 \chi_{\Omega_n} \\ &= \sum_{l > \frac{\log(\alpha_n)}{\log(1/100)}} (y, u_l)^2 \chi_{\Omega_n} \geq \frac{\log(1/100)}{\log(\alpha_n)} \chi_{\Omega_n} \\ &\implies \alpha_n \chi_{\Omega_n} < \frac{1}{100^n}. \end{aligned}$$

It follows that

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

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 $\mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \rightarrow \infty$ for $n \rightarrow \infty$.

3.2 Convergence in probability of the discrepancy principle

In this section we show, that the discrepancy principle yields convergence in probability, matching asymptotically the optimal deterministic rate. The proofs of the Theorems 3 and 4 and of Corollary 3 are given in the following section.

Theorem 3 (Convergence of the discrepancy principle) *Assume that K is a compact operator with dense range between Hilbert spaces \mathcal{X} and \mathcal{Y} and that Y_1, Y_2, \dots are i.i.d. \mathcal{Y} -valued random variables with $\mathbb{E}Y_1 = \hat{y} \in \mathcal{R}(K)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. Let R_α be induced by a filter fulfilling Assumption 1 with $v_0 > 1$. Applying Algorithm 1 with or without the emergency stop yields a sequence $(\alpha_n)_n$. Then we have that for all $\varepsilon > 0$*

$$\mathbb{P}(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \varepsilon) \xrightarrow{n \rightarrow \infty} 1,$$

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

Remark 4 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 $\mathbb{E}\|(KR_\alpha - Id)\bar{Y}_n\|^2 \approx \delta_n^{\varepsilon_{st}}$, 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 under an additional source condition: Assuming that there is a $0 < v \leq v_0 - 1$ (where v_0 is the qualification of the chosen regularisation method) such that $K^+\hat{y} = (K^*K)^{\frac{v}{2}} w$ for a $w \in \mathcal{X}$ with $\|w\| \leq \rho$, then

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

for some constant $C > 0$. The next theorem shows a concentration result for the discrepancy principle as implemented in Algorithm 1, with a bound similar to (2).

Theorem 4 (Rate of convergence of the discrepancy principle) Assume that K is a compact operator with dense range between Hilbert spaces \mathcal{X} and \mathcal{Y} . Moreover, Y_1, Y_2, \dots are i.i.d. \mathcal{Y} -valued random variables with $\mathbb{E}Y_1 = \hat{y} \in \mathcal{R}(K)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. Let R_α be induced by a filter fulfilling Assumption 1 with $v_0 > 1$. Moreover, assume that there is a $0 < v \leq v_0 - 1$ and a $\rho > 0$ such that $K^+\hat{y} = (K^*K)^{v/2}w$ for some $w \in \mathcal{X}$ with $\|w\| \leq \rho$. Applying Algorithm 1 with or without the emergency stop yields a sequence $(\alpha_n)_{n \in \mathbb{N}}$. Then there is a constant L , such that

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

We deduce a deterministic bound for $\|R_{\alpha_n} \bar{Y}_n - K^+\hat{y}\|$ (for n large).

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

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

Proof (Corollary 2) By the second assertion in Lemma 1 and Markov's inequality, for any $c, \varepsilon > 0$,

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

□

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 $\mathbb{E}\|R_{\alpha(\delta_n^{est})} \bar{Y}_n - K^+\hat{y}\|^2 \rightarrow 0$ for $n \rightarrow \infty$.

The speed of the convergence will depend on how fast $\mathbb{P}(\Omega_n)$ converges to 1 and might be very slow.

3.3 Almost sure convergence

The results so far delivered 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, \dots is an i.i.d sequence with values in some separable Hilbert space \mathcal{Y} . Moreover, assume that $\mathbb{E}Y_1 = 0$ and $\mathbb{E}\|Y_1\|^2 < \infty$. Then we have that

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

Proof This is a simple consequence of Corollary 8.8 in [24].

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

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

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 $\mathbb{P}(\lim_{n \rightarrow \infty} s_n^2 = \mathbb{E}\|Y_1 - \hat{y}\|^2) = 1$ and $\tau > 1$ it holds that $\sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2} \leq \tau s_n$ 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

4.1 Proofs without emergency case

We will multiple times use the Pythagorean theorem for independent separable Hilbert space valued random variables Z_i with $\mathbb{E}\|Z_i\|^2 < \infty$ and $\mathbb{E}Z_i = 0$,

$$\mathbb{E} \left\| \sum_{i=1}^n Z_i \right\|^2 = \sum_{i=1}^n \sum_{l,l'=1}^{\infty} \mathbb{E}[(Z_i, e_l)(Z_i, e_{l'})] = \sum_{i=1}^n \mathbb{E} \left[\sum_{j=1}^{\infty} (Z_i, e_j)^2 \right] = \sum_{i=1}^n \mathbb{E} \|Z_i\|^2, \quad (3)$$

where $(e_l)_{l \in \mathbb{N}}$ is an orthonormal basis. Based on this, the central ingredient will be the following lemma, which strengthens the pointwise worst case error bound $\|(KR_\alpha - Id)(\bar{Y}_n - \hat{y})\| \leq C_0 \delta_n^{true}$ in some sense.

Lemma 1 *For all $\varepsilon > 0$ and (deterministic) sequences $(q_n)_{n \in \mathbb{N}}$ with $q_n > 0$ and $\lim_{n \rightarrow \infty} q_n = 0$, it holds that*

$$\mathbb{P}(\|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| \geq \varepsilon / \sqrt{n}) \rightarrow 0$$

and

$$\mathbb{P}(|\sqrt{n} \delta_n^{est} - \gamma| \geq \varepsilon) \rightarrow 0$$

for $n \rightarrow \infty$, where $\gamma = 1$ or $\gamma = \sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2}$, depending on if we used the sample variance or not.

Proof By Tschebyscheff's inequality and (3)

$$\begin{aligned}\mathbb{P}(\|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| \geq \varepsilon/\sqrt{n}) &\leq \frac{n}{\varepsilon^2} \mathbb{E}\|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\|^2 \\ &= \frac{1}{\varepsilon^2} \mathbb{E}\|(KR_{q_n} - Id)(Y_1 - \hat{y})\|^2.\end{aligned}$$

Since K has dense range, $KR_{q_n} - Id$ converges to 0 pointwise for $n \rightarrow \infty$ and it follows that $(KR_{q_n} - Id)(Y_1 - \hat{y})$ also converges pointwise to 0. Moreover by (6), $\|(KR_{q_n} - Id)(Y_1 - \hat{y})\|^2 \leq C_0\|Y_1 - \hat{y}\|^2$, so $\mathbb{E}\|(KR_{q_n} - Id)(Y_1 - \hat{y})\|^2 \rightarrow 0$ for $n \rightarrow \infty$ by the dominated convergence theorem. The second assertion only needs a proof for $\gamma = \sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2}$ and then

$$\begin{aligned}n\delta_n^{est2} &= \frac{1}{n-1} \sum_{i=1}^n \|Y_i - \bar{Y}_n\|^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \|Y_i\|^2 - \|\bar{Y}_n\|^2 \right) \\ &\rightarrow \mathbb{E}\|Y_1\|^2 - \|\hat{y}\|^2 = \mathbb{E}\|Y_1 - \hat{y}\|^2 = \gamma^2\end{aligned}$$

almost surely (thus in particular in probability) for $n \rightarrow \infty$ by the strong law of large numbers (Corollary 7.10 in [24]) and the bias-variance-decomposition. Therefore $\sqrt{n}\delta_n^{est} \rightarrow \gamma$ in probability for $n \rightarrow \infty$. \square

For convergence in probability it does not matter how large the error is on sets with diminishing probability and with Lemma 1 we will show, that the probability of certain 'good events' is 1 in the limit of infinitely many measurements.

Define for $q \in (0, 1)$ (as chosen in Algorithm 1)

$$\begin{aligned}\psi_q : \mathbb{R}^+ &\rightarrow \left\{ q^k : k \in \mathbb{N}_0 \right\} \\ \alpha &\mapsto \max \left\{ q^k : q^k \leq \alpha \right\}.\end{aligned}\tag{4}$$

So $\min(q\alpha, 1) \leq \psi_q(\alpha) \leq \alpha$ and by definition, if $\|(KR_{\psi_q(\alpha)} - Id)\bar{Y}_n\| < \delta_n^{est}$, it holds that $\alpha_n \geq \min(q\alpha, 1)$, where α_n is the output of Algorithm 1.

We will also need some well known properties of regularisations defined by filters which fullfill Assumption 1. These are mostly easy modifications from [12].

Proposition 1 *The constants in the following are defined as in Assumption 1. We assume, that K is bounded and linear with non-closed range. 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*

$$\|R_\alpha\| \leq C_R \sqrt{C_F}/\sqrt{\alpha}\tag{5}$$

$$\|I - KR_\alpha\| \leq C_0\tag{6}$$

for all $\alpha > 0$, with $C_0 \geq 1$. If moreover, the filter has qualification $\nu_0 > 0$ and there is a $w \in \mathcal{X}$ with $\|w\| \leq \rho$ such that $K^+ \hat{y} = (K^* K)^{\frac{\nu}{2}} w$ for some $0 < \nu \leq \nu_0$, then

$$\|R_\alpha \hat{y} - K^+ \hat{y}\| \leq C_\nu \rho \alpha^{\nu/2} \quad (7)$$

$$\|R_\alpha \hat{y} - K^+ \hat{y}\| \leq \|KR_\alpha \hat{y} - KK^+ \hat{y}\|^{\frac{\nu}{\nu+1}} C_0^{\frac{1}{\nu+1}} \rho^{\frac{1}{\nu+1}} \quad (8)$$

for all $\alpha > 0$. If additionally, $\nu_0 \geq \nu - 1 > 0$, then

$$\|KR_\alpha \hat{y} - KK^+ \hat{y}\| \leq C_{\nu+1} \rho \alpha^{\frac{\nu+1}{2}}. \quad (9)$$

Moreover, if additionally K is compact, than there is a function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $g(\alpha) \rightarrow \infty$ for $\alpha \rightarrow 0$, such that

$$\lim_{\alpha \rightarrow 0} \|(KR_{\psi_q(\alpha g(\alpha))} - Id)Kx\|/\sqrt{\alpha} = 0. \quad (10)$$

for all $x \in \mathcal{X}$, where ψ_q is given above.

Proof (Proposition 1) (5) and (8) are shown in the proofs of Theorem 4.2 and Theorem 4.17 in [12]. (7) and (8) are Theorem 4.3 in [12]. (6) follows directly from Assumption 1. For (10), we first show that

$$\lim_{\alpha \rightarrow 0} \|(KR_\alpha - Id)Kx\|/\sqrt{\alpha} = 0. \quad (11)$$

We mimic the proof of Theorem 3.1.17 of [31] and set $\varepsilon > 0$. We fix L , such that $C_1^2 \sum_{l=L+1}^\infty (\hat{x}, v_l)^2 < \varepsilon$. Then

$$\begin{aligned} \|(KR_\alpha - Id)K\hat{x}\|^2/\alpha &= \sum_{l=1}^\infty (F_\alpha(\sigma_l^2)\sigma_l^2 - 1)^2 \frac{\sigma_l^2}{\alpha} (\hat{x}, v_l)^2 \\ &\leq \left(\sup_{\lambda > 0} \lambda^{\frac{\nu_0}{2}} |F_\alpha(\lambda)\lambda - 1| \right)^2 \|\hat{x}\|^2 \sum_{l=1}^L \frac{\sigma_l^{2(1-\nu_0)}}{\alpha} \\ &\quad + \left(\sup_{\lambda > 0} \lambda^{\frac{1}{2}} |F_\alpha(\lambda)\lambda - 1| \right)^2 \frac{\sum_{l=L+1}^\infty (\hat{x}, v_l)^2}{\alpha} \\ &\leq C_{\nu_0}^2 L \sigma_L^{2(1-\nu_0)} \|\hat{x}\|^2 \alpha^{\nu_0-1} + C_1^2 \sum_{l=L+1}^\infty (\hat{x}, v_l)^2 < 2\varepsilon \end{aligned}$$

for all $\alpha < \left(\varepsilon^{-1} C_{\nu_0}^2 L \sigma_L^{2(1-\nu_0)} \|\hat{x}\|^2 \right)^{-\frac{1}{\nu_0-1}}$, so $\|(KR_\alpha - Id)Kx\|/\sqrt{\alpha} = 0$ for $\alpha \rightarrow 0$. Now set $\tilde{g}(\alpha) := \sup \left\{ t > 0 : \left\| (KR_{\psi_q(\alpha t)} - Id)Kx \right\|/\sqrt{\alpha} \leq t^{-1} \right\}$. W.l.o.g. \tilde{g} is finite for any $\alpha > 0$. Also, for any $t > 0$

$$\left\| (KR_{\psi_q(\alpha t)} - Id)Kx \right\|/\sqrt{\alpha} = \sqrt{\frac{\psi_q(\alpha t)}{\alpha}} \left\| (KR_{\psi_q(\alpha t)} - Id)Kx \right\|/\sqrt{\psi_q(\alpha t)} \leq \frac{1}{t}$$

for α small enough, because of (11) and since $\psi_q(\alpha t) \leq \alpha t$. So $\tilde{g}(\alpha) \rightarrow \infty$ for $\alpha \rightarrow 0$ and by definition of \tilde{g} the claim holds for $g(\alpha) := \tilde{g}(\alpha) - 1$ (g is well defined for α small enough). \square

Proof (Theorem 4) Set $q_n := \psi_q(b_n)$ where $b_n := \left(\frac{1}{\rho} \frac{\gamma}{4C_{v+1}\sqrt{n}}\right)^{\frac{2}{v+1}}$ with $\gamma = 1$ or $\gamma = \sqrt{\mathbb{E}\|Y_1 - \hat{y}\|^2}$, depending on if we used the sample variance or not, and ψ_q given in (4). Define

$$\Omega_n := \Omega_n(q_n, \gamma) := \left\{ |\sqrt{n}\delta_n^{est} - \gamma| < \gamma/2, \|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| < \gamma/\sqrt{16n} \right\}. \quad (12)$$

Then by (9) and since $q_n \leq b_n$,

$$\begin{aligned} \|(KR_{q_n} - Id)\bar{Y}_n\| \chi_{\Omega_n} &\leq \|(KR_{q_n} - Id)\hat{y}\| \chi_{\Omega_n} + \|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| \chi_{\Omega_n} \quad (13) \\ &\leq C_{v+1} \rho b_n^{\frac{v+1}{2}} \chi_{\Omega_n} + \frac{\gamma}{4\sqrt{n}} \chi_{\Omega_n} = \frac{\gamma}{2\sqrt{n}} \chi_{\Omega_n} < \delta_n^{est} \chi_{\Omega_n}, \end{aligned}$$

so $\alpha_n \chi_{\Omega_n} \geq q b_n \chi_{\Omega_n} \geq q \left(\frac{\delta_n^{est}}{6C_{v+1}}\right)^{\frac{2}{v+1}} \chi_{\Omega_n}$ for n large enough. By (8), (6) and since K has dense image,

$$\begin{aligned} \|R_{\alpha_n} \hat{y} - K^+ \hat{y}\| &\leq \|KR_{\alpha_n} \hat{y} - KK^+ \hat{y}\| \frac{1}{\rho^{\frac{1}{v+1}}} C_0^{\frac{1}{v+1}} \rho^{\frac{1}{v+1}} = \|KR_{\alpha_n} \hat{y} - \hat{y}\| \frac{1}{\rho^{\frac{1}{v+1}}} C_0^{\frac{1}{v+1}} \rho^{\frac{1}{v+1}} \\ &\leq (\|(KR_{\alpha_n} - Id)\bar{Y}_n\| + \|(KR_{\alpha_n} - Id)(\hat{y} - \bar{Y}_n)\|) \frac{1}{\rho^{\frac{1}{v+1}}} C_0^{\frac{1}{v+1}} \rho^{\frac{1}{v+1}} \\ &\leq (\delta_n^{est} + \|(KR_{\alpha_n} - Id)(\hat{y} - \bar{Y}_n)\|) \frac{1}{\rho^{\frac{1}{v+1}}} C_0^{\frac{1}{v+1}} \rho^{\frac{1}{v+1}} \\ &\leq (\delta_n^{est} + C_0 \delta_n^{true}) \frac{1}{\rho^{\frac{1}{v+1}}} C_0^{\frac{1}{v+1}} \rho^{\frac{1}{v+1}} \leq (\delta_n^{est} + \delta_n^{true}) \frac{1}{\rho^{\frac{1}{v+1}}} C_0 \rho^{\frac{1}{v+1}}. \end{aligned}$$

Finally,

$$\begin{aligned} &\|R_{\alpha_n} \bar{Y}_n - K^+ \hat{y}\| \chi_{\Omega_n} \\ &\leq \|R_{\alpha_n} \hat{y} - K^+ \hat{y}\| \chi_{\Omega_n} + \|R_{\alpha_n} \bar{Y}_n - R_{\alpha_n} \hat{y}\| \chi_{\Omega_n} \\ &\leq (\delta_n^{est} + \delta_n^{true}) \frac{1}{\rho^{\frac{1}{v+1}}} C_0 \rho^{\frac{1}{v+1}} \chi_{\Omega_n} + C_R \sqrt{C_F} \frac{\delta_n^{true}}{\sqrt{\alpha_n}} \chi_{\Omega_n} \\ &\leq (2 \max(\delta_n^{est}, \delta_n^{true})) \frac{1}{\rho^{\frac{1}{v+1}}} C_0 \rho^{\frac{1}{v+1}} \chi_{\Omega_n} + C_R \sqrt{C_F} \rho^{\frac{1}{v+1}} \left(\frac{6C_{v+1}}{\delta_n^{est}}\right)^{\frac{1}{v+1}} \frac{\delta_n^{true}}{\sqrt{q}} \chi_{\Omega_n} \\ &\leq L \max \left\{ \delta_n^{est} \frac{1}{\rho^{\frac{1}{v+1}}}, \delta_n^{true} \frac{1}{\rho^{\frac{1}{v+1}}} \left(\frac{\delta_n^{true}}{\delta_n^{est}}\right)^{\frac{1}{v+1}} \right\}, \end{aligned}$$

with $L := 2 \frac{1}{\rho^{\frac{1}{v+1}}} C_0 \rho^{\frac{1}{v+1}} + C_R \sqrt{C_F/q} (6C_{v+1})^{\frac{1}{v+1}}$ and the proof is finished, because $\mathbb{P}(\Omega_n) \rightarrow 1$ for $n \rightarrow \infty$ by Lemma 1. \square

Proof (Proof of Theorem 3) W.l.o.g. we may assume that there are arbitrarily large $l \in \mathbb{N}$ with $(\hat{y}, u_l) \neq 0$, since otherwise we could apply Theorem 4 with any $v > 0$.

Let $\varepsilon' > 0$. Then there is a $L \in \mathbb{N}$ such that $(\hat{y}, u_L) \neq 0$ and $\left(F_{q^k}(\sigma_L^2)\sigma_L^2 - 1\right)^2 > 1/2$ for all $k \in \mathbb{N}_0$ with $q^k \geq \varepsilon'$ (because the F_{q^k} are bounded and $\sigma_l \rightarrow 0$ for $l \rightarrow \infty$). Set

$$\Omega_n := \left\{ |\sqrt{n}\delta_n^{est} - \gamma| < \gamma, (\bar{Y}_n, u_L)^2 \geq (\hat{y}, u_L)^2/2 \right\}. \quad (14)$$

Then for $n \geq 16\gamma^2/(\hat{y}, u_L)^2$,

$$\begin{aligned} \delta_n^{est} \chi_{\Omega_n} &\leq \frac{2\gamma}{\sqrt{n}} \chi_{\Omega_n} < \sqrt{\frac{(\hat{y}, u_L)^2}{4}} \chi_{\Omega_n} \leq \sqrt{\left(F_{q^k}(\sigma_L^2)\sigma_L^2 - 1\right)^2 (\bar{Y}_n, u_L)^2} \chi_{\Omega_n} \\ &\leq \sqrt{\sum_{l=1}^{\infty} \left(F_{q^k}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\bar{Y}_n, u_l)^2} \chi_{\Omega_n} = \|(KR_{q^k} - Id)\bar{Y}_n\| \chi_{\Omega_n} \end{aligned}$$

for all $k \in \mathbb{N}_0$ with $q^k \geq \varepsilon'$. Thus for Ω_n given in (14)

$$\lim_{n \rightarrow \infty} \mathbb{P}(\alpha_n \leq \varepsilon') \geq \lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = 1 \quad (15)$$

by Lemma 1 and since $(\bar{Y}_n, u_L) = \sum_{i=1}^n (Y_i, u_L)/n \rightarrow \mathbb{E}(Y_1, u_L) = (\hat{y}, u_L) \neq 0$ almost surely for $n \rightarrow \infty$. Set $q_n := \psi_q(b_n)$ with $b_n := n^{-1}g(n^{-1})$ and g and ψ_q given in (10) and (4). Define

$$\Omega_n := \left\{ |\sqrt{n}\delta_n^{est} - \gamma| < \gamma/2, \|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| < \gamma/4\sqrt{n} \right\}. \quad (16)$$

Then for n large enough (such that $\|(KR_{q_n} - Id)\hat{y}\|\sqrt{n} \leq \gamma/4$, see (10) with $\alpha = n^{-1}$),

$$\begin{aligned} \|(KR_{q_n} - Id)\bar{Y}_n\| \chi_{\Omega_n} &\leq \frac{1}{\sqrt{n}} \sqrt{n} \|(KR_{q_n} - Id)\hat{y}\| \chi_{\Omega_n} + \|(KR_{q_n} - Id)(\bar{Y}_n - \hat{y})\| \chi_{\Omega_n} \\ &\leq \frac{\gamma}{4\sqrt{n}} \chi_{\Omega_n} + \frac{\gamma}{4\sqrt{n}} \chi_{\Omega_n} \leq \frac{\gamma}{2\sqrt{n}} \chi_{\Omega_n} \leq \delta_n^{est} \chi_{\Omega_n}. \end{aligned} \quad (17)$$

That is $\alpha_n \chi_{\Omega_n} \geq qb_n \chi_{\Omega_n} \geq qn^{-1}g(n^{-1}) \chi_{\Omega_n}$ for n large enough. Finally set

$$\tilde{\Omega}_n := \left\{ \delta_n^{true} \leq \sqrt{\frac{g(n^{-1})}{n}}, \|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \leq \frac{\varepsilon}{2} \right\} \cap \Omega_n,$$

with Ω_n given in (16). So $\mathbb{P}(\tilde{\Omega}_n) \rightarrow 1$ for $n \rightarrow \infty$, since $\mathbb{P}\left(\delta_n^{true} \leq \sqrt{g(n^{-1})/n}\right) \rightarrow 1$, because of $g(n^{-1}) \rightarrow \infty$, $\mathbb{P}(\Omega_n) \rightarrow 1$ by Lemma 1 and $\mathbb{P}\left(\|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \leq \frac{\varepsilon}{2}\right) \rightarrow 1$ by (15) ($\varepsilon' > 0$ is arbitrary). Thus for n large enough (so that $C_R^2 C_F / q \sqrt{g(n^{-1})} \leq \frac{\varepsilon^2}{4}$)

$$\begin{aligned} \|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \chi_{\tilde{\Omega}_n} &\leq \|R_{\alpha_n}\hat{y} - K^+\hat{y}\| \chi_{\tilde{\Omega}_n} + \|R_{\alpha_n}(\bar{Y}_n - \hat{y})\| \chi_{\tilde{\Omega}_n} \\ &\leq \frac{\varepsilon}{2} + \sqrt{\frac{C_R^2 C_F}{\alpha_n}} \delta_n^{true} \chi_{\tilde{\Omega}_n} \leq \frac{\varepsilon}{2} + \sqrt{\frac{C_R^2 C_F}{q \sqrt{g(n^{-1})}}} \leq \varepsilon, \end{aligned}$$

and $\mathbb{P}(\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \varepsilon) \geq \mathbb{P}(\tilde{\Omega}_n) \rightarrow 1$ for $n \rightarrow \infty$. \square

4.2 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 $\|R_{\max\{\alpha_n, 1/n\}}\bar{Y}_n - K^+\hat{y}\|$. It suffices to show that $\mathbb{P}(\alpha_n \geq 1/n) \rightarrow 1$ for $n \rightarrow \infty$.

First assume that $K^+\hat{y} = (K^*K)^{\frac{\nu}{2}}w$ for some $w \in \mathcal{X}$ with $\|w\| \leq \rho$ and $0 < \nu \leq \nu_0 - 1$. With (13) it follows that

$$\mathbb{P}\left(\alpha_n \geq \frac{1}{\rho} \left(\frac{\gamma}{4C_{\nu+1}\sqrt{n}}\right)^{\frac{2}{\nu+1}}\right) \geq \mathbb{P}(\Omega_n) \rightarrow 1 \quad (18)$$

for $n \rightarrow \infty$, with Ω_n given in (12). Otherwise, if there are no such ν, ρ and w , then (17) implies that for all $\varepsilon > 0$

$$\mathbb{P}(\alpha_n \geq qg(n^{-1})/n) \geq \mathbb{P}(\Omega_n) \rightarrow 1 \quad (19)$$

for $n \rightarrow \infty$, with $g(n^{-1}) \rightarrow \infty$ and Ω_n given in (16). Then (18) and (19) together yield $\mathbb{P}(\alpha_n \geq 1/n) \rightarrow 1$ for $n \rightarrow \infty$ and therefore the result. \square

4.3 Proof of Corollary 3

Proof (Corollary 3) Fix $\varepsilon > 0$. Denote by α_n the output of the discrepancy principle with emergency stop and set

$$\Omega_n := \{\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \leq \varepsilon\}. \quad (20)$$

It is

$$\|R_{\alpha}\hat{y} - K^+\hat{y}\| \leq \|R_{\alpha}K - Id\|\|\hat{x}\| \leq \|KR_{\alpha} - Id\|\|\hat{x}\| \leq C \quad (21)$$

for all $\alpha > 0$. By bias-variance decomposition it holds that

$$\begin{aligned} \mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\|R_{\alpha_n}\bar{Y}_n - R_{\alpha_n}\hat{y}\|^2 + \mathbb{E}\|R_{\alpha_n}\hat{y} - K^+\hat{y}\|^2 \\ &\leq \mathbb{E}\left[\|R_{\alpha_n}\|^2 \delta_n^{true2}\right] + C^2 \leq C_R^2 C_F \mathbb{E}\left[\delta_n^{true2} / \alpha_n\right] + C^2 \\ &\leq n C_R^2 C_F \mathbb{E}\delta_n^{true2} + C^2 = C_R^2 C_F \mathbb{E}\|Y_1 - \hat{y}\|^2 + C^2 \leq C', \end{aligned}$$

where C' does not depend on n and where we used $\alpha_n \leq 1$ and (21) in the second step and $\alpha_n \geq 1/n$ in the fourth. By (20) there holds $\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\| \chi_{\Omega_n} \leq \varepsilon$, so

$$\begin{aligned} \mathbb{E}\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \chi_{\Omega_n}\right] + \mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \chi_{\Omega_n^c}\right] \\ &\leq \varepsilon^2 + \mathbb{E}\left[\|R_{\alpha_n}\bar{Y}_n - K^+\hat{y}\|^2 \chi_{\Omega_n^c}\right]. \end{aligned}$$

We apply Cauchy-Schwartz to the second term

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

and we claim that there is a constant A with $\mathbb{E} \|R_{\alpha_n} \bar{Y}_n - K^+ \hat{y}\|^4 \leq A$ for all $n \in \mathbb{N}$.

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

for some constant B , where we used (21) in the second step. First,

$$\begin{aligned}&\mathbb{E} [\delta_n^{true4} / \alpha_n^2] \\ &\leq n^2 \mathbb{E} \|\bar{Y}_n - \hat{y}\|^4 = n^2 \mathbb{E} \left[\sum_{j,j' \geq 1} (\bar{Y}_n - \hat{y}, u_j)^2 (\bar{Y}_n - \hat{y}, u_{j'})^2 \right] \\ &= \frac{1}{n^2} \left(\sum_{j,j' \geq 1} \sum_{i,i',l,l'=1}^n \mathbb{E} [(Y_i - \hat{y}, u_j) (Y_l - \hat{y}, u_j) (Y_{i'} - \hat{y}, u_{j'}) (Y_{l'} - \hat{y}, u_{j'})] \right) \\ &\leq \frac{1}{n^2} \sum_{j,j' \geq 1} \left(n \mathbb{E} [(Y_1 - \hat{y}, u_j)^2 (Y_1 - \hat{y}, u_{j'})^2] + n^2 \mathbb{E} [(Y_1 - \hat{y}, u_j)^2] \mathbb{E} [(Y_1 - \hat{y}, u_{j'})^2] \right. \\ &\quad \left. + 2n^2 (\mathbb{E} [(Y_1 - \hat{y}, u_j) (Y_1 - \hat{y}, u_{j'})])^2 \right) \\ &\leq \frac{n+2n^2}{n^2} \mathbb{E} \left[\sum_{j,j' \geq 1} (Y_1 - \hat{y}, u_j)^2 (Y_1 - \hat{y}, u_{j'})^2 \right] \\ &\quad + \mathbb{E} \left[\sum_{j \geq 1} (Y_1 - \hat{y}, u_j)^2 \right] \mathbb{E} \left[\sum_{j' \geq 1} (Y_1 - \hat{y}, u_{j'})^2 \right] \\ &\leq \frac{n+2n^2}{n^2} \mathbb{E} \left[\left(\sum_{j \geq 1} (Y_1 - \hat{y}, u_j)^2 \right)^2 \right] + \left(\mathbb{E} \left[\sum_{j \geq 1} (Y_1 - \hat{y}, u_j)^2 \right] \right)^2 \\ &= \frac{n+2n^2}{n^2} \mathbb{E} \|Y_1 - \hat{y}\|^4 + (\mathbb{E} [\|Y_1 - \hat{y}\|^2])^2 \leq B_1\end{aligned}$$

for some constant B_1 , where in the fourth step we used that the Y_i are i.i.d, that $\mathbb{E} (Y_1 - \hat{y}, u_j) = (\mathbb{E} [Y_1] - \hat{y}, u_j) = 0$ and that $\mathbb{E} [XY] = \mathbb{E} [X] \mathbb{E} [Y]$ for independent (and integrable) random variables (so the relevant cases are the ones where either all indices i, i', l, l' are equal or exactly pairwise two). Then we used Jensen's inequality in the fifth step. Moreover,

$\mathbb{E} \left[\delta_n^{true^2} / \alpha_n \right] \leq n \mathbb{E} \left[\delta_n^{true^2} \right] = \mathbb{E} \|Y_1 - \hat{y}\|^2 = B_2$, so the claim holds for $A = B(B_1 + B_2 + 1)$. By Theorem 3, for n large enough, $\mathbb{P}(\Omega_n) \geq 1 - \varepsilon^4/A$, so

$$\mathbb{E} \|R_{\alpha_n} \bar{Y}_n - K^+ \hat{y}\|^2 \leq \varepsilon \mathbb{E} [\chi_{\Omega_n}] + \sqrt{\mathbb{E} \|R_{\alpha_n} \bar{Y}_n - K^+ \hat{y}\|^4 \mathbb{P}(\Omega_n^C)} \leq \varepsilon^2 + \varepsilon^2 \rightarrow 0$$

for $\varepsilon \rightarrow 0$. \square

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 [22])

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

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

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

Ultimately 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

$$K : L^2([0, 1]) \rightarrow L^2([0, 1])$$

$$f \mapsto Af = g : x \mapsto \int_0^x f(y) dy.$$

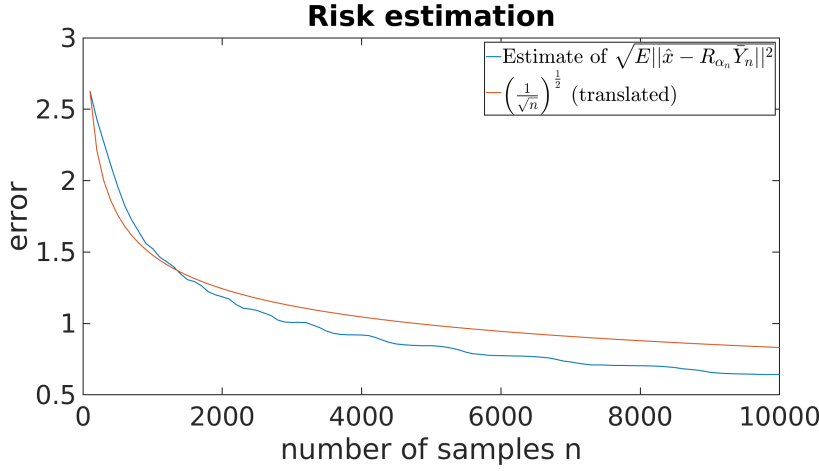


Fig. 1 Estimated Risk of a binary option.

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, \dots identically distributed to s and set $Y_i := e^{-rT} Q\chi_{\{S_0 e^{TZ_i} \geq K\}}$. Then we have $\mathbb{E}Y_i = e^{-rT} Q\mathbb{P}(S_0 e^{TZ_i} \geq K) = e^{-rT} Qp(S_0) = V(S_0)$ 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 $\hat{x} = K^+ \hat{y} \in \mathcal{X}_V$ for all $v > 0$ using the transformation $z(\xi) = 0, 5e^{\sqrt{0.3}\xi} - 0, 15$. Since the qualification of the Tikhonov regularisation is 2, Theorem 4 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 [14]) or use specific regularization methods for numerical differentiation [18].

5.2 Inverse heat equation

We consider the toy problem 'heat' from [19]. We chose the discretisation level $m = 100$ and set $\sigma = 0.7$. Under this choice, the last seven singular values (calculated

Fig. 2 Estimated relative L^2 error for 'heat', that is $\sqrt{\sum_{i=1}^{200} e_i^2 / 200}$ where e_i is the relative error $\|R_{\alpha_n} \bar{Y}_n - K^+ \hat{y}\| / \|K^+ \hat{y}\|$ of the i -th run.

	dp	dp+es	a priori
$n = 10^3$	572.49	0.66	0.83
$n = 10^4$	79.45	0.49	0.76
$n = 10^5$	107.19	0.31	0.69

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 $1, 1/2^{3/4}, \dots, 1/m^{3/4}$. 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 principle 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 similar with increasing sample size - with the crucial difference, that the outliers as such we denote the red crosses above the blue box, thus the cases where the method 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 18 for $n = 10^5$, indicating the (slow) convergence in probability of the discrepancy principle. Note that $\delta_n^{true} / \delta_n^{est} \approx 1.9$ (in average), if we only consider the runs yielding outliers. This illustrates, that the lack of convergence in L^2 is caused by the occasional underestimation of the data error.

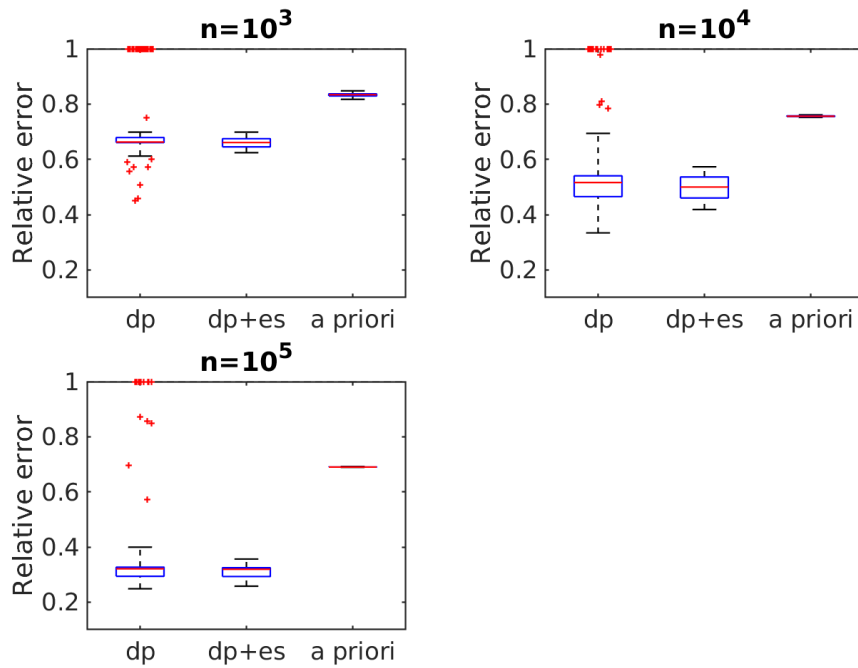


Fig. 3 Comparison of Tikhonov regularisation with discrepancy principle (dp, Algorithm 1), discrepancy principle with emergency stop (dp+es, Algorithm 1 (optional)) and a priori choice for 'heat'. Boxplots of the relative errors $\|R_{\alpha_n} \tilde{Y}_n - K^+ \hat{y}\| / \|K^+ \hat{y}\|$ for 200 simulations 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. Bakushinskii, 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 Bakushinskii 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. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
13. 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)
14. Gerstner, T., Harrach, B., Roth, D.: Monte Carlo pathwise sensitivities for barrier options. arXiv preprint arXiv:1804.03975 (2018)
15. 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)
16. Ghosal, S., Van der Vaart, A.: Fundamentals of nonparametric Bayesian inference, vol. 44. Cambridge University Press (2017)
17. Giné, E., Nickl, R.: Mathematical foundations of infinite-dimensional statistical models, vol. 40. Cambridge University Press (2016)
18. Hanke, M., Scherzer, O.: Inverse problems light: numerical differentiation. *The American Mathematical Monthly* **108**(6), 512–521 (2001)
19. Hansen, P.C.: Discrete inverse problems: insight and algorithms, vol. 7. Siam (2010)
20. Hassan, U., Anwar, M.S.: Reducing noise by repetition: introduction to signal averaging. *European Journal of Physics* **31**(3), 453 (2010)
21. Hofinger, A.: Ill-posed problems: Extending the deterministic theory to a stochastic setup. Trauner (2006)
22. Hull, J.C., Basu, S.: Options, futures, and other derivatives. Pearson Education India (2016)
23. Kaipio, J., Somersalo, E.: Statistical and computational inverse problems, vol. 160. Springer Science & Business Media (2006)
24. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes, vol. 23. Springer Science & Business Media (1991)
25. Lu, S., Mathé, P.: Discrepancy based model selection in statistical inverse problems. *Journal of Complexity* **30**(3), 290–308 (2014)
26. 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)
27. Lyons, R.G.: Understanding digital signal processing, 3/E. Pearson Education India (2004)
28. 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)
29. Mathé, P., Pereverzev, S.V.: Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse problems* **19**(3), 789 (2003)
30. 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)
31. 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>
32. Rieder, A.: Keine Probleme mit inversen Problemen: eine Einführung in ihre stabile Lösung. Springer-Verlag (2013)
33. Tikhonov, A., Arsenin, V.Y.: Methods for solving ill-posed problems. John Wiley and Sons, Inc (1977)
34. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis* **14**(4), 651–667 (1977)
35. 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)