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 is usually estimated through averaging of multiple measurements. In this paper, we integrate the probably most 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 $Y_1, Y_2, \ldots$ of it, which are independent, identical distributed and unbiased ($\mathbb{E}Y_i = \hat{y}$) random variables [1]. The 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 $K^+\bar{Y}_n \not\to K^+\hat{y}$ for $n \to \infty$, because the generalised inverse (section 6.2, Definition 6) of a compact operator is not continuous. So the inverse is replaced with a family of continuous approximations $(R_\alpha)_{\alpha>0}$, called regularisation.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. A natural guess is

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

where either $s_n = 1$ is constant or $s_n = \sqrt{\sum_{i\leq n} \|Y_i - \bar{Y}_n\|^2/(n-1)}$ is the square root of the sample variance. The 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, 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 $Y_1, Y_2, \ldots$ are i.i.d. $\mathscr{Y}-$valued random variables which fullfill $\mathbb{E}[Y_1] = \hat{y} \in \mathscr{D}(K^+)$ and $0 < \mathbb{E}\|Y_1 - \hat{y}\|^2 < \infty$. 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*

[1] Rigorous definitions are given in section 6.1 and 6.2.

principle (see Algorithm 1 *on page* 8*). 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 $K^+\hat{y}$ *satisfies* $\|K^+\hat{y}\|_\nu \leq \rho$ *for* $\rho > 0$ *and* $0 < \nu < \nu_0$ *(section 6.2, Definition 11 and 16), 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 does not yield convergence in L^2 for 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.
The main difference to existing results is, that we have no assumptions for the error distribution. The natural approach can be easily used by everyone, who can measure multiple times.
Stochastic inverse problems are an active field of research with close ties to high dimensional statistics ([14],[13],[25]). 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 [18]). This opposes the frequentist setting, where the inverse problem is assumed to have a deterministic, exact solution ([7],[5]). We are working in the frequentist setting, but compared to the existing literature we are staying much closer to the classic deterministic theory of linear inverse problems ([10],[26],[27]). The error is often modelled as a Hilbert space process (for example Gaussian white noise), that is in contrast to our, more classic error model, it is not an element of the Hilbert space itself. Still, usually there are made additional distribution specific assumptions. This more general error model makes it impossible to determine the regularisation parameter through the discrepancy principle [24], which is known to work optimal in the classic deterministic setting. Instead, typical methods to determine the regularisation parameter are cross validation [29], Lepski's balancing principle [23] or penalised empirical risk minimisation [8]. Methods motivated by the discrepancy principle were studied recently ([21],[6],[22]). In the frequentist literature, people are overwhelmingly aiming to prove (optimal) convergence in L^2 (often called convergence of the mean squared error) via so called oracle inequalities. Another approach is to transfer results from the classical deterministic theory using the Ky-Fan metric, which metrises convergence in probability ([16],[12]). Finally, aspects of the Bakushinskii veto for stochastic inverse problems are discussed in ([3],[4],[30]) under assumptions for the noise distribution.
It may be surprising that the natural approach has not been studied yet. We conclude the introduction with a small insight in the main difficulties, which are in our opinion the reason for this.
By the strong law of large numbers (section 6.1, Theorem 10) for integrable independent random variables it holds that

$$\mathbb{P}\left(\lim_{n\to\infty}\delta_n^{true}=0\right)=\mathbb{P}\left(\lim_{n\to\infty}\|\bar{Y}_n-\hat{y}\|=0\right)=\mathbb{P}\left(\lim_{n\to\infty}\|\bar{Y}_n-\mathbb{E}Y_1\|=0\right)=1. \quad (1)$$

Then the definition of a regularisation method gives

$$\mathbb{P}\left(\lim_{n\to\infty}\|R_{\alpha(\delta_n^{true},\bar{Y}_n)}\bar{Y}_n-K^+\hat{y}\|=0\right)=1. \quad (2)$$

Infact, by Bakushinskii's veto [2] it is not possible to achieve convergence without knowing an upper bound for the data error δ_n^{true}.
We quickly motivate why δ_n^{est} is a natural guess for δ_n^{true}. In our independent setting, the central limit theorem (section 6.1, Theorem 9) gives us the asymptotic distribution of δ_n^{true}. Roughly speaking, for large n,

$$\delta_n^{true}=\|\bar{Y}_n-\hat{y}\|=\frac{1}{\sqrt{n}}\left\|\frac{\sum_{i\leq n}(Y_i-\hat{y})}{\sqrt{n}}\right\|\approx\frac{\|Z\|}{\sqrt{n}}, \quad (3)$$

where Z is a centred Gaussian random variable with the same covariance structure as Y_1 (in particular, $\mathbb{E}\|Z\|^2=\mathbb{E}\|Y_1-\hat{y}\|^2$). Moreover, we can calculate the second moment of δ_n^{true} for any n:

$$\sqrt{\mathbb{E}[\delta_n^{true2}]}=\sqrt{\mathbb{E}\left\|\left(\frac{\sum_{i\leq n}(Y_i-\hat{y})}{n}\right)^2\right\|}=\sqrt{\frac{\sum_{i\leq n}\mathbb{E}\|Y_i-\hat{y}\|^2}{n^2}}=\frac{\sqrt{\mathbb{E}\|Y_1-\hat{y}\|^2}}{\sqrt{n}}. \quad (4)$$

So, in some sense $\delta_n^{true}\sim 1/\sqrt{n}$. But, for $n\to\infty$, $\sqrt{n}\delta_n^{true}$ converges (weakly) to a Gaussian random variable, while $\sqrt{n}\delta_n^{est}$ converges (almost surely) to a constant, see Figure 1.
That is, the estimator δ_n^{est} often underestimates the true data error δ_n^{true}. This is one reason, that this naive approach is not used widely in the community. It is more popular to choose estimators, which rather overestimate the true data error ([9],[28]). But we believe that many people in applications, especially those without a mathematical background, use the naive way presented here to estimate their data (error). In fact, the underestimating is only a minor problem for a priori regularisation, since it is clear from the deterministic theory, that we are allowed to underestimate the error by a constant and in our given case, we underestimate by a random constant. On the other hand, using the discrepancy principle together with an underestimated error yields non convergence in a general classic setting. It turns out that with our error model the problem of underestimating is compensated by the following: In deterministic regularisation one always considers the worst case error, that is the noise may come from arbitrarily bad directions (i.e. it is supported mainly on eigenvectors corresponding to arbitrarily small eigenvalues). Here, our estimator $\bar{Y}_n$, when multiplied by $\sqrt{n}$, somehow stabilises along (random) directions determined by Z and (thus determined by the distribution of Y_1).
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

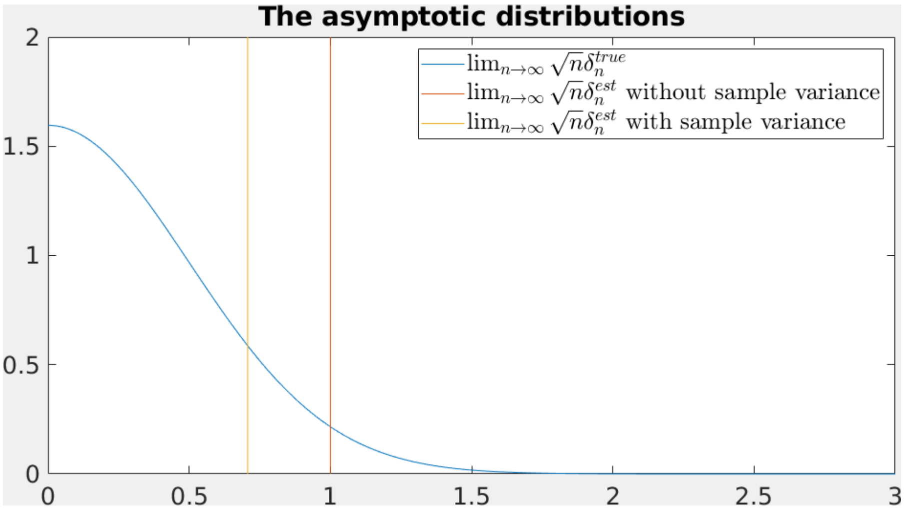

Fig. 1 Here we see the distributions of $\lim_{n\to\infty}\sqrt{n}\delta_n^{true}$ (blue) and $\lim_{n\to\infty}\sqrt{n}\delta_n^{est}$ (red and yellow), for the one-dimensional case and $\sqrt{\mathbb{E}|Y_1-\hat{y}|^2}=1/\sqrt{2}$. The first one is a halfnormal distribution, while the latter ones are in fact degenerated delta distributions, due to the almost sure convergence. Moreover, if we use the sample variance for δ_n^{est}, we can quantify the asymptotic probability of underestimating the true data error δ_n^{true} with Lemma (4) (section 6.2)). For large n and any $\tau>0$ it holds that $\mathbb{P}(\tau\delta_n^{est}>\delta_n^{true})\approx\mathbb{P}\left(\frac{\|Z\|}{\sqrt{\mathbb{E}\|Y_1-\hat{y}\|^2}}<\tau\right)\geq\Phi(\tau)-\Phi(-\tau)$, where Φ is the cumulative distribution function of a standard Gaussian. In particular, for τ=1 we have that $\mathbb{P}(\delta_n^{est}<\delta_n^{true})\approx 0.31$, which equals the area under the blue graph on the right of the yellow line.

δ_n^{est} to obtain almost sure convergence and we give some numerical demonstrations of the main Theorems 3 and 4. In section 6 we collected definitions and facts from probability theory and deterministic regularisation theory.

1.1 The error model

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 $0<\mathbb{E}\|Y_1\|^2<\infty$. A justification for this model is given in Theorem 6 (section 6.1). 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.

2 A priori regularisation

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 we also know that $\mathbb{E}\delta_n^{true2} = \mathbb{E}\|Y_1 - \hat{y}\|^2/n$ (see equation (4)), 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.

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 in fact it suffice to assume K to be bounded and linear. Remember that R_α is linear. Thus $\mathbb{E}[R_\alpha Y_1] = R_\alpha \mathbb{E}[Y_1] = R_\alpha \hat{y}$ and with Corollary 3 (section 6.1)

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

Therefore,

$$\begin{aligned}
\mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - K^+\hat{y}\|^2 &= \mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - R_{\alpha(\delta_n^{est})}\hat{y} + R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \mathbb{E}\|R_{\alpha(\delta_n^{est})}\bar{Y}_n - R_{\alpha(\delta_n^{est})}\hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \frac{\mathbb{E}\|R_{\alpha(\delta_n^{est})}Y_1 - R_{\alpha(\delta_n^{est})}\hat{y}\|^2}{n} + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&\le \frac{\|R_{\alpha(\delta_n^{est})}\|^2}{n}\mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&= \|R_{\alpha(\delta_n^{est})}\|^2 \delta_n^{est2}\mathbb{E}\|Y_1 - \hat{y}\|^2 + \|R_{\alpha(\delta_n^{est})}\hat{y} - K^+\hat{y}\|^2 \\
&\to 0 \qquad \text{for} \quad n \to \infty.
\end{aligned}$$

□

As in the deterministic case, under source conditions (section 6.2, Definition 11) we can prove convergence rates. It is common to consider regularisations induced by a filter (section 6.2, Definition 14). We make the following standard assumptions for the filter:

Assumption 1 *$(F_\alpha)_{\alpha>0}$ is a regularising filter with qualification $\nu_0 > 0$. Moreover, there is a constant $C_F > 0$ such that $|F_\alpha(\lambda)| \le C_F/\alpha$ for all $0 < \lambda \le \|K\|^2$.*

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

1. Tikhonov regularisation,
2. generalised Tikhonov regularisation,
3. truncated singular value regularisation,
4. Landweber iteration.

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* $\bar{Y}_n := \sum_{i\le n} Y_i/n$ *and* $\delta_n^{est} = n^{-1/2}$*. Assume that for* $0 < \nu \le \nu_0$ *and* $\rho > 0$ *we have that* $K^+\hat{y} \in \mathscr{X}_\nu, \|K^+\hat{y}\|_\nu \le \rho$*. Then for*

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

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

Proof We proceed similiary to the proof of Theorem 1, using additionally Proposition 2 and 3 of section 6.2.

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

□

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

$$\|(KR_{\alpha_n} - Id)\bar{Y}_n\| \approx \delta_n^{true} = \|\bar{Y}_n - \hat{y}\|, \tag{5}$$

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 with Algorithm 1.

Algorithm 1 Discrepancy principle with estimated data error

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

Algorithm 1 converges under two assumptions. The first one is also necessary in the classical deterministic setting. Equation (5) 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, in fact the distribution of Y_1 is usually absolutely continuous wich implies that $\mathbb{P}(Y_1 = ... = Y_n) = 0$ for all $n \in \mathbb{N}$.
Under this assumptions, at first glance, one may try to prove convergence in squared expectation, similiar to the previous section. But here we have the problem, that α_n is random, since it depends by defintion on the random data. Indeed, we show that Algorithm 1 does not yield such a convergence. The rough idea is the following: with a diminishing probability p we are underestimating the data error significantly, thus the discrepancy principle gives a way too small α and we still have $p\|R_\alpha\| \gg 1$. We give a small rigorous example. As a compact operator, K admits a singular value decomposition (section 6.2, Definition 10) which we denote by (σ_l, v_l, u_l).

3.1 A counter example for L^2 convergence

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

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

It follows that

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

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

We saw that, unlike in the a priori setting, we have no convergence in squared expectation. The following main theorem of the section assures convergence in probability. 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 for the filter (section 6.2, Definition 15).

Assumption 2 *$(F_\alpha)_{\alpha>0}$ is a regularising and monotone filter with qualification $\nu_0 > 1$. Moreover, there is a constant $C_F > 0$ such that $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all $0 < \lambda \leq \|K\|^2$.*

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

Remark 3 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 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 4 If one tried to argue as in Remark 1 to show convergence in squared expectation, one would have to determine the regularisation parameter not as given by equation (5), 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

$$\sup_{y^\delta : \|y^\delta - \hat{y}\| \leq \delta} \|R_{\alpha(y^\delta,\delta)}y^\delta - K^+\hat{y}\| \leq C\rho^{\frac{1}{\nu+1}}\delta^{\frac{\nu}{\nu+1}}. \tag{6}$$

The next theorem shows the analogous result for the natural approach: A similiar bound to (6) 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 (6) 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 (6) 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 $K^+\hat{y} \in \mathscr{X}_\nu$ and $\|K^+\hat{y}\|_\nu \leq \rho$. Applying Algorithm 3 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 $\varepsilon > \varepsilon' > 0$. □

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 [20].

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

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

where σ_n is the square root of the sample variance. Since $\mathbb{P}(\lim_{n\to\infty}\sigma_n^2 = \mathbb{E}\|Y_1 - \hat{y}\|^2) = 1$ and $\tau > 1$ it holds that $\sqrt{\mathbb{E}\|Y_1-\hat{y}\|} \leq \tau\sigma_n$ 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

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. 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

$$\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \approx \sqrt{n}\|(KR_{\alpha_n} - Id)Z\| \leq c\sqrt{n} \sim c\delta_n^{est} \tag{7}$$

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 (section 6.2, Definition 15), which is natural and 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 (7) 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

$$\|(KR_\alpha - Id)y\|^2 = \sum_l \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 (y,u_l)^2 \tag{8}$$

$$\leq \sum_l \left(F_\beta(\sigma_l^2)\sigma_l^2 - 1\right)^2 (y,u_l)^2 = \|(KR_\beta - Id)y\|^2 \tag{9}$$

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

$$\frac{1}{\sigma_l^2} \geq F_\alpha(\sigma_l^2) \geq F_\beta(\sigma_l^2)$$
$$\iff |1 - F_\alpha(\sigma_l^2)\sigma_l^2| = 1 - F_\alpha(\sigma_l^2)\sigma_l^2 \leq 1 - F_\beta(\sigma_l^2)\sigma_l^2 = |1 - F_\beta(\sigma_l^2)\sigma_l^2|.$$

We begin with the special case, where the problem is in fact 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 $m \leq n$

$$\mathbb{P}(\|(KR_{x_n} - Id)\bar{Y}_n\| > \delta_n^{est}) = \mathbb{P}(\|(KR_{x_n} - Id)(\bar{Y}_n - \hat{y})\| > \delta_n^{est}) \tag{10}$$

$$= \mathbb{P}\left(\left\|(KR_{x_n} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| > \sqrt{n}\delta_n^{est}\right) \tag{11}$$

$$\leq \mathbb{P}\left(\left\|(KR_{x_m} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\right\| > \sqrt{n}\delta_n^{est}\right) \tag{12}$$

because of equation (8) and (9). Using the central limit theorem gives $\sum_{i\leq n}(Y_i - \hat{y})\sqrt{n} \xrightarrow{\text{w}} Z$ 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 $\|(KR_{x_m} - Id)\sum_{i\leq n}(Y_i - \hat{y})/\sqrt{n}\| \xrightarrow{\text{w}} \|(KR_{x_m} - Id)Z\|$ for $n \to \infty$ by the continuous mapping theorem (section 6.1, Theorem 7). 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 (section 6.1, Theorem 8) implies that

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

Finally, by Portemanteau's lemma (section 6.1, Lemma 3)

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

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 $\varepsilon>0$,

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

Thus finally

$$\limsup_{n\to\infty}\mathbb{P}(\|(KR_{x_n}-Id)\bar{Y}_n\|>\delta_n^{est})\le\mathbb{P}(\|(KR_{x_m}-Id)Z\|\ge\gamma),\ \ \forall\, m\in\mathbb{N} \tag{13}$$

$$\overset{m\to\infty}{\longrightarrow}0. \tag{14}$$

Now we set

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

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

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

So equation (13) and (14) yield (for n large enough)

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

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 $\alpha\le a_0$, the boundedness of F_α implies that there is a $L\in\mathbb{N}$, such that $(\hat{y},u_l)=0$ for all $l\ge L$. So

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

We deduce that

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

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

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

□

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 $\mathbb{P}(\Omega_{n,c}) \overset{n\to\infty}{\longrightarrow} 1$ 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

$$
\mathbb{P}\left((\bar{Y}_n,u_1)^2 \ge \frac{1}{2}(\hat{y},u_1)^2,\ldots,(\bar{Y}_n,u_k)^2 \ge \frac{1}{2}(\hat{y},u_k)^2\right) \overset{n\to\infty}{\longrightarrow} 1. \tag{15}
$$

So we define

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

By equation (15) N_k is well defined with $1 = N_1 \le N_2 \le N_3 \le \ldots$. For all $n\in\mathbb{N}$ we set

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

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

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

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.

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

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

$$\frac{4\gamma^2}{n}\chi_{\Omega_{n,c}^1} \geq {\delta_n^{est}}^2 \chi_{\Omega_{n,c}^1} \geq \|(KR_{\alpha_n} - Id)\bar{Y}_n\|^2 \chi_{\Omega_{n,c}^1} \tag{16}$$

$$= \sum_{l\geq 1}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\bar{Y}_n, u_l)^2 \chi_{\Omega_{n,c}^1} \tag{17}$$

$$\geq \sum_{l\leq K_n}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\bar{Y}_n, u_l)^2 \chi_{\Omega_{n,c}^1} \tag{18}$$

$$\geq \frac{1}{2}\sum_{l\leq K_n}\left(F_{\alpha_n}(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\hat{y}, u_l)^2 \chi_{\Omega_{n,c}^1} \tag{19}$$

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 (16) to (19)

$$\mathbb{P}(\alpha_n \leq g(n)) \geq \mathbb{P}(\Omega_{n,c}^1) \to 1. \tag{20}$$

So we see, that for n large enough

$$\mathbb{P}(\|(KR_{a_0} - Id)\bar{Y}_n\| > \delta_n^{est}, \lim_{k\to\infty}\alpha_k = 0) \geq \mathbb{P}(\Omega_{n,c}^1), \tag{21}$$

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

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

Now we follow the ideas of the proof of Lemma 1(equation (10) to (14)). For $m \leq n$

$$\mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta_n^{est}\right) \tag{22}$$
$$\leq \mathbb{P}\left(\|(KR_{g(m)} - Id)\frac{\sum_{i\leq n}(Y_i - \hat{y})}{\sqrt{n}}\| \geq c\sqrt{n}\delta_n^{est}\right) \tag{23}$$
$$\overset{n\to\infty}{\longrightarrow} \mathbb{P}\left(\|(KR_{g(m)} - Id)Z\| \geq c\gamma\right) \overset{m\to\infty}{\longrightarrow} 1. \tag{24}$$

We conclude that

$$\mathbb{P}\left(\Omega_{n,c}^2\right) = \mathbb{P}\left(\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta_n^{est}\right) \tag{25}$$
$$\geq \mathbb{P}\left(\|(KR_{\alpha_n} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta_n^{est}, \alpha_n \leq g(n)\right) \tag{26}$$
$$\geq \mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| \leq c\delta_n^{est}, \alpha_n \leq g(n)\right) \tag{27}$$
$$= 1 - \mathbb{P}\left(\{\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta_n^{est}\} \cup \{\alpha_n > g(n)\}\right) \tag{28}$$
$$\geq 1 - \mathbb{P}\left(\|(KR_{g(n)} - Id)(\bar{Y}_n - \hat{y})\| > c\delta_n^{est}\right) - \mathbb{P}\left(\alpha_n > g(n)\right) \tag{29}$$
$$\overset{n\to\infty}{\longrightarrow} 1, \tag{30}$$

where we used the monotonicity of $\mathbb{P}$ in line(26), the monotinicity of F_α in line (27), the subadditivity of $\mathbb{P}$ in line (29) and equation (20) and (22) to (24) in line (30). The same argumentation shows that $\lim_{n\to\infty}\mathbb{P}(\Omega_{n,c}^3) = 1$ too. From the inequality (21) it follows that for n large enough

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

The proof is concluded by (for n large enough)

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

□

Now we can give the proofs for the main theorems.

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

$$\|R_{\alpha_n}\hat{y} - K^+\hat{y}\|\chi_{\Omega_{n,c}} \to 0 \tag{31}$$

Since $\left\|\frac{\sum_{i\le n} Y_i - \hat{y}}{\sqrt{n}}\right\| \xrightarrow{\text{w}} \|Z\|$ 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 (section 6.1, Theorem 8). 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 in fact 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

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

So by Proposition 2,

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

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

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

and

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

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)

$$
\begin{aligned}
&\sum_l \left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2-1\right)^2\frac{\sigma_l^2}{\alpha_n/q}(\hat{x},v_l)^2\chi_{\Omega_{n,c}}\\
&=\sum_{l\le N}\left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2-1\right)^2\frac{\sigma_l^2}{\alpha_n/q}(\hat{x},v_l)^2\chi_{\Omega_{n,c}}\\
&\quad+\sum_{l>N}\left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2-1\right)^2\frac{\sigma_l^2}{\alpha_n/q}(\hat{x},v_l)^2\chi_{\Omega_{n,c}}\\
&\le\sum_{l\le N}\left(F_{\alpha_n/q}(\sigma_l^2)\sigma_l^2-1\right)^2\frac{\sigma_l^2}{\alpha_n/q}(\hat{x},v_l)^2\chi_{\Omega_{n,c}}+C_1^2\sum_{l>N}(\hat{x},v_l)^2\\
&\le\sum_{l\le N}C_{\nu_0}^2\sigma_l^{2-2\nu_0}(\alpha_n/q)^{\nu_0-1}(\hat{x},v_l)^2\chi_{\Omega_{n,c}}+\varepsilon/2\\
&\le C_{\nu_0}^2\sigma_N^{2-2\nu_0}(\alpha_n/q)^{\nu_0-1}\chi_{\Omega_{n,c}}+\varepsilon/2\le\varepsilon\quad\text{for } n \text{ large enough.}
\end{aligned}
$$

Thus $\lim_{n\to\infty}\|R_{\alpha_n}\|\delta_n^{est}\chi_{\Omega_{n,c}} = 0$ also. Together with (31) 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 3

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

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

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

Now we treat the second term. Proposition 2 yields

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

By Proposition 3 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

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

Putting it all together

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

□

5 Numerical demonstration

In this section we give examples for each of the Theorems 4 and 3. We state the idealised infinte dimensional problems. The operators K in the examples will be Fredholm integral operators of the first kind and hence compact. In the numerical simulation we use high dimensional discretisations. The chosen regularisation method is Tikhonov regularisation together with the discrepancy principle. The discretised large systems of linear equations are solved iteratively using the conjugate gradient method. In order to use Tikhonov regularisation we need K'. As a discretisation for this we use $(K_{disc})'$, that is the conjugate of our discretisation K_{disc} of K. Another option would be to discretise K' directly, but then $K_{disc}(K')_{disc}$ would not be symmetric and positive semidefinite.

5.1 Derivation of a binary option

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 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 [17])

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

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 derivation. Set $\mathscr{X} = \mathscr{Y} = L^2([0,1] =$ and define

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

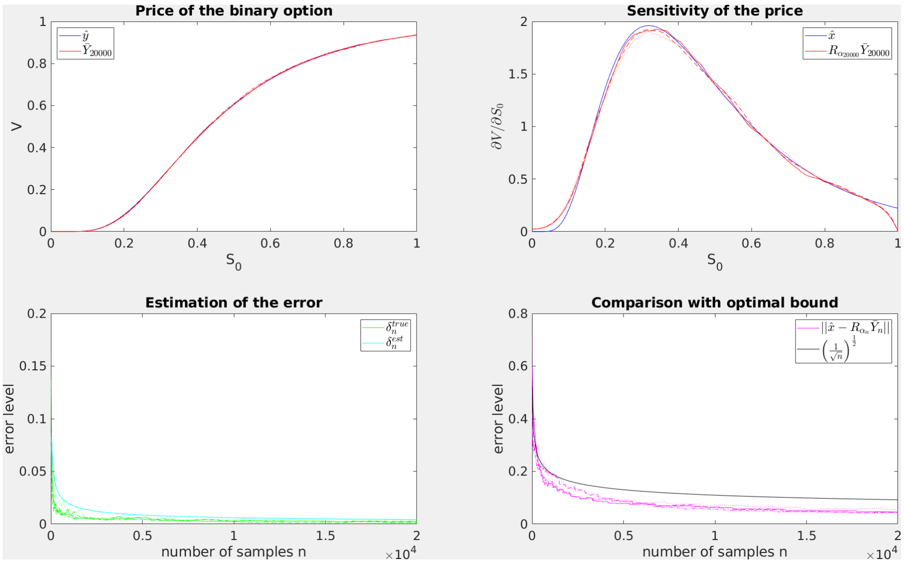

Fig. 2 Risk estimation of a binary option. Four simulations are plotted in each small window. (*Upper left*: The true data $\hat{y}$ and the corrupted data $\bar{Y}$ for the last step $n = 20000$. *Upper right*: The true solution $\hat{x}$ and the approximation $R_\alpha \bar{Y}$ for the last step $n = 20000$. *Lower left*: Evolution of δ_n^{est} and δ_n^{true} with respect to the number of samples. *Lower right*: Comparison of $\sim (1/\sqrt{n})^{\frac{1}{2}}$ and the actual error $\|\hat{x} - R_{\alpha_n}\bar{Y}_n\|$. Here seemingly $\mathbb{P}(\delta_n^{true} \leq \delta_n^{est})$ is larger than the lower bound from Lemma 4, which is approximately 0.6827.

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 $m = 100000$ elements (we can calculate Kg exactly for such a spline g). We use in total $n = 20000$ random variables. 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 $\nu \in \mathbb{R}$ 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 $(1/\sqrt{n})^{\frac{1}{2}}$. The numerical results are visualised in Figure 5.1.
Note that this is only an adademic example for demonstration. In the given situation one would rather proceed as in [15],[1] or [11].

5.2 Image deblurring

When taking a picture, sharp edges of objects are usually blurred. An idealised monocoloured picture may be modelled as a function $v \in L^2([0,1]^2)$ and the blurring process is written as

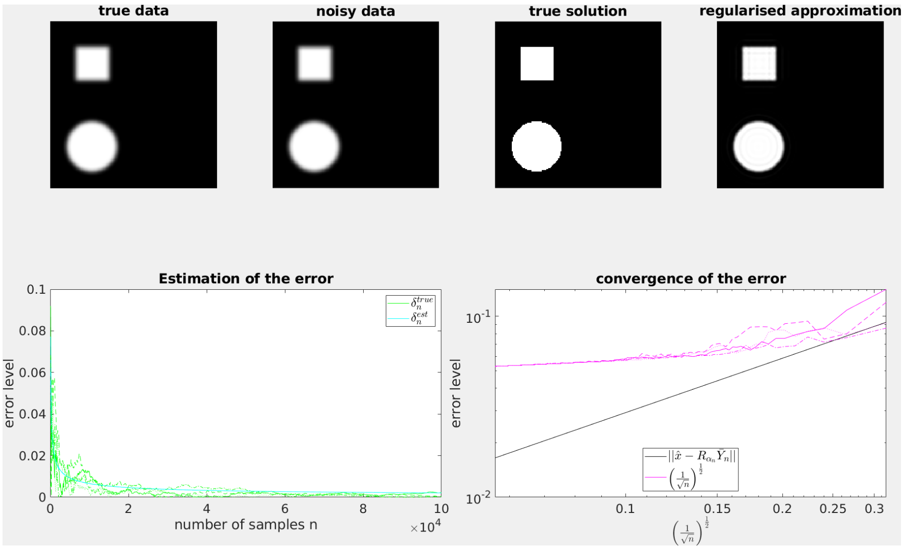

Fig. 3 Deblurring of a blurred image, we did four simulations. In the upper row only the results of the first simulation are plotted. (*Upper row*: The true data $\hat{y}$ and the corrupted data $\bar{Y}$ for the last step on the left side, the true solution $\hat{x}$ and the approximation $R_\alpha \bar{Y}$ for the last step on the right side. *Lower left*: Evolution of δ^{est} and δ^{true}. *Lower right*: Evolution of the asymptotics of the actual error $\|\hat{x} - R_\alpha \bar{Y}\|$ in a logarithmic plot together with the (shifted) identity line, to demonstrate the subpolynomial convergence.

$$
\begin{aligned}
K :& L^2\left([0,1]^2\right) \to L^2\left([0,1]^2\right) \\
& v_{sharp} \mapsto v_{blur} : v_{blur}(x) = \int_{[0,1]^2} k(x-y) v_{sharp}(y) dy,
\end{aligned}
$$

with a Gaussian kernel $k(z) := e^{-\frac{\|z\|^2}{2\sigma^2}} / 2\pi\sigma^2$. We use

$$
\hat{x} = v_{sharp}(x_1, x_2) = \chi_{\max(|x_1-0.25|, x_2-0.25|)<0.1} + \chi_{(x_1-0.25)^2+(x_2-0.75)^2<0.15^2}
$$

and

$$
\hat{y} = K\hat{x},
$$

compare to figure 5.2. Monocoloured pictures are stored on a computer as vectors $(a_{ij})_{ij} \in \mathbb{R}^{k \times l}$, where (a_{ij}) is the intensity of the (i,j)-th pixel. Hence we replace $L^2\left([0,1]^2\right)$ with step functions on a homogenous grid with $m = kl = 150^2 = 22500$ elements. For a step function g we calculate Kg using the Matlab function **imgaussfilt** with $\sigma = 2$. We added uniform noise, e.g. $Y_i := \hat{y} + U_i \mathbb{1}$, where U_i is uniformly distributed on $[-1/2, 1/2]$. We use $n = 20000$ samples. Clearly the true solution $\hat{x}$ has no smoothness at all, that is we can expect only subpolynomial convergence.

6 Basics from probability in Hilbert spaces and regularisation theory

Here we collect the required definitions and results from probability theory and classical regularisation theory.

6.1 Probability theory

We start with the definitions of a probability space, random variables and independency.

Definition 1 (measurable space,probability space) A set Ω together with a sigma algebra $\mathscr{A}$ is called a measurable space $(\Omega,\mathscr{A})$. A measurable space $(\Omega,\mathscr{A})$ together with a probability measure $\mathbb{P}$ on $\mathscr{A}$ is called a probability space $(\Omega,\mathscr{A},\mathbb{P})$. A toplogical space $\mathscr{Y}$ naturally carries the Borel sigma algebra $\mathscr{B}$ generated by the family of open sets in $\mathscr{Y}$.

Definition 2 (random variable) A measurable function $X:\Omega\to\mathscr{Y}$ from a probability space $(\Omega,\mathscr{A},\mathbb{P})$ to a measurable space $(\mathscr{Y},\mathscr{G})$ is called a random variable. A random variable Y naturally induces a probability measure $\mathbb{P}_Y$ on $(\mathscr{Y},\mathscr{G})$ through $\mathbb{P}_{\mathscr{Y}}(G):=\mathbb{P}(Y^{-1}(G))$ for $G\in\mathscr{G}$ (the pushforward). A random variable generates a sub sigma algebra $\{Y^{-1}(G):G\in\mathscr{G}\}$ of $\mathscr{A}$. For a Banach space $\mathscr{Y}$ we call the random variable $Y:\Omega\to\mathscr{Y}$ integrable, if $\int_\Omega \|Y\|d\mathbb{P}<\infty$. For an integrable random variable we define the expectation $\mathbb{E}Y:=\int_\Omega Y d\mathbb{P}$.

Definition 3 (independency) Let I be some index set. The family of sub sigma algebras $(\mathscr{A})_{\alpha\in I}$ on the probability space $(\Omega,\mathscr{A},\mathbb{P})$ is called independent, if for all $J\subset I$ with $|J|<\infty$ it holds that

$$\mathbb{P}(\cap_{j\in J}A_j)=\prod_{j\in J}\mathbb{P}(A_j)\qquad \forall A_j\in\mathscr{A}_j.$$

A family of random variables is called independent, if the family of its generated sigma algebras are independent.

The following theorem assures, that our error model is well defined.

Theorem 6 (i.i.d. sequence of random variables) *Let $(\mathscr{Y},\mathscr{G})$ be a measurable space and μ be a probability measure on $(\mathscr{Y},\mathscr{G})$. Then there exists a probability space $(\Omega,\mathscr{A},\mathbb{P})$ and random variables $Y_i:\Omega\to\mathscr{Y}$ such that the family $(Y_i)_{i\in\mathbb{N}}$ is independent and $\mathbb{P}_{Y_i}=\mu$ for all $i\in\mathbb{N}$.*

Proof Theorem 14.32 from [19]. □

Here we list the common types of convergence for random variables.

Definition 4 (convergence of random variables) Consider random variables $X_n,X:(\Omega,\mathscr{K},\mathbb{P})\to(M,\mathscr{B})$, where the former is a probability space and the latter a separable, metric space (with metric d and the borel sigma algebra $\mathscr{B}$).

1. The sequence $(X_n)_n$ is said to converge $\mathbb{P}-$almost surely to X ($X_n\xrightarrow{\text{a.s.}}X$), if there is a measurable set $\Omega'\subset\Omega$, with $\mathbb{P}(\Omega')=1$ such that for all $\omega\in\Omega'$

$$d(X_n(\omega),X(\omega))\to 0\quad\text{for } n\to\infty. \tag{32}$$

2. The sequence $(X_n)_n$ is said to converge in L^p for $p \in [1,\infty]$ to X ($X_n \xrightarrow{\mathrm{L^p}} X$), if $\mathbb{E}\|X_n\|^p, \mathbb{E}\|X\|^p < \infty$ and
$$(\mathbb{E}\|X_n - X\|^p)^{1/p} \to 0 \quad \text{for } n \to \infty.$$
3. The sequence $(X_n)_n$ is said to converge in probability to X ($X_n \xrightarrow{\mathbb{P}} X$), if for all $\varepsilon > 0$
$$\mathbb{P}(d(X_n, X) > \varepsilon) \to 0 \quad \text{for } n \to \infty.$$
So $X_n \xrightarrow{\mathbb{P}} X$ is equivalent to $d(X_n, X) \xrightarrow{\mathbb{P}} 0$ ($d(X_n, X)$ is measurable).
4. The sequence $(X_n)_n$ is said to converge weakly (or in distribution) to X ($X_n \xrightarrow{\mathrm{w}} X$), if for all continuous and bounded $f : M \to \mathbb{R}$
$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \text{for } n \to \infty.$$

Note that there is no topology inducing almost sure convergence. Almost sure convergence and convergence in L^p each imply convergence in probability, which implies weak convergence.
The following three statements are well known facts for weak convergence, see for example [19] for the proofs.

Lemma 3 (Portementau's leamma) *Suppose we have random variables X_n, X, taking values in a separable Hilbert space. Then the following are equivalent:*

1. $X_n \xrightarrow{\mathrm{w}} X$,
2. $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *for all closed F,*
3. $\liminf_{n\to\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$ *for all open O.*

Proof Theorem 13.1 from [19]. □

Theorem 7 (Continuous mapping theorem) *Assume that $T : \mathscr{X} \to \mathscr{Y}$ is a continuous map between separable Hilbert spaces and $X_n, X : \Omega \to \mathscr{X}$ are random variables with $X_n \xrightarrow{\mathrm{w}} X$. Then it follows that $T(X_n) \xrightarrow{\mathrm{w}} T(X)$.*

Proof Theorem 13.25 from [19]. □

Theorem 8 (Slutsky's theorem) *Assume that $(X_n)_n$ and $(Y_n)_n$ are sequences of random variables on a separable Hilbert space $\mathscr{X}$. If $X_n \xrightarrow{\mathrm{w}} X$ for some random variable $X : \Omega \to \mathscr{X}$ and $Y_n \xrightarrow{\mathbb{P}} c$ for some constant $c \in \mathscr{X}$, it holds that $X_n + Y_n \xrightarrow{\mathrm{w}} X + c$. Moreover, for $\mathscr{X} = \mathbb{R}$, if $\mathbb{P}(Y_n = 0) = 0$, for all $n \in \mathbb{N}$ and $c \neq 0$, then $X_n / Y_n \xrightarrow{\mathrm{w}} X / c$.*

Proof Theorem 13.18 from [19]. □

We conclude with some generalisations of familiar facts for real valued random variables to random variables in a Hilbert space.

Proposition 1 *For independent random variables $X,Y : \Omega \to \mathscr{X}$ on a separable Hilbert space with $\mathbb{E}\|X\|^2, \mathbb{E}\|Y\|^2 < \infty$ it holds that*

$$\mathbb{E}(X,Y) = (\mathbb{E}X, \mathbb{E}Y). \tag{33}$$

Proof Take an orthonormal basis $\{e_l\}_{l\in\mathbb{N}}$, then by dominated convergence

$$\begin{aligned}
\mathbb{E}(X,Y) &= \mathbb{E}\left(\sum_l (X,e_l)e_l, \sum_m (Y,e_m)e_m\right) \\
&= \mathbb{E}\sum_{l,m}(X,e_l)(Y,e_m)(e_l,e_m) \\
&= \sum_l \mathbb{E}\left[(X,e_l)(Y,e_l)\right] = \sum_l \mathbb{E}(X,e_l)\mathbb{E}(Y,e_l) \\
&= \sum_l (\mathbb{E}X,e_l)(\mathbb{E}Y,e_l) \\
&= \sum_{l,m}(\mathbb{E}X,e_l)(\mathbb{E}Y,e_m)(e_l,e_m) \\
&= \left(\sum_l (\mathbb{E}_l,e_l)e_l, \sum_m (EY,e_m)e_m\right) \\
&= (\mathbb{E}X,\mathbb{E}Y).
\end{aligned}$$

□

Corollary 3 *Consider independent random variables X,Y in a separable Hilbert space $\mathscr{X}$ with $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ and a constant $c \in \mathscr{X}$. Then*

1. $\mathbb{E}\|X+Y\|^2 = \mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2$ *and*
2. $\mathbb{E}\|X+c\|^2 = \mathbb{E}\|X\|^2 + \|c\|^2$.

Proof

$$\mathbb{E}\|X+Y\|^2 = \mathbb{E}\|X\|^2 + 2\langle \mathbb{E}X, \mathbb{E}Y\rangle + \mathbb{E}\|Y\|^2.$$

□

Definition 5 A random variabel X on a Banach space $\mathscr{X}$ is called Gaussian, if for all linear and bounded $f : \mathscr{X} \to \mathbb{R}$ it holds that

1. $f(X)$ is measurable,
2. $f(X)$ is a real valued normal variable.

Lemma 4 *Let X be a centred Gaussian with values in a Banach space and $\sigma^2 := \mathbb{E}\|X\|^2$. Denote by Ψ the distribution function of a standard normal variable. Then, for all $t > 0$*

$$\mathbb{P}(\|X\| > t) \le 2(1 - \Psi(t/\sigma)),$$

in particular, $\mathbb{P}(\|X\| \le \sigma) \ge 0.6827$.

Proof See chapter 3.1 of [20]. □

Theorem 9 (Central limit theorem) *Let X be a random variable with values in a separable Hilbert space $\mathscr{X}$, such that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. Then for i.i.d. copies $(X_n)_n$ it holds that*

$$\frac{\sum_{i\leq n} X_i}{\sqrt{n}} \overset{n\to\infty}{\longrightarrow} Z \qquad \textit{weakly}$$

where Z is a centred Gaussian with the same covariance structure as X. That is, for all $x,x' \in \mathscr{X}$ it holds that $\mathbb{E}[(Z,x)(Z,x')] = \mathbb{E}[(X,x)(X,x')]$. Note, that this fully determines the distribution of Z.

Proof Theorem 10.5 in [20]. □

Theorem 10 (Strong law of large numbers) *Assume X to be a random variable with values in a separable Hilbert space $\mathscr{X}$ and X_n to be i.i.d. copies of X. If $\mathbb{E}\|X\| < \infty$, then*

$$\frac{1}{n}\sum_{i\leq n} X_i \overset{\text{a.s.}}{\longrightarrow} \mathbb{E}X \quad \textit{for } n \to \infty.$$

Proof Corollary 7.10 in [20]. □

6.2 Regularisation theory

We use the following standard definitions from deterministic linear regularisation theory.

Definition 6 (Moore-Penrose inverse) For a compact operator $K : \mathscr{X} \to \mathscr{Y}$ between separable Hilbert spaces we define the Moore-Penrose inverse $K^+ : \mathscr{R}(K) \oplus \mathscr{R}(K)^\perp \to \mathscr{X}$ as the unique solution of the normal equation $K^*Kx = K^*y$ in $\mathscr{N}(K)^\perp$. K^+ is linear. K^+ is continuous if and only if the range $\mathscr{R}(K) := \{Kx \in \mathscr{Y} : x \in \mathscr{X}\}$ of K is closed.

Definition 7 (regularisation) A family of linear operators $R_\alpha : \mathscr{Y} \to \mathscr{X}$ is called regularisation of $K^+ : \mathscr{Y} \to \mathscr{X}$ for $\alpha \to 0$, if

1. $R_\alpha \in \mathscr{L}(\mathscr{Y},\mathscr{X})$ for all $\alpha > 0$,
2. $R_\alpha y \overset{\alpha\to 0}{\longrightarrow} K^+ y$ for all $y \in \mathscr{D}(K^+)$.

Definition 8 (parameter choice strategy) A function $\alpha : \mathbb{R}^+ \times \mathscr{Y} \to \mathbb{R}^+$ is called parameter choice strategy. If $\alpha(\delta, y^\delta) = \alpha(\delta)$ for all $\delta > 0$ and $y^\delta \in \mathscr{Y}$, it is called a priori.

Definition 9 (regularisation scheme) A combination of a regularisation and a parameter choice strategy is called regularisation scheme, if for all $y \in \mathscr{D}(K^+)$ and for all $(y^\delta)_{\delta>0} \subset \mathscr{Y}$ with $\|y^\delta - y\| \leq \delta$

$$R_{\alpha(\delta,y^\delta)} y^\delta \overset{\delta\to 0}{\longrightarrow} K^+ y.$$

Definition 10 (singular value decomposition) For a compact operator $K : \mathscr{X} \to \mathscr{Y}$ there exists a monotone sequence $\|K\| = \sigma_1 \geq \sigma_2 \geq ... > 0$ with either $\sigma_i \overset{i\to\infty}{\to} 0$ or there exists a $N \in \mathbb{N}$ with $\sigma_i = \sigma_N$ for all $i \geq N$. Moreover there are families of orthonormal vectors $(u_i)_{i\in\mathbb{N}}$ and $(v_i)_{i\in\mathbb{N}}$ with $span(u_i : i \in \mathbb{N}) = \overline{\mathscr{R}(K)}$, $span(v_i : i \in \mathbb{N}) = \mathscr{N}(K)^\perp$ such that

$$Kv_i = \sigma_i v_i, \quad K^* u_i = \sigma_i v_i.$$

In order to guarantee certain rates of convergence, the true solution $\hat{x}$ has to fullfill a source condition.

Definition 11 (source conditions) For $\nu > 0$ we define $(K^*K)^\nu \in \mathscr{L}(\mathscr{X})$ through

$$(K^*K)^\nu x := \sum_n \sigma_n^{2\nu}(v_n, x)v_n.$$

Write $|K|^\nu := (K^*K)^{\nu/2}$ and define $\mathscr{X}_\nu := \mathscr{R}(|K^\nu|)$ and $\|x\|_\nu := \sum_n \sigma_n^{-2\nu}(x, v_n)^2$.

Many popular linear regularisation methods are of the following class.

6.2.1 Regularisation via filter

Definition 12 (filter) A filter is a family $(F_\alpha)_{\alpha>0}$ of bounded real valued functions on $(0, \|K\|^2]$, with

$$\lim_{\alpha\to 0} F_\alpha(\lambda) = \frac{1}{\lambda} \quad \text{for all} \quad 0 < \lambda \leq \|K\|^2.$$

Definition 13 (regularising filter) A filter $(F_\alpha)_{\alpha>0}$ is called regularising filter, if there is a constant $C_R > 0$ such that

$$\lambda F_\alpha(\lambda) < C_R \quad \text{for all} \quad \alpha > 0,\ \ 0 < \lambda \leq \|K\|^2.$$

Definition 14 (regularisation induced by a regularising filter) A regularising filter $(F_\alpha)_{\alpha>0}$ defines a regularisation via

$$R_\alpha : \mathscr{Y} \to \mathscr{X}, \quad y \mapsto \sum_{i<\infty} F_\alpha(\sigma_i^2)\sigma_i(y, u_i)v_i.$$

In this case we write $R_\alpha := F_\alpha(K^*K)K^*$.

Popular methods of this kind are

1. Tikhonov regularisation, induced by

$$F_\alpha(\lambda) = \frac{1}{\lambda + \alpha},$$

2. Truncated singular value regularisation, induced by

$$F_\alpha(\lambda) = \begin{cases} 0, & \text{for} \quad \lambda < \alpha \\ 1/\lambda, & \text{for} \quad \lambda \geq \alpha \end{cases},$$

3. Landweber iteration, induced by

$$F_\alpha(\lambda) = \frac{1-(1-\omega\lambda)^{\frac{1}{\alpha}}}{\lambda},$$

where the relaxation parameter ω is chosen between 0 and $\frac{1}{\|K\|^2}$.

Definition 15 (monotone filter) A filter $(F_\alpha)_{\alpha>0}$ is called monotone, if

$$F_\alpha(\lambda) \geq F_\beta(\lambda) \quad \text{for all} \quad 0 < \lambda \leq \|K\|^2 \quad \text{and} \quad \alpha \leq \beta.$$

Definition 16 (qualification of a filter) For a filter $(F_\alpha)_{\alpha>0}$, the maximal $\nu_0 > 0$, such that for all $\nu \in (0, \nu_0]$ there exist a constant $C_\nu > 0$ with

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

is called the qualification of the filter (respectively the qualification of the induced regularisation). For example, $\nu_0 = 2$ for the Tikhonov regularisation and $\nu_0 = \infty$ for the Landweber iteration or the truncated singular value regularisation.

We give some properties of regularisations defined by filters which fullfill Assumption 1 or 2.

Proposition 2 *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*

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

Proof

$$\begin{aligned}
\|R_\alpha\| &= \sup_{\|y\|=1} \sqrt{\sum_{i\geq 1} F_\alpha^2(\sigma_i^2)\sigma_i^2 (y,u_i)^2} \\
&\leq \sqrt{\sup_{\sigma_i^2} F_\alpha(\sigma_i^2)\sigma_i^2 \sup_{\sigma_i^2} F_\alpha(\sigma_i^2) \sup_{\|y\|=1} \sum_{i\geq 1} (y,u_i)^2} \leq \sqrt{C_R C_F}/\sqrt{\alpha}
\end{aligned}$$

□

Proposition 3 *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*

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

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

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

Proof We have that $(\hat{y}, u_l) = (K\hat{x}, u_l) = (\hat{x}, K^* u_l) = \sigma_l(\hat{x}, v_l)$, so

$$\begin{aligned} R_\alpha \hat{y} - K^+ \hat{y} &= \sum_{l\geq 1} \left(F_\alpha(\sigma_l^2)\sigma_l - 1/\sigma_l\right)(\hat{y}, u_l) v_l \\ &= \sum_{l\geq 1} \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 (\hat{x}, v_l) v_l \\ &= \sum_{l\geq 1} \sigma_l^\nu \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right) \sigma_l^{-\nu} (\hat{x}, v_l) v_l \end{aligned}$$

For the first statement

$$\begin{aligned} \|R_\alpha \hat{y} - K^+ \hat{y}\| &= \sqrt{\sum_{l\geq 1} \sigma_l^{2\nu} \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 \sigma_l^{-2\nu} (\hat{x}, v_l)^2} \\ &\leq \sup_{0<\lambda\leq\|K\|^2} |F_\alpha(\lambda)\lambda - 1| \lambda^{\nu/2} \|\hat{x}\|_\nu \leq C_\nu \alpha^{\nu/2} \rho \end{aligned}$$

For the second statement set $\xi := \sum_{l\geq 1}(F_\alpha(\sigma_l^2)\sigma_l^2 - 1)\sigma_l^{-\nu}(\hat{x}, v_l)v_l$, so

$$R_\alpha \hat{y} - K^+ \hat{y} = \sum_{l\geq 1} \sigma_l^\nu \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right) \sigma_l^{-\nu} (\hat{x}, v_l) v_l = |K|^\nu \xi.$$

The interpolation inequality (Theorem 2.4.2 in [26]) gives

$$\|K^+ \hat{y} - R_\alpha \hat{y}\| = \||K^\nu|\xi\| \leq \left\||K^{\nu+1}|\xi\right\|^{\frac{\nu}{\nu+1}} \|\xi\|^{\frac{1}{\nu+1}}$$

For the second factor we get

$$\|\xi\| = \sqrt{\sum_{l\geq 1} \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right)^2 \sigma_l^{-2\nu} (\hat{x}, v_l)^2} \leq C\|\hat{x}\|_\nu \leq C\rho.$$

For the first it follows that,

$$\begin{aligned} \||K|^{\nu+1}\xi\| &= \||K|\,(|K|^\nu \xi)\,\| = \sqrt{(|K|(K^+\hat{y} - R_\alpha \hat{y}), |K|(K^+\hat{y} - R_\alpha \hat{y}))} \\ &= \sqrt{(K^*K(K^+\hat{y} - R_\alpha \hat{y}), K^+\hat{y} - R_\alpha \hat{y})} \\ &= \|KK^+\hat{y} - KR_\alpha \hat{y}\| \end{aligned}$$

For the third statement

$$\begin{aligned} KR_\alpha \hat{y} - KK^+ \hat{y} &= K(R_\alpha \hat{y} - K^+ \hat{y}) = \sum_{l\geq 1} \sigma_l^\nu \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right) \sigma_l^{-\nu} (\hat{x}, v_l) K v_l \\ &= \sum_{l\geq 1} \sigma_l^{\nu+1} \left(F_\alpha(\sigma_l^2)\sigma_l^2 - 1\right) \sigma_l^{-\nu} (\hat{x}, v_l) u_l \\ \Longrightarrow \|KR_\alpha \hat{y} - KK^+ \hat{y}\| &\leq \sup_{0<\lambda\leq\|K\|^2} |F_\alpha(\lambda)\lambda - 1| \lambda^{\frac{\nu+1}{2}} \rho = C_{\nu+1} \alpha^{\frac{\nu+1}{2}} \rho. \end{aligned}$$

□

References

1. 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)
2. 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)
3. Bauer, F., Reiß, M.: Regularization independent of the noise level: an analysis of quasi-optimality. Inverse Problems **24**(5), 055009 (2008)
4. Becker, S.: Regularization of statistical inverse problems and the bakushinskiĭ veto. Inverse Problems **27**(11), 115010 (2011)
5. 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)
6. 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)
7. Cavalier, L.: Inverse problems in statistics. In: Inverse problems and high-dimensional estimation, pp. 3–96. Springer (2011)
8. 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)
9. 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)
10. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
11. Gerstner, T., Harrach, B., Roth, D.: Monte carlo pathwise sensitivities for barrier options. arXiv preprint arXiv:1804.03975 (2018)
12. 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)
13. Ghosal, S., Van der Vaart, A.: Fundamentals of nonparametric Bayesian inference, vol. 44. Cambridge University Press (2017)
14. Giné, E., Nickl, R.: Mathematical foundations of infinite-dimensional statistical models, vol. 40. Cambridge University Press (2016)
15. Hanke, M., Scherzer, O.: Inverse problems light: numerical differentiation. The American Mathematical Monthly **108**(6), 512–521 (2001)
16. Hofinger, A.: Ill-posed problems: Extending the deterministic theory to a stochastic setup. Trauner (2006)
17. Hull, J.C., Basu, S.: Options, futures, and other derivatives. Pearson Education India (2016)
18. Kaipio, J., Somersalo, E.: Statistical and computational inverse problems, vol. 160. Springer Science & Business Media (2006)
19. Klenke, A.: Probability theory: a comprehensive course. Springer Science & Business Media (2013)
20. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: isoperimetry and processes. Springer Science & Business Media (2013)
21. Lu, S., Mathé, P.: Discrepancy based model selection in statistical inverse problems. Journal of Complexity **30**(3), 290–308 (2014)
22. 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)
23. Mathé, P., Pereverzev, S.V.: Geometry of linear ill-posed problems in variable hilbert scales. Inverse problems **19**(3), 789 (2003)
24. 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)
25. 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
26. Rieder, A.: Keine Probleme mit inversen Problemen: eine Einführung in ihre stabile Lösung. Springer-Verlag (2013)
27. Tikhonov, A., Arsenin, V.Y.: Methods for solving ill-posed problems. John Wiley and Sons, Inc (1977)
28. 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)

29. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. SIAM Journal on Numerical Analysis **14**(4), 651–667 (1977)
30. 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)