Optimal Regularization for Ill-Posed Problems in Metric Spaces

Frank Bauer, Axel Munk

Institute for Mathematical Stochastics
University of Göttingen
Maschmühlenweg 8-10
37073 Göttingen
Germany
Email: bauer@mathematik.uni-kl.de, munk@math.uni-goettingen.de

Abstract

We present a strategy for choosing the regularization parameter (Lepskij-type balancing principle) for ill-posed problems in metric spaces with deterministic or stochastic noise. Additionally we improve the strategy in comparison to the previously used version for Hilbert spaces in some ways.

AMS-Classification: 47A52, 65J22, 49J35, 93E25

Keywords: Regularization in normed spaces, Lepskij-type balancing principle, l^p-spaces

Supported by the DFG: "Graduiertenkolleg 1023: Identification in Mathematical Models: Synergy of Stochastic and Numerical Methods", University of Göttingen

1 Introduction

In this paper we will be concerned with the numerical solution of linear inverse problems. These are operator equations where the operator is not continuously invertible; mostly A is a linear compact operator $A : \mathcal{X} \to \mathcal{Y}$ acting between two metric spaces,

$$Ax = y^0. \tag{1}$$

Here x has to be recovered form y^0. Moreover, due to measurement errors we can just access noisy data $y^\delta = y^0 + \delta\xi$. The vector ξ will be a standardized random element in the space $\mathcal{Y}$ to be specified later.

Because the inversion of (1) yields an unbounded and hence not continuous operation various regularization techniques have been developed over the last decades, we mention [1], [2] and [3]. Among these are so called spectral methods, as Tikhonov-Phillips regularization, spectral cut-off regularization, Landweber iteration or regularization methods which explicitly penalize certain measures of roughness of the solution of $Ax = y^\delta$, such as BV norms or the number 1 modes [4]. Recently, also non-convex norms such as L^p-norms for $0 < p < 1$ have become popular [5] as well as regularization in Banach spaces [6].

There is an extensive analysis of the convergence of regularization methods in the deterministic context as well as in a setting with random noise ξ (see e.g. [1], [7] and [8]). However, most of these results are formulated in terms of convergence rates depending on a proper choice of the regularization parameter. Unfortunately these results can hardly be utilized in practice due to the unknown information required on x, which determines the choice of the regularization parameter and the noise level. Hence selection of a proper regularization parameter is one of the most challenging tasks when performing a particular regularization scheme in practice.

Among various parameter selection strategies which have been advocated during the past (e.g. Morozov's discrepancy principle [9], generalized cross-validation [10], the L-curve method [11]) we want to emphasize the Lepskii-balancing principle [12] which has been shown to be adaptive in the sense that it adapts automatically within a scale of Hilbert spaces in order to select the optimal regularization parameter in a minimax sense. We mention in particular [13] and [14].

Nevertheless, consistency of the Lepskii principle has been only provided in the context of Hilbert spaces and it remains unclear whether it can also be applied for regularization schemes which habe to be formulated in a more general context as it is required for TV or L^p with $p \neq 2$ penalties.

The aim of this paper is to transfer the Lepskii balancing principle to general metric spaces where we provide a full convergence analysis for l^p spaces, $p > 0$.

The paper is organized as follows: In section 2 we introduce the model, notation and set up. In section 3 we introduce a variant of the Lepskii balancing principle and provide a convergence analysis in a general framework of metric spaces. In section 4 wie discuss the l^p spaces in detail where we restrict ourselves to spectral cut-off regularization and a Gaussian random element ξ.

2 Assumptions

2.1 Preliminaries and Notation

Let $\mathcal{X}$ a metric vector space with distance function $d(\cdot,\cdot)$. Let $\mathcal{Y}$ a topological vector space. Additionally, assume $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ if not stated otherwise.

Throughout this article "$C_{\boxdot}$" will denote some generic constants which may depend from use to use.

2.2 Smoothness and Noise Behavior

Assume that we have a sequence $\{A_n\}_{n\in\mathbb{N}}$ of continuously invertible operators $A_n : \mathcal{X} \to \mathcal{Y}$, the regularization operators. Define according to these the regularized solutions

$$x_n^\delta = A_n^{-1} y^\delta.$$

First of all we assume that the regularization method is consistent, i.e. the regularized solutions converge to the true solution x in the noise free case.

Assumption 1 (Approximation Error)**.** *Let x the solution of (1) and assume that there exists for x a monotonically decreasing continuous function $\psi_x : \mathbb{R} \to \mathbb{R}$ (approximation error function) with $\lim_{n\to\infty} \psi_x(n) = 0$ and the approximation error is bounded as*

$$d(x, x_n^0) \le \psi_x(n). \tag{2}$$

Remark 2. Although the existence of ψ_x is only required for the particular x, in most situations this will be a function ψ independent of x.

In the case of Hilbert spaces (2) depends on the smoothness of x, given by source conditions. In general, the smoother the solution x is with respect to the operator A, the faster decays ψ_x. However, this is also largely influenced by the regularization method measured in terms of the qualification number. A detailed description of this in various contexts can be found in [13],[1] and [8].

The next assumption guarantees that the regularization method actually controls the error in expectation.

Assumption 3 (Stochastic Error Behavior)**.** *Let $k \in \mathbb{N}$. Assume additionally that $\mathbb{E}d(x_n^0, x_n^\delta)^k < \infty$ for all n and that there is a monotonically decreasing continuous function $\rho : \mathbb{R} \to \mathbb{R}$ with $\lim_{n\to\infty} \rho(n) = 0$ and*

$$\mathbb{E}d(x_n^0, x_n^\delta)^k \leq \frac{\delta^k}{\rho(n)^k}. \tag{3}$$

Furthermore assume $\rho(n+1) \geq C_{\text{speed}}\rho(n)$ for some positive real constant C_{speed}.

Remark 4. The property $\rho(n+1) \geq C_{\text{speed}}\rho(n)$ assures that the sequence of regularized solutions is such that subsequent regularized solutions are close enough for our purposes. This is a technical condition required in the proof of Lemma 8.

Definition 5 (Optimal Rate and Regularization Parameter)**.** Define the optimal regularization parameter as

$$n_{\text{opt}} = \min\left\{n : \ \psi_x(n) \leq \frac{\delta}{\rho(n)}\right\}. \tag{4}$$

Define the optimal rate function $\text{OptRate} : \mathbb{R}_+ \to \mathbb{R}_+$

$$\text{OptRate}(\delta) = \psi_x((\psi_x\rho)^{-1}(\delta)). \tag{5}$$

In the sequel we require an exponential bound for $d(x_n^0, x_n^\delta)$.

Assumption 6 (Exponential Bound)**.** *We assume to have positive constants $\tilde{c}$ and l and a function $f(\cdot,\cdot)$, such that for all $N > n_{opt}$ and $\tau > 1$*

$$\mathbb{P}_\xi\left\{\max_{n_{opt}\leq n\leq N} d(x_n^0, x_n^\delta)\rho(n)\delta^{-1} > \tau\right\} \leq \exp\left(-2\tilde{c}\tau^l\right) f(n_{opt}, N). \tag{6}$$

Remark 7. The deterministic case $d(x_n^0, x_n^\delta) \leq \frac{\delta}{\rho(n)}$ is obtained in this stochastic setting as a special case with $f(\cdot,\cdot) = 0$. For Gaussian noise we will compute $f(\cdot,\cdot)$ in Section 4.

3 Rates

3.1 Optimal Rate

Knowing ψ_x and ρ we can choose the regularization parameter a-priorily using the bounds (2) and (3). This is just of theoretical use and cannot be applied in practice. However in the sequel we present a parameter selection rule which achieves an almost optimal rate of convergence without the knowledge of ψ_x and hence is adaptive.

Lemma 8. *Assume (2) and (3). It holds that*

$$\mathbb{E}d(x, x^\delta_{n_{\mathrm{opt}}})^k \leq C_{\mathrm{opt}} \left(\mathrm{OptRate}(\delta)\right)^k$$

and that this bound is rate optimal.

Proof. We have

$$\mathbb{E}d(x, x^\delta_n)^k \leq c_{\mathcal{X},k} \left(\mathbb{E}d(x, x^0_n)^k + \mathbb{E}d(x^0_n, x^\delta_n)^k\right) \leq c_{\mathcal{X},k} \left(\psi(n)^k + \frac{\delta^k}{\rho(n)^k}\right),$$

where ψ_x is a decreasing and $1/\rho$ an increasing real valued function. In order to achieve a rate optimal solution we have to show that

$$\mathbb{E}d(x, x^\delta_{n_{\mathrm{opt}}})^k \leq C_{rate} \min_n \left(\psi(n)^k + \frac{\delta^k}{\rho(n)^k}\right).$$

The intersection point of the functions ψ_x and δ/ρ exists by continuity and monotonicity of ψ_x and ρ. It will be called $(\psi_x \rho)^{-1}(\delta) = n_0$. It holds

$$2\,\mathrm{OptRate}(\delta)^k = \psi(n_0)^k + \frac{\delta^k}{\rho(n_0)^k} \leq 2 \min_n \left(\psi(n)^k + \frac{\delta^k}{\rho(n)^k}\right),$$

and hence it is sufficient to concentrate on a bound in OptRate. On the one hand we have

$$\psi_x(n_{\mathrm{opt}})\rho(n_{\mathrm{opt}}) \leq \delta = \psi_x(n_0)\rho(n_0),$$

and on the other hand

$$\psi_x(n_{\mathrm{opt}} - 1)\rho(n_{\mathrm{opt}} - 1) \geq \delta = \psi_x(n_0)\rho(n_0).$$

This yields

$$\begin{aligned}
\mathbb{E}d(x, x^\delta_{n_{\mathrm{opt}}})^k \leq & c_{\mathcal{X},k} \left(\psi_x(n_{\mathrm{opt}})^k + \frac{\delta^k}{\rho(n_{\mathrm{opt}})}\right) \leq 2c_{\mathcal{X},k} \frac{\delta^k}{\rho(n_{\mathrm{opt}})^k} \\
\leq & 2c_{\mathcal{X},k} \frac{\psi_x(n_0)^k \rho(n_0)^k}{\rho(n_0+1)^k} \leq 2c_{\mathcal{X},k} C^k_{\mathrm{speed}} \psi_x(n_0)^k \\
= & C_{\mathrm{opt}} \left(\mathrm{OptRate}(\delta)\right)^k.
\end{aligned}$$

3.2 Balancing Principle

Now we will introduce an adapted version of the balancing principle which is less computationally demanding than the original version (see e.g. [14]).

This algorithm is an extended version of the algorithm in [13] which now works for general regularization methods. Note that we do not require explicit knowledge of ψ_x.

Definition 9 (Look-Ahead)**.** Let $\sigma > 1$ a positive real number and $N > n_{\text{opt}}$ a real number denoting an upper bound. Define the look ahead function by

$$l_{N,\sigma}(n) = \min\{\min\{m|\rho(n) > \sigma\rho(m)\}, N\}$$

Remark 10. By definition we have that $l_{N,\sigma}(n) > n$ for all $n < N$. The former method [14] will be obtained for setting σ to ∞.

Now we can define the balancing functional $b_{N,\sigma}(n)$:

Definition 11 (Balancing Functional)**.** The balancing functional is defined as

$$b_{N,\sigma}(n) = \max_{n<m\leq l_{N,\sigma}(n)} \left\{4^{-1} d(x_n, x_m)\rho(m)\delta^{-1}\right\}.$$

The smoothed balancing functional is defined as

$$B_{N,\sigma}(n) = \max_{n\leq m\leq N} \{b_{N,\sigma}(n)\}. \tag{7}$$

Remark 12. Note, that $B_{N,\sigma}(n)$ is a monotonically decreasing function.

Definition 13 (Balancing Stopping Index)**.** The balancing stopping index is defined as

$$n_{N,\sigma,\kappa} = \min_{n\leq N} \{B_{N,\sigma}(n) \leq \kappa\}. \tag{8}$$

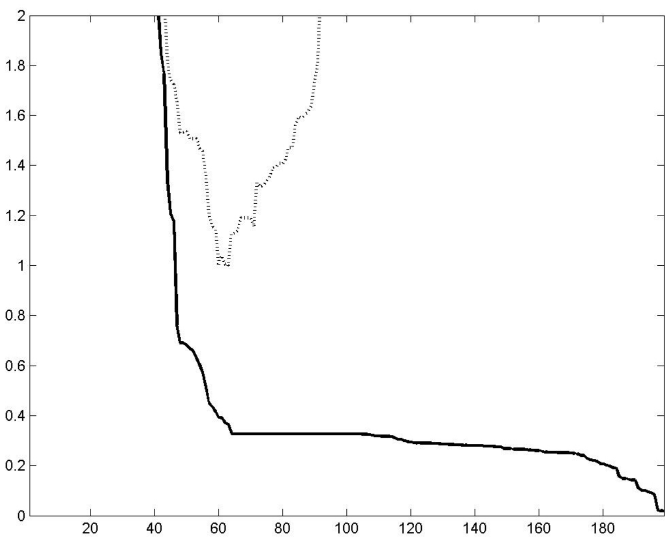

Figure 1: x-axis: cut-off parameter; y-axis: Balancing functional (solid line), $d(x, x_n^\delta)/\min_n d(x, x_n^\delta)$ (dotted line)

In figure 1 we display an example of the balancing functional. In this special case we chose

- $\mathcal{X} = \mathcal{Y} = \mathbb{R}^{200}$ with standard basis $\{u_k\}_{k\in\{1,\dots,200\}}$; $d(\cdot,\cdot)$ standard l^2-norm, i.e. $d(x_1,x_2) = \|x_1 - x_2\|_2$
- $A = \mathrm{diag}(k^{-3})$
- The Fourier coefficients $\langle x, u_k\rangle$ of x are independently drawn according to $\mathcal{N}(0, k^{-5})$.
- The Fourier coefficients $\langle \xi, u_k\rangle$ of the noise ξ are independently drawn according to $\mathcal{N}(0, \delta^2)$; $\delta = 10^{-9}$.

As regularization method we used spectral cut-off. For determining the balancing functional we assumed $\sigma = 3$. In our experience the displayed graph is a good prototype for all balancing functionals observed in practice, more or less independent of the regularization method or the used metric.

Lemma 14 (Balancing Lemma)**.** *Assume (2), (3) and (6). For the balancing stopping index (8) we obtain*

$$\mathbb{E}d(x, x^\delta_{n_{N,\sigma,\kappa}})^k \leq C_{\mathrm{tail}}(\sigma) f(n_{opt}, N) \left(\frac{\delta}{\rho(N)}\right)^k \exp\left(-\tilde{c}\kappa^l\right) + C_{\mathrm{main}}(\sigma)\kappa^k \,\mathrm{OptRate}(\delta)^k \tag{9}$$

Proof. Define the random variable Ξ by

$$\Xi = \max_{n_{opt}\leq n\leq N} d(x^0_n, x^\delta_n)\rho(n)\delta^{-1}$$

and define $\Omega_\kappa = \{\xi : \Xi \leq \kappa\}$ and its complement by $\overline{\Omega_\kappa}$.

We will distinguish two cases

Main Behavior: ($\Xi \leq \kappa$, i.e. $\xi \in \Omega_\kappa$)

For all (n,m) fulfilling $n_{\mathrm{opt}} \leq n \leq m$ we have

$$\begin{aligned} d(x^\delta_n, x^\delta_m) &\leq d(x, x^0_n) + d(x^0_n, x^\delta_n) + d(x, x^0_m) + d(x^0_m, x^\delta_m) \\ &\leq \psi(n) + \frac{\kappa\delta}{\rho(n)} + \psi(m) + \frac{\kappa\delta}{\rho(m)} \leq \frac{4\kappa\delta}{\rho(m)}, \end{aligned}$$

and hence $b_{N,\sigma}(n) \leq \kappa$. This implies in particular that $B_{N,\sigma}(n_{\mathrm{opt}}) \leq \kappa$ and thus $n_{N,\sigma,\kappa} \leq n_{\mathrm{opt}}$ (see (7) and (8)).

Now define $n_0 = n_{N,\sigma,\kappa}$ and $n_{k+1} = l_{N,\sigma}(n_k)$ stopping if $n_K > n_{\mathrm{opt}}$ or $n_K = N$. n_K is now defined as $n_K = n_{\mathrm{opt}}$. Due to the definition of $n_{N,\sigma,\kappa}$ and the monotonicity of $B_{N,\sigma}$ we obtain for all $0 \leq k \leq K$ that we have

$B_{N,\sigma}(n_k) \leq \kappa$. This gives

$$
\begin{aligned}
d(x - x^\delta_{n_{N,\sigma,\kappa}}) &\leq d(x, x^\delta_{n_{\text{opt}}}) + \sum_{k=0}^{K-1} d(x^\delta_{n_k}, x^\delta_{n_{k+1}}) \\
&\leq C_{\text{opt}} \operatorname{OptRate}(\delta) + \sum_{k=0}^{K-1} \frac{4\kappa\delta}{\rho(n_{k+1})} \\
&\leq C_{\text{opt}} \operatorname{OptRate}(\delta) + \frac{4\kappa\delta}{\rho(n_{\text{opt}})} \sum_{k=0}^{K-1} (\sigma^{-1})^{K-1-k} \\
&\leq C_{\text{opt}} \operatorname{OptRate}(\delta) + 4\kappa C_{\text{opt}} \frac{1}{1-\sigma^{-1}} \operatorname{OptRate}(\delta) \\
&\leq (C_{\text{main}}(\sigma))^{1/k} \kappa \operatorname{OptRate}(\delta).
\end{aligned}
$$

Tail Behavior: ($\Xi > \kappa$, i.e. $\xi \in \overline{\Omega_\kappa}$)

Like beforehand we define $n_0 = n_{N,\sigma,\kappa}$ and $n_{k+1} = l_{N,\sigma}(n_k)$ stopping if $n_K = N$. Then we have

$$
\begin{aligned}
d(x, x^\delta_{n_{N,\sigma,\kappa}}) \leq& d(x, x^0_N) + d(x^\delta_N, x^0_N) + \sum_{k=0}^{K-1} d(x^\delta_{n_k}, x^\delta_{n_{k+1}}) \\
\leq& \psi(N) + \frac{\Xi\delta}{\rho(N)} + \frac{4\kappa\delta}{\rho(N)} \frac{1}{1-\sigma^{-1}} \leq 6 \frac{1}{1-\sigma^{-1}} \Xi \frac{\delta}{\rho(N)}.
\end{aligned}
\tag{10}
$$

Using this result we obtain:

$$
\begin{aligned}
\int_{\overline{\Omega_\kappa}} d(x, x^\delta_{n_{N,\sigma,\kappa}})^k d\mathbb{P}(\xi) &\leq \left(3 \frac{1}{1-\sigma^{-1}} \frac{\delta}{\rho(N)}\right)^k \int_{\overline{\Omega_\kappa}} \Xi^k d\mathbb{P}(\xi) \\
&\leq \left(3 \frac{1}{1-\sigma^{-1}} \frac{\delta}{\rho(N)}\right)^k \left(\int_{\overline{\Omega_\kappa}} \Xi^{2k} d\mathbb{P}(\xi) \int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}(\xi)\right)^{1/2}
\end{aligned}
\tag{11}
$$

Now we estimate the two parts separately.

$$
\begin{aligned}
\int_{\overline{\Omega_\kappa}} \Xi^{2k} d\mathbb{P}(\xi) &\leq - \int_\kappa^\infty \tau^{2k} d\left(\exp\left(-2\tilde{c}\tau^l\right) f(n_{opt}, N)\right) \\
&\leq - \int_0^\infty \tau^{2k} d\left(\exp\left(-2\tilde{c}\tau^l\right) f(n_{opt}, N)\right) \\
&= -\tau^{2k} \left(\exp\left(-2\tilde{c}\tau^l\right) f(n_{opt}, N)\right)\Big|_0^\infty \\
&\quad + 2k \int_0^\infty \tau^{2k-1} \left(\exp\left(-2\tilde{c}\tau^l\right) f(n_{opt}, N)\right) d\tau \\
&= 2k f(n_{opt}, N) \int_0^\infty \tau^{2k-1} \left(\exp\left(-2\tilde{c}\tau^l\right)\right) d\tau \\
&\leq C_l f(n_{opt}, N),
\end{aligned}
$$

and the second factor in (11) is estimated as

$$\int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}(\xi) \leq \exp\left(-2\tilde{c}\kappa^l\right) f(n_{opt}, N).$$

Hence we get

$$\int_{\overline{\Omega_\kappa}} \Xi^k d\mathbb{P}(\xi) \leq C_{\text{tail}}(\sigma) f(n_{opt}, N) \left(\frac{\delta}{\rho(N)}\right)^k \exp\left(-\tilde{c}\kappa^l\right).$$

This yields

$$\begin{aligned} \mathbb{E}d(x, x^\delta_{n_{N,\sigma,\kappa}})^k \leq{}& C_{\text{tail}}(\sigma) f(n_{opt}, N) \left(\frac{\delta}{\rho(N)}\right)^k \exp\left(-\tilde{c}\kappa^l\right) \\ &+ C_{\text{main}}(\sigma)\kappa^k \operatorname{OptRate}(\delta)^k. \end{aligned}$$

Theorem 15. *Assume (2), (3) and (6). There is a δ_0 such that for all $\delta < \delta_0$ we can choose $n_{opt} \leq N = \rho^{-1}(\delta)$ and have*

$$\mathbb{E}d(x, x^\delta_{n_{N,\sigma,\kappa}})^k \leq C_{\text{tail}}(\sigma) f(n_{opt}, N) \exp\left(-\tilde{c}\kappa^l\right) + C_{\text{main}}(\sigma)\kappa^k \operatorname{OptRate}(\delta)^k.$$

Proof. As A is compact $\delta / \operatorname{OptRate}(\delta) \to 0$ as $\delta \to 0$. Hence due to $\rho(n+1) \geq C_{\text{speed}}\rho(n)$ there is a δ_0 s.t. for all $\delta < \delta_0$ it always holds $N = \rho^{-1}(\delta) > n_{\text{opt}}$.

The inequality follows by insertion in (9).

Remark 16. When setting the look-ahead parameter σ to infinity all proofs also hold provided for all $x_1, x_2, x \in \mathcal{X}$

$$d(x_1, x_2)^k \leq c_{\mathcal{X},k}\left(d(x_1, x)^k + d(x_2, x)^k\right). \tag{12}$$

for some constant $c_{\mathcal{X},k}$. This includes l^p with $0 < p < 1$ with $d(\cdot,\cdot)^p = ||\cdot - \cdot||_p^p$.

4 The Lepskij Balancing Principle in l^p spaces

In the following section we will restrict ourselves to a special situation. We will assume that we are operating in separable Hilbert spaces $\mathcal{X}$, $\mathcal{Y}$ and (see e.g. [15]) and hence we have a singular value decomposition of the linear compact operator $A : \mathcal{X} \to \mathcal{Y}$ with the corresponding basis $\{u_k\}_{k\in\mathbb{N}}$ and $\{v_k\}_{k\in\mathbb{N}}$ and the singular values $\{\lambda_k\}_{k\in\mathbb{N}}$ where the λ_k are forming a monotonically decreasing sequence tending to zero, where It holds that $Ax = \sum_{k=1}^\infty \lambda_k \langle x, u_k\rangle v_k$.

4.1 Regularization Operators

In the sequel we will restrict our analysis to the spectral cut-off operators A_n defined by

$$A_n x = \sum_{k=1}^{n} \lambda_k \langle x, u_k \rangle v_k.$$

4.2 Noise Model

We assume that $y^\delta = y^0 + \delta\xi$ where ξ is a zero-mean weak Gaussian random element (see e.g. [16]). This specifically means that for every element $g \in \mathcal{Y}$ we have $\langle g, y^\delta \rangle = \langle g, Ax \rangle + \delta \langle g, \xi \rangle$, where $\langle g, \xi \rangle$ is a centered Gaussian random variable on a probability space $(\Omega, \Sigma, \mathbb{P}_\xi)$ with variance $||g||^2$. Hence we have in addition $\mathbb{E}\left(\langle g_1, \xi \rangle \langle g_2, \xi \rangle\right) = \langle g_1, g_2 \rangle$ for all $g_1, g_2 \in \mathcal{Y}$, i.e. the white noise element ξ is generated by a stochastic process with the identity covariance operator. In the sequel we will denote the $\mathcal{N}(0,1)$ random variables $\langle \xi, u_k \rangle$ by R_k.

4.3 Norms and Expectations

As normed spaces we will now consider the subspace $l^p \cap l^2$ of the previously defined Hilbert space l^2 equipped with the p-norm $(1 \le p \le \infty)$

$$||x||_p = \left(\sum_{k=1}^{\infty} |\langle x, u_k \rangle|^p \right)^{\frac{1}{p}}.$$

Note that all considerations will hold for l^p and not just for $l^p \cap l^2$.

Straightforward calculation shows

$$\mathbb{E}||x_n^0 - x_n^\delta||_p^p = \delta^p C_p \sum_{k=1}^{n} \left|\lambda_k^{-1}\right|^p, \tag{13}$$

where $C_p = 2^{\frac{p}{2}} \Gamma(\frac{1+p}{2}) / \Gamma(\frac{1}{2})$.

4.4 Probabilities

Assume that Z_k are i.i.d. random variables with distribution $\mathcal{N}(0,1)$. Now we define

$$S_n = \sum_{k=1}^{n} |\lambda_k^{-1}|^p |Z_k|^p$$

Furthermore, let

$$S_n^* = \frac{S_n}{\mathbb{E} S_n} = C_p^{-1} \sum_{k=1}^{n} |Z_k|^p \frac{|\lambda_k^{-1}|^p}{\sum_{j=1}^{n} |\lambda_j^{-1}|^p} =: C_p^{-1} \sum_{k=1}^{n} |Z_k|^p \alpha_{n,k}^p. \tag{14}$$

This implies that $\alpha_{n,k} > 0$ and $\sum_{k=1}^{n} \alpha_{n,k}^p = 1$, and because of the monotonicity of the λ_k, additionally $\max_{k=1,\dots,n} \alpha_{n,k} = \alpha_{n,n}$.

First we need two auxiliary lemmas

Lemma 17. *Let a_k positive real numbers and $W_k = a_k Z_k$ independent random variables where the Z_k are $\mathcal{N}(0,1)$ distributed. Then*

$$\mathbb{E}\left\{\max_{k=1,\dots,n} |W_k|\right\} \le \sqrt{2\ln 2n} \max_{k=1,\dots,n} |a_k|. \tag{15}$$

The proof follows the lines of [17] and is therefore omitted.

Lemma 18 (Borel's inequality, see e.g. [18])**.** *Let a_k positive real numbers and $W_k = a_k Z_k$ independent random variables where the Z_k are $\mathcal{N}(0,1)$ distributed. Then*

$$\mathbb{P}\left\{\max_{k=1,\dots,n} |W_k| - \mathbb{E}\left\{\max_{k=1,\dots,n} |W_k|\right\} > \tau\right\} \le \exp\left(-\frac{\tau^2}{2\max_{k=1,\dots,n} a_k^2}\right). \tag{16}$$

Lemma 19. *It holds for $C_p^{-1/p}\tau \ge 1 + \sqrt{2\ln 2N}$ and $N \ge 1$*

$$\mathbb{P}\left\{\max_{n_{opt}\le n\le N} S_n^* > \tau^p\right\} \le \exp\left(-2\frac{1}{2C_p^{1/p}}\tau\right) N^2.$$

Proof. Define

$$\overline{\tau}^p = C_p^{-1}\tau^p.$$

One has, using (16), (15) and $\sum_{k=1}^{n} \alpha_{n,k}^p = 1$,

$$\begin{aligned}
\mathbb{P}\left\{\max_{n_{opt}\le n\le N} S_n^* > \tau^p\right\} =&\mathbb{P}\left\{\max_{n_{opt}\le n\le N} \sum_{k=1}^{n} |Z_k|^p \alpha_{n,k}^p > \overline{\tau}^p\right\}\\
\le&\mathbb{P}\left\{\max_{n_{opt}\le n\le N} \max_{k=1,\dots,n} |Z_k|^p > \overline{\tau}^p\right\}\\
=&\mathbb{P}\left\{\max_{k=1,\dots,N} |Z_k| > \overline{\tau}\right\}\\
=&\mathbb{P}\left\{\max_{k=1,\dots,N} |Z_k| - \mathbb{E}\left(\max_{k=1,\dots,N} |Z_k|\right) > \overline{\tau} - \mathbb{E}\left(\max_{k=1,\dots,N} |Z_k|\right)\right\}\\
\stackrel{(16)}{\le}& \exp\left(-\frac{\left(\overline{\tau} - \mathbb{E}\left(\max_{k=1,\dots,N} |Z_k|\right)\right)^2}{2}\right)\\
\stackrel{(15)}{\le}& \exp\left(-\frac{\left(\overline{\tau} - \sqrt{2\ln 2N}\right)^2}{2}\right)\\
\le& \exp\left(-2\frac{1}{2C_p^{1/p}}\tau\right) N^2.
\end{aligned}$$

This yields our main theorem:

Theorem 20. *Assume that we have an inverse problem in the space l^p fulfilling assumption (2) with Gaussian white noise. Furthermore assume that it holds $\kappa = 2C_k^{1/k}\sqrt{\ln 2N}$ for $N = \rho^{-1}(\delta)$. Then it holds for a fixed constant μ*

$$\mathbb{E}d(x, x^\delta_{n_{N,\sigma,\kappa}})_p^p \leq C_{\text{all}}(\sigma)\left(\ln \mathrm{OptRate}(\delta)^{-1}\right)^k \mathrm{OptRate}(\delta)^k.$$

Proof. Due to $C_p^{-1/p}\kappa = 2\sqrt{\ln 2N} \geq 1 + \sqrt{2\ln 2N}$ we can apply the last lemma and so it holds (3) and (6). Hence the requirements for Lemma 19 hold.

Using the same arguments as in [14] we obtain the above result by Theorem 15.

References

[1] H. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.

[2] A. Tikhonov, V. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.

[3] B. Hofmann, Regularization of Applied Inverse and Ill-Posed Problems, Teubner, Leipzig, 1986.

[4] O. Scherzer, Taut-string algorithm and regularization programs with G-norm data fit, J. Math. Imaging Vision 23 (2) (2005) 135–143.

[5] M. Nikolova, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares, Multiscale Model. Simul. 4 (3) (2005) 960–991 (electronic).

[6] F. Schöpfer, A. K. Louis, T. Schuster, Nonlinear iterative methods for linear ill-posed problems in Banach spaces, Inverse Problems 22 (1) (2006) 311–329.

[7] B. A. Mair, F. H. Ruymgaart, Statistical inverse estimation in Hilbert scales, SIAM J. Appl. Math. 56 (5) (1996) 1424–1444.

[8] N. Bissantz, T. Hohage, A. Munk, Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise, Inverse Problems 20 (2004) 1773–1791.

[9] V. A. Morozov, Methods for solving incorrectly posed problems, Springer-Verlag, New York, 1984.

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

[11] P. Hansen, Rank-deficient and discrete ill-posed problems, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998, numerical aspects of linear inversion.

[12] O. Lepskiĭ, On a problem of adaptive estimation in Gaussian white noise., Theory of Probability and its Applications 35 (3) (1990) 454–466.

[13] P. Mathé, S. Pereverzev, Geometry of linear ill-posed problems in variable hilbert spaces, Inverse Problems 19 (3) (2003) 789–803.

[14] F. Bauer, S. Pereverzev, Regularization without preliminary knowledge of smoothness and error behavior, European J. Appl. Math. 16 (3) (2005) 303 – 317.

[15] W. Rudin, Functional Analysis, McGraw-Hill Series in Higher Mathematics, McGraw-Hill Book Company, New York, 1973.

[16] A. V. Skorokhod, Integration in Hilbert space, Springer, Berlin, 1974.

[17] L. Devroye, G. Lugosi, Combinatorial Methods in Density Estimation, Springer Series in Statistics, Springer, New York, 2001.

[18] A. van der Vaart, J. Wellner, Weak Convergence and Empirical Processes. With Applications to Statistics, Springer Series in Statistics, Springer, New York, 1996.