Assessing Uncertainty in Linear Inverse Problems with the Metrics of Ky Fan and Prokhorov

Andreas Hofinger*

August 11, 2006

Abstract

A convergence rate analysis is an important tool to assess the quality of regularization methods for inverse problems.

In this work, we show that results that were previously obtained for the Prokhorov metric and Tikhonov regularization can be extended to the metric of Ky Fan and to more general regularization methods.

1 Introduction

In the following, we consider a stochastic linear inverse problem by assuming that the operator and the right-hand side of the original problem $Ax = y$ are influenced by some random parameter ω, element of a probability space $(\Omega, \mathcal{A}, \mu)$. I.e., for fixed ω, we consider the equation

$$A(\omega)x(\omega) = y(\omega), \qquad \omega \in \Omega. \tag{1}$$

In [9], we derived convergence results for this problem in terms of the Prokhorov metric. In this work, we extend [9] in two ways:

On the one hand, we are going to consider more general regularization methods in place of Tikhonov regularization. Via spectral theory, many regularization methods can be treated at once, in particular the new results apply to Landweber iteration as well (sections 4 and 5).

*Spezialforschungsbereich SFB F013/F1308, Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, A-4040 Linz, Austria email: andreas.hofinger@oeaw.ac.at

On the other hand, we derive analogous results as in [9] for a different metric. Let us first of all recall the setup of [9]: Suppose we are given two random variables $x(\omega)$ and $\tilde{x}(\omega) \in X$ for $\omega \in \Omega$ and probability space $(\Omega, \mathcal{A}, \mu)$. Via the measure μ on Ω we can define two corresponding measures μ_x and $\mu_{\tilde{x}}$ (the so-called *distributions* of x and $\tilde{x}$) on the space X: For a Borel set $B \subset X$ we define

$$\mu_x(B) := \mu(x^{-1}(B)) := \mu\{\omega \in \Omega \mid x(\omega) \in B\}\,,$$

and $\mu_{\tilde{x}}(B)$ accordingly. Instead of measuring the pointwise distance of $x(\omega)$ and $\tilde{x}(\omega)$ directly, we can use this lifting onto spaces of probability measures and compute the distance of the respective measures there, using an appropriate metric. In [9, 10], the *Prokhorov metric* was used to determine convergence and convergence rates for inverse problems.

Definition 1.1 (Prokhorov metric)**.** *The distance of two measures μ_x, $\mu_{\tilde{x}}$ in the Prokhorov metric is defined as (see, e. g., [4, 6, 16, 20])*

$$\rho_{\text{P}}(\mu_x,\, \mu_{\tilde{x}}) := \inf\{\varepsilon > 0 \mid \mu_x(B) \le \mu_{\tilde{x}}(B^\varepsilon) + \varepsilon,\, \forall\, \textit{Borel sets}\, B \subset \Omega\}\,. \tag{2}$$

Here, $B^\varepsilon = \{x \mid d(x, B) < \varepsilon\}$, where $d(x, B)$ is the distance of x to B, i. e., $d(x, B) = \inf_{z \in B} \|x - z\|$.

In this work, in addition to the Prokhorov metric, we also consider a closely related measure of distance—the *Ky Fan metric*. While the Prokhorov metric works with distributions, the Ky Fan metric uses random variables to define distances. This metric is defined as follows.

Definition 1.2 (Ky Fan metric)**.** *The distance of two random variables x, $\tilde{x}$ in the Ky Fan metric is defined as ([11], see also [6])*

$$\rho_{\text{K}}(x,\, \tilde{x}) := \inf\{\varepsilon > 0 \mid \mu\{\omega \in \Omega \mid d(x(\omega), \tilde{x}(\omega)) > \varepsilon\} < \varepsilon\}\,. \tag{3}$$

It can be shown that for random variables x, $\tilde{x}$ and corresponding distributions μ_x, $\mu_{\tilde{x}}$ always

$$\rho_{\text{P}}(\mu_x,\, \mu_{\tilde{x}}) \le \rho_{\text{K}}(x,\, \tilde{x}). \tag{4}$$

Moreover, for given measures μ_x, $\mu_{\tilde{x}}$ one can always find appropriate random variables x, $\tilde{x}$ such that equality holds in (4) (see [5, 6, 21]). For some additional background and a detailed comparison of these metrics with other notions of convergence (in particular, convergence in expectation) see, e. g., [14].

The outline of this paper is as follows. In the next section, we show that any regularization method that leads to convergence in the Ky Fan or Prokhorov metric must be using information about the noise level in the respective metric. This is the equivalent of Bakushinsky's result on non-convergence of error-free methods ([1], cf. [8]). When a deterministic operator is lifted to an operator between probability spaces, Hölder-continuity estimates translate to new bounds. Such bounds are derived in section 3 and used in section 4, the main part of this work, to derive results on convergence and convergence rates. As in [9], *stochastic source conditions* play an important role in obtaining these convergence results. Finally, we show in section 5 how the derived estimates simplify, when they are applied to Landweber's method.

2 Bakushinsky's result on error-free methods

Since solving (1) is ill-posed, we need to apply a regularization technique to obtain a stable solution when the data are noisy. Let us denote such a method by R_α. In general, in a regularization method one replaces (1) by a family of neighbouring well-posed problems (cf. [8, Ch.3]). Via the regularization parameter α, one can vary between problems that approximate (1) well, but are therefore also unstable, and problems that are poor approximations of (1), but are less susceptible to noise. But how should this parameter α be chosen for an optimal compromise between accuracy and stability?

If we ask for convergence of x_α^δ to $x^\dagger$ when y^δ converges to y, certainly the parameter choice α must somehow depend on the data y^δ and/or the noise level δ. Methods that depend on δ only are called *a-priori*, those that depend on δ and y^δ, *a-posteriori* parameter choice rules. Methods that use y^δ only, but no information about δ, are called *error-free*. An example for such an error-free method is the L-curve method (see Remark 2.2). However, a classical result of Bakushinsky [1] states that a convergent parameter choice rule must be using information about the noise level δ. Otherwise $x_\alpha^\delta = R_\alpha(A, y^\delta)$ can converge to $x^\dagger$ only for operators with bounded inverse, i. e., in cases when solving (1) is well-posed. For the L-curve method, non-convergence was explicitly demonstrated in [7] and [24].

In the following, we show that Bakushinsky's result on non-convergence of error-free methods remains valid also on the space of probability measures, i. e., we show that a regularization method of the form $R_{\alpha(y^\delta)}(A, y^\delta)$ can lead to convergence only when the operator A in (1) has a bounded inverse.

Theorem 2.1 (Non-Convergence of Error-Free Methods)**.** *Suppose the regularization method $R_{\alpha(y^\delta)}(A, y^\delta)$ does not utilize information about the noise level δ, measured in the Ky Fan metric. Suppose further, that this method leads to convergence, i. e., for all y, $y(\omega) \in \mathcal{D}(A(\omega))$ for almost all ω, and* *noisy data y^δ with $\rho_{\text{K}}(y,\, y^\delta) \le \delta$ we have*

$$\rho_{\text{K}}\left(R_{\alpha(y^\delta)}(A, y^\delta),\, A^\dagger y\right) \to 0 \quad \textit{for} \quad \delta \to 0.$$

Then $A^\dagger(\omega)$ is bounded for almost all ω. The analogous statement holds for the Prokhorov metric.

Proof. By assumption, the method converges for arbitrary measurements $y^\delta(\omega)$ with $\rho_{\text{K}}(y,\, y^\delta) \to 0$. Now choose $y^\delta(\omega) = y(\omega) \in \mathcal{D}(A^\dagger(\omega))$ for almost all ω. Since the method is convergent, we obtain

$$\rho_{\text{K}}(R_{\alpha(y)}(A, y),\, A^\dagger y) = 0\,. \tag{5}$$

Let $y_n(\omega) \in \mathcal{D}(A^\dagger(\omega))$ for almost all ω. Then by (5) and the triangle inequality we can estimate

$$\rho_{\text{K}}(A^\dagger y_n,\, A^\dagger y) \le \rho_{\text{K}}(R_{\alpha(y_n)}(A, y_n),\, A^\dagger y). \tag{6}$$

The analogous bound also holds for the Prokhorov metric. This estimate is valid for arbitrary $y(\omega) \in \mathcal{D}(A^\dagger(\omega))$, in particular also for the element $y(\omega) = 0$. With this choice the Prokhorov distance and the Ky Fan distance are equal[1]. Therefore the following argument also holds for the Prokhorov metric.

Let Ω_u denote the subset of Ω on which $A^\dagger(\omega)$ is unbounded. For $\omega \in \Omega_u$ and arbitrary $n \in \mathbb{N}$ we can find $z_n(\omega) \in \mathcal{D}(A^\dagger(\omega))$ with

$$\|z_n(\omega)\| \le 1, \quad \text{and} \quad \left\|A^\dagger(\omega) z_n(\omega)\right\| \ge n\,.$$

Define y_n via $y_n(\omega) := z_n(\omega)/\sqrt{n}$ for $\omega \in \Omega_u$, and $y_n(\omega) = 0$ elsewhere. Clearly, $y_n \to 0$ almost surely, since $\sup_\omega \|y_n(\omega)\| \le 1/\sqrt{n}$. Convergence of the regularization method implies via (6) existence of a sequence ε_n, $\varepsilon_n \to 0$, with

$$\rho_{\text{K}}(A^\dagger y_n,\, 0) < \varepsilon_n\,.$$

Let n be sufficiently large to have $\varepsilon_n \le 1$. Then we can estimate

$$\begin{aligned}
\mu\left(\Omega_u\right) &= \mu\{\omega \in \Omega \mid \left\|A^\dagger(\omega) y_n(\omega) - A^\dagger(\omega) 0\right\| \ge \sqrt{n}\} \\
&\le \mu\{\omega \in \Omega \mid \left\|A^\dagger(\omega) y_n(\omega) - A^\dagger(\omega) 0\right\| \ge \varepsilon_n\} \le \varepsilon_n\,.
\end{aligned}$$

So the set Ω_u where $A^\dagger(\omega)$ is unbounded (on its domain), is a null set.

[1] see e. g., [13, 14].

It remains to be shown that $A^\dagger(\omega)$ is not only continuous almost surely on its domain, but that this domain also coincides with the whole space Y, and that therefore $A^\dagger(\omega)$ is continuous on Y almost surely. Fix $\omega \in \Omega \setminus \Omega_u$. Since $A^\dagger(\omega)$ is continuous on $\mathcal{D}(A^\dagger(\omega))$, it can be uniquely extended to a linear continuous operator $\overline{A^\dagger(\omega)}$ on $\overline{\mathcal{D}(A^\dagger(\omega))}$.

The Moore-Penrose inverse satisfies $A(\omega)A^\dagger(\omega) = Q(\omega)|_{\mathcal{R}(A(\omega))}$, where $Q(\omega)$ is the orthogonal projector from Y onto $\overline{\mathcal{R}(A(\omega))}$ (see [8]). The operator $A(\omega)\overline{A^\dagger(\omega)}$ defines a linear continuous extension of $Q(\omega)|_{\mathcal{R}(A(\omega))}$ to $\overline{\mathcal{R}(A(\omega))}$. Since $Q(\omega)$ itself is continuous, we find

$$A(\omega)\overline{A^\dagger(\omega)} = Q(\omega)\,.$$

The left hand side of this equality clearly maps to $\mathcal{R}(A(\omega))$, while the right-hand side maps (by definition) *onto* the closure $\overline{\mathcal{R}(A(\omega))}$. Thus the two spaces must be equal, i. e., $\mathcal{R}(A(\omega))$ must be closed.

Hence we found that for arbitrary $\omega \in \Omega \setminus \Omega_u$, $A^\dagger(\omega)$ is bounded on $\mathcal{D}(A^\dagger(\omega))$ as well as $\mathcal{D}(A^\dagger(\omega)) = \mathcal{R}(A(\omega)) + \mathcal{R}(A(\omega))^\perp = Y$, so $A^\dagger(\omega)$ is bounded on Y for almost all ω. $\square$

Remark 2.2. Typical convergent parameter choice rules are *a-priori* rules such as

$$\alpha \sim \delta,$$

which gives convergence for Tikhonov regularization in linear and nonlinear problems[2] or *a-posteriori* rules such as, e. g., Morozov's discrepancy principle ([19]). Besides these convergent parameter choice rules, there are also other methods in use, one popular being the L-curve method. There, a plot of $\|x_\alpha^\delta\|$ vs. $\|F(x_\alpha^\delta) - y^\delta\|$ is generated by picking several values of α. Often the corresponding graphs look like the letter "L". Hansen proposed to choose the regularization parameter α corresponding to the corner of this "L" (see, e. g., [12]). The classical result due to Bakushinsky [1] already shows that this method cannot yield convergence for noise level tending to 0 in the deterministic setup. (Although for finite noise levels, the choice may still yield good results.) Now using a stochastic notion of convergence, one could imagine that error-free methods might converge, but as Theorem 2.1 shows, this is neither the case for the Ky Fan metric, nor for the Prokhorov metric. Also in these concepts, knowledge about the noise is necessary. Nevertheless, the required knowledge is now considerably weaker than before, since it is phrased in probabilistic terms, and not in terms of the Hilbert-space norm, i. e., in a worst case sense. In particular, it can be of the form

[2] See e. g., Corollary 4.4 and Theorem 10.4 in [8]. Confer also Theorem 4.4 and Remark 5.2.

"with 95% probability we have $\|y(\omega) - y^\delta(\omega)\| \le \delta$"

which is not possible in a deterministic setting, and which also need not imply a bound in expectation (see examples in [13, 14]).

3 Lifting of Locally Hölder-continuous Maps

An important tool to obtain the convergence results in [9, 10] was a lifting technique. In the following we show that an analogous lifting result as [9, Thm. 2.1] also holds for the Ky Fan metric. This allows to use similar proof techniques as in [9] to obtain convergence results in sections 4.1 and 4.2.

For $F : X \to Y$, we define the lifting $\tilde{F}$ to the spaces of probability measures on metric spaces X and Y, equipped with the metrics d and $\tilde{d}$, respectively, in the following way: For a probability measure μ_x on X, $\tilde{F}(\mu_x)$ defines a probability measure on Y via

$$\tilde{F}(\mu_x)(B) := \mu_x\left(F^{-1}(B)\right) = \mu_x\{x \mid F(x) \in B\} \quad (B \subset Y, \text{ Borel set})\,.$$

In the following let F be locally Hölder-continuous with Hölder exponent $0 < \gamma \le 1$, where the Hölder-constant is controlled by a monotonically increasing, right continuous function $h : \mathbb{R}^+ \to \mathbb{R}^+$,

$$\tilde{d}(F(z), F(\tilde{z})) \le h\big(\max\{d(z,0), d(\tilde{z},0)\}\big)\, d(z,\tilde{z})^\gamma \quad (z, \tilde{z} \in X\,). \tag{7}$$

In the lifting we want to obtain an analogous estimate for $\tilde{F}$, with z, $\tilde{z} \in X$ replaced by some corresponding measures μ_x and $\mu_{\tilde{x}}$ on X, respectively.

Since the Hölder-constant depends on the size of z and $\tilde{z}$, it is reasonable that for lifting such an estimate to probability measures, we have to introduce some kind of balancing, which ensures that the probability of large elements is small. This is a requirement on the *tail behavior* of the involved measures. We assume that the measure $\mu_{\tilde{x}}$ fulfills, with some monotonically decreasing, right continuous function κ ($\kappa(\vartheta) \to 0$ for $\vartheta \to \infty$), the following decay condition:

$$\mu_{\tilde{x}}(\mathcal{C}\{\mathcal{B}_\vartheta(z_0)\}) \le \kappa(\vartheta) \quad (\vartheta > 0)\,. \tag{8}$$

Here $\mathcal{B}_\vartheta(z_0)$ denotes a ball with radius ϑ around some specific element z_0; by $\mathcal{C}\{B\}$ we denote the complement of a set B, i. e., for $B \subset X$, $\mathcal{C}\{B\} := \{x \in X \mid x \notin B\}$. Using this decay condition we can later split the domain X into one part where the Hölder-constant is uniformly bounded, and another part where this constant may be large, but which has low probability.

For the Prokhorov metric the following theorem was obtained in [9] as an extension of Theorem 1 in [10]. We now prove that an analogous result is valid for the Ky Fan metric. The theorem shows that the lifted operator $\tilde{F}$ fulfills a Hölder-condition with the same exponent as the original operator F.

Theorem 3.1. *Let x, $\tilde{x}$ be random variables with distributions μ_x, $\mu_{\tilde{x}}$ respectively. Let assumptions* (7) *and* (8) *hold. Then with $\rho_{\mu_x,\mu_{\tilde{x}}} := \rho_{\text{P}}(\mu_x,\, \mu_{\tilde{x}})$ and $\rho_{x,\tilde{x}} := \rho_{\text{K}}(x,\, \tilde{x})$*

$$\begin{aligned} &\rho_{\text{P}}(\tilde{F}(\mu_x),\, \tilde{F}(\mu_{\tilde{x}})) \leq \\ &\qquad \leq \inf_{\vartheta>0} \max\{\rho^{\gamma}_{\mu_x,\mu_{\tilde{x}}} h\big(d(z_0,0)+\vartheta+\rho_{\mu_x,\mu_{\tilde{x}}}\big), \rho_{\mu_x,\mu_{\tilde{x}}}+\kappa(\vartheta)\}\,. \\ &\rho_{\text{K}}(F(x),\, F(\tilde{x})) \leq \\ &\qquad \leq \inf_{\vartheta>0} \max\{\rho^{\gamma}_{x,\tilde{x}} h\big(d(z_0,0)+\vartheta+\rho_{x,\tilde{x}}\big), \rho_{x,\tilde{x}}+\kappa(\vartheta)\}\,. \end{aligned} \tag{9}$$

Especially, with $\kappa^{-1}(\tau) := \inf\{\vartheta \geq 0 \mid \kappa(\vartheta) \leq \tau\}$ we obtain (for $\rho_{\mu_x,\mu_{\tilde{x}}} \neq 0$ and $\rho_{x,\tilde{x}} \neq 0$ respectively) the Hölder-estimates

$$\begin{aligned} &\rho_{\text{P}}(\tilde{F}(\mu_x),\, \tilde{F}(\mu_{\tilde{x}})) \leq \\ &\qquad \leq \max\{h\left(d(z_0,0)+\kappa^{-1}(\rho_{\mu_x,\mu_{\tilde{x}}})+\rho_{\mu_x,\mu_{\tilde{x}}}\right), 2\,\rho^{1-\gamma}_{\mu_x,\mu_{\tilde{x}}}\}\rho^{\gamma}_{\mu_x,\mu_{\tilde{x}}} \\ &\rho_{\text{K}}(F(x),\, F(\tilde{x})) \leq \\ &\qquad \leq \max\{h\left(d(z_0,0)+\kappa^{-1}(\rho_{x,\tilde{x}})+\rho_{x,\tilde{x}}\right), 2\,\rho^{1-\gamma}_{x,\tilde{x}}\}\rho^{\gamma}_{x,\tilde{x}}. \end{aligned} \tag{10}$$

Proof. For the Prokhorov metric the proof is given in [9]. An important step in the proof is the estimate

$$\begin{aligned} \tilde{d}(F(z),F(\tilde{z})) &\leq h(\max\{d(z,0),d(\tilde{z},0)\})\, d(z,\tilde{z})^{\gamma} \\ &\leq h(d(z_0,0)+\max\{d(z,z_0),d(\tilde{z},z_0)\})\,\delta^{\gamma} \\ &\leq h(d(z_0,0)+\vartheta+\delta)\,\delta^{\gamma}\,. \end{aligned} \tag{11}$$

which holds whenever $d(z,z_0) \leq \vartheta$ and $d(z,\tilde{z}) \leq \delta$.

To show the result for the Ky Fan metric we first choose $\delta > \rho_{\text{K}}(x,\, \tilde{x})$ and $\vartheta > 0$. Then by (3) and (8)

$$\mu\{\omega \in \Omega \mid d(x(\omega),\tilde{x}(\omega)) > \delta \vee d(\tilde{x}(\omega),z_0) > \vartheta\} < \delta + \kappa(\vartheta)\,. \tag{12}$$

Now choose ω with $d(x(\omega),\tilde{x}(\omega)) \leq \delta$ and $d(\tilde{x}(\omega),z_0) \leq \vartheta$. By (12) the probability for this event is at least $1-\delta-\kappa(\vartheta)$. Via (11) we obtain

$$\tilde{d}(F(x(\omega)),F(\tilde{x}(\omega))) \leq h(d(z_0,0)+\vartheta+\delta)\,\delta^{\gamma}$$

and consequently

$$\mu\{\omega \in \Omega \mid \tilde{d}(F(x(\omega)), F(\tilde{x}(\omega))) \leq h(d(z_0, 0) + \vartheta + \delta)\, \delta^\gamma\} \geq 1 - \delta - \kappa(\vartheta)\,.$$

Since h is continuous from the right, this bound holds also for the case $\delta = \rho_\kappa(x, \tilde{x})$. Taking the infimum with respect to ϑ gives the desired result. $\square$

Remark 3.2. In, e.g., (9) we obtained an analogous estimate for the Ky Fan metric as for the Prokhorov metric. Note that, although these results look so similar, neither of them implies the other. The first estimate applies to probability measures, the second one to random variables.

Since Theorem 3.1 is the basis for various results in the following section, we will observe such similarities for several theorems. Whenever it is clear how to translate one result into the other, we will not state both estimates, but only the result for the Prokhorov metric, and note that the analogous result is true as well in the Ky Fan metric to avoid unnecessary repetition.

4 A Convergence Analysis for General Regularization Methods

In this section we transfer the results from [9, 10], which were obtained for Tikhonov regularization, to general regularization methods. Due to Theorem 3.1 we are able to derive all these results for the Prokhorov metric and the Ky Fan metric at the same time. To show the applicability of the new results, we finally derive a full convergence rate result for Landweber's method in section 5.

Using the framework of spectral theory, general convergence results for regularization methods have been obtained in the literature (for some background on spectral theory and its use in regularization theory see, e.g., [8]). Via this approach a large class of regularization schemes can be written as

$$x_\alpha^\delta = R_\alpha y^\delta = g_\alpha\,(A^*A)\,A^* y^\delta\,. \tag{13}$$

Here $g_\alpha(A^*A)$ defines the regularization operator, using the spectral family of the self-adjoint operator A^*A. In general, the *filter function* $g_\alpha(\lambda)$ is assumed to be bounded, but converges point-wise to the unbounded function $1/\lambda$ when $\alpha \to 0$ (see Theorem 4.1). The choice of $g_\alpha(\cdot)$ for various well-known regularization methods is shown in Table 1 (cf. [8]).

The stability properties of the regularization method R_α can now be deduced from properties of $g_\alpha(\cdot)$. For our considerations we assume that

Tikhonov regularization	$g_\alpha(\lambda) = \dfrac{1}{\lambda + \alpha}$
Iterated Tikhonov regularization of order n	$g_\alpha(\lambda) = \dfrac{(\lambda+\alpha)^n - \alpha^n}{\lambda(\lambda+\alpha)^n}$
Truncated singular value decomposition	$g_\alpha(\lambda) = \begin{cases} \lambda^{-1}, & \lambda \geq \alpha \\ 0, & \lambda < \alpha \end{cases}$
Landweber iteration	$g_k(\lambda) = \sum_{j=0}^{k-1} (1-\lambda)^j$

Table 1: The function $g_\alpha(\lambda)$ for some common regularization methods ([8])

there are constants C and G_α such that the following estimates hold

$$|\lambda g_\alpha(\lambda)| \leq C \qquad \forall \lambda \in (0, M^2], \tag{14a}$$
$$|g_\alpha(\lambda)| \leq G_\alpha \qquad \forall \lambda \in [0, M^2]. \tag{14b}$$

If the operator under consideration is random, we must ensure that (14) is satisfied for (almost) all appearing realizations $A(\omega)$, i.e., that (14a) and (14b) hold up to $M = \operatorname{ess\,sup} \|A(\omega)\|^2$, cf. Theorem 5.1 and the subsequent discussion.

4.1 Convergence

We now apply the lifting technique of section 3 to investigate the stability properties of general regularization methods. When only a noisy version A^ξ of the operator A is available, we define the regularized solution $x^\delta_{\alpha(\delta)}$ pointwise as

$$x^\delta_{\alpha,A^\xi}(\omega) := g_\alpha\left(A^\xi(\omega)^* A^\xi(\omega)\right) A^\xi(\omega)^* y^\delta(\omega). \tag{15}$$

To investigate convergence of $x^\delta_{\alpha,A^\xi}(\omega)$ to $x^\dagger(\omega)$ in the Prokhorov metric (or analogously the Ky Fan metric) in dependence of the noise in A and y we split the total error $\rho_{\mathrm{P}}(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}})$ into three parts: the approximation error $\rho_{\mathrm{P}}(\mu_{x^\dagger}, \mu_{x_\alpha})$ in Theorems 4.1 and 4.7, the modelling error $\rho_{\mathrm{P}}(\mu_{x_\alpha}, \mu_{x_{\alpha,A^\xi}})$ in Lemma 4.2, and the propagated data error $\rho_{\mathrm{P}}(\mu_{x_{\alpha,A^\xi}}, \mu_{x^\delta_{\alpha,A^\xi}})$ in Lemma 4.3. The convergence result is given in Theorem 4.4. Imposing an additional source condition on $x^\dagger(\omega)$ we obtain bounds for $\rho_{\mathrm{P}}(\mu_{x^\dagger}, \mu_{x_\alpha})$ in Theorem 4.7

and a convergence rate result for the total error in Theorem 4.8. The appearance of this result for Landweber's method is shown in Theorem 5.1. The following result is a modification of [10, p.100] to the setup of general regularization methods and the Ky Fan metric.

Theorem 4.1. *Let $y(\omega) \in \mathcal{D}(A^\dagger(\omega))$ for almost all ω, where $\mathcal{D}(A^\dagger(\omega)) := \mathcal{R}(A(\omega)) + \mathcal{R}(A(\omega))^\perp$. Let* (14a) *hold and for all $\lambda \in (0, \text{ess sup}\, \|A(\omega)\|^2]$*

$$\lim_{\alpha\to 0} g_\alpha(\lambda) = \frac{1}{\lambda}$$

Then $x_\alpha(\omega) := g_\alpha(A(\omega)^* A(\omega))A(\omega)^* y(\omega)$ converges to $x^\dagger(\omega)$ in the Ky Fan metric as well as in the Prokhorov metric, in particular

$$\rho_{\text{P}}(\mu_{x_\alpha}, \mu_{x^\dagger}) \le \rho_{\text{K}}(x_\alpha, x^\dagger) \to 0 \quad \textit{as} \quad \alpha \to 0\,.$$

Proof. Because $y(\omega) \in \mathcal{D}(A^\dagger(\omega))$ for almost all ω, x_α converges to $x^\dagger$ almost surely, as follows from the deterministic theory [8, Thm. 4.1]. Since almost sure convergence implies convergence in the Ky Fan metric (see [6, 14]) and due to (4) we obtain the result. □

Next, we turn to stability issues that appear when data or operator are noisy. The following technical assumption is necessary only when we want to consider inexact operators $A^\xi(\omega)$. In that case, we must impose an additional decay condition controlling the probability that $y(\omega)$ and $A(\omega)$ differ from some deterministic quantities (e. g., their means) A_0 and y_0. Let $\varphi_{\text{op}}(\cdot)$ be some monotonically decreasing, right continuous function with $\varphi_{\text{op}}(\vartheta) \to 0$ as $\vartheta \to \infty$ such that

$$\mu\{\omega \in \Omega \mid \max\{\|A(\omega) - A_0\|\,, \|y(\omega) - y_0\|\} > \vartheta\} \le \varphi_{\text{op}}(\vartheta)\,. \tag{16}$$

To obtain a stability bound for noise in the operator we suppose furthermore there is a function $\tilde{h}_\alpha(\lambda)$ such that

$$\begin{aligned} &\left\| \left(g_\alpha(A(\omega)^* A(\omega))A(\omega)^* - g_\alpha(A^\xi(\omega)^* A^\xi(\omega))A^\xi(\omega)^* \right) y \right\| \le \\ &\qquad \tilde{h}_\alpha\left(\max\left\{\|y\|\,, \|A(\omega)\|\,, \left\|A^\xi(\omega)\right\|\right\}\right) \left\|A(\omega) - A^\xi(\omega)\right\| \end{aligned} \tag{17}$$

(cf. (25)). For a treatment of noisy operators in the framework of spectral theory see [2, 3, 23].

Lemma 4.2. *Let $x_\alpha(\omega) := g_\alpha(A(\omega)^* A(\omega))A(\omega)^* y(\omega)$ and $x_{\alpha,A^\xi}(\omega)$ denote the regularized solution for exact data and noisy operator, i. e., $x_{\alpha,A^\xi}(\omega) :=$*

$g_\alpha(A^\xi(\omega)^* A^\xi(\omega)) A^\xi(\omega)^* y(\omega)$. *Furthermore, let* $A(\omega)$ *and* $y(\omega)$ *fulfill the decay condition* (16). *Then we have, with* $\rho_A := \rho_{\text{P}}(\mu_{(A,y)},\, \mu_{(A^\xi,y)})$,

$$\begin{aligned}
&\rho_{\text{P}}(\mu_{x_\alpha},\, \mu_{x_{\alpha,A^\xi}}) \\
&\quad \le \inf_{\vartheta>0} \max\{\rho_A\, \tilde h_\alpha\big(\max\{\|A_0\| + \rho_A,\, \|y_0\|\} + \vartheta\big),\, \rho_A + \varphi_{\text{op}}(\vartheta)\} \\
&\quad \le \rho_A \max\{\tilde h_\alpha\big(\max\{\|A_0\| + \rho_A,\, \|y_0\|\} + \varphi_{\text{op}}^{-1}(\rho_A)\big),\, 2\}\,,
\end{aligned} \tag{18}$$

where $\tilde h_\alpha$ *is defined via* (17). *The analogous bound holds for the Ky Fan metric with* $\rho_A := \rho_{\text{K}}(A,\, A^\xi)$ *(cf. Remark 3.2).*

Proof. We estimate $\|x_\alpha(\omega) - x_{\alpha,A^\xi}(\omega)\|$ pointwise for fixed ω and lift the resulting bound using Theorem 3.1 afterwards. Therefore we observe that $x_\alpha(\omega) - x_{\alpha,A^\xi}(\omega)$ can be written as $F(A(\omega), y(\omega)) - F(A^\xi(\omega), y(\omega))$, where the nonlinear (in (A, y)) operator F is defined as $F(A, y) := g_\alpha(A^*A)A^*y$. The bound in (17) yields the local Lipschitz estimate

$$\begin{aligned}
&\big\|F(A(\omega), y) - F(A^\xi(\omega), y)\big\| \le \\
&\qquad\qquad \le \tilde h_\alpha\big(\max\left\{\|A(\omega)\|\,, \big\|A^\xi(\omega)\big\|\,, \|y(\omega)\|\right\}\big)\, \big\|A(\omega) - A^\xi(\omega)\big\|
\end{aligned}$$

Introducing the distance $d\big((A, y), (\tilde A, \tilde y)\big) := \max(\|A - \tilde A\|, \|y - \tilde y\|)$ this estimate attains a form as in (7).

Direct application of Theorem 3.1 to lift this local Lipschitz estimate now immediately yields (18), with $\|y_0\|$ replaced by $\|y_0\| + \rho_A$. Examination of (11) shows that also the slightly better estimate (18) is satisfied. Finally we note that for $d(\cdot, \cdot)$ as above, $\rho_{\text{K}}((A, y),\, (A^\xi, y)) = \rho_{\text{K}}(A,\, A^\xi)$. □

The next lemma deals with instabilities due to noise in the data.

Lemma 4.3. *Let* $x_{\alpha,A^\xi}(\omega)$ *be defined as in Lemma 4.2 and* $x^\delta_{\alpha,A^\xi}(\omega)$ *be as in* (15), *and let* (14) *hold. Then with* $\rho_y := \rho_{\text{P}}(\mu_{(A^\xi,y)},\, \mu_{(A^\xi,y^\delta)})$

$$\rho_{\text{P}}(\mu_{x_{\alpha,A^\xi}},\, \mu_{x^\delta_{\alpha,A^\xi}}) \le \max\left\{\sqrt{CG_\alpha},\, 1\right\} \rho_y\,. \tag{19}$$

The analogous estimate holds in the Ky Fan metric with $\rho_y := \rho_{\text{K}}(y,\, y^\delta)$ *(cf. Remark 3.2).*

Proof. The regularization method we consider is such, that for (almost) all realizations of $A(\omega)$

$$\|g_\alpha(A(\omega)^* A(\omega)) A(\omega)^*\| \le \sqrt{CG_\alpha}$$

(cf. [8, Thm. 4.2]) which yields

$$\left\|F(A^{\xi}(\omega), y(\omega)) - F(A^{\xi}(\omega), y^{\delta}(\omega))\right\| \leq \sqrt{CG_{\alpha}} \left\|y(\omega) - y^{\delta}(\omega)\right\| ,$$

(a. s.) with F as in Lemma 4.2. Using the distance $d\big((A, y), (\tilde{A}, \tilde{y})\big) := \max(\|A - \tilde{A}\|, \|y - \tilde{y}\|)$, we can transfer this estimate to a form as in (7) with h being constant. The proof now follows by applying Theorem 3.1 and the observation that the infimum in (9) is attained for $\vartheta \to \infty$. $\square$

A combination of the previous results gives the following convergence theorem.

Theorem 4.4. *Let $y(\omega) \in \mathcal{D}(A^{\dagger}(\omega))$ for almost all ω. Let ρ_y and ρ_A be defined as in Lemma 4.2 and 4.3. Let $\alpha = \alpha(\rho_y, \rho_A)$ be chosen such that α, $\sqrt{CG_{\alpha}}\rho_y$ as well as (18) tend to 0 as $\rho_y \to 0$ and $\rho_A \to 0$. Then*

$$\rho_{\mathrm{P}}(\mu_{x^{\delta}_{\alpha,A^{\xi}}}, \mu_{x^{\dagger}}) \to 0$$

as $\rho_y \to 0$ and $\rho_A \to 0$. The analogous statement holds for the Ky Fan metric (cf. Remark 3.2).

Proof. The full error $\rho_{\mathrm{P}}(\mu_{x^{\delta}_{\alpha,A^{\xi}}}, \mu_{x^{\dagger}})$ can be bounded via the estimates in Theorem 4.1, Lemma 4.2 and 4.3. To obtain the convergence result we must show only that a parameter choice rule $\alpha(\rho_y, \rho_A)$ with the required properties can indeed be found. Consider first of all (18) and assume that ρ_A tends to 0. Letting α shrink and ϑ grow sufficiently slowly, we can ensure that the first part in the max-expression still tends to 0. But since ϑ is growing, also the second part will start to shrink, and in total we find that a choice of $\alpha \to 0$ such that (18) tends to 0 when $\rho_A \to 0$ is possible. For $\alpha > 0$, $G_{\alpha} < \infty$. Therefore also a choice of α with $\sqrt{G_{\alpha}}\rho_y \to 0$ (cf. (19)) is possible. After these observation, the proof follows immediately by combining Theorem 4.1 with Lemmas 4.2 and 4.3 $\square$

4.2 Convergence Rates

The stability results in the previous section are all based on properties of the regularization method only, in particular, they are independent of the quality (e. g., smoothness) of the exact solution $x^{\dagger}$. It is well known, that convergence rate results for deterministic inverse problems can be obtained only if the solution $x^{\dagger}$ satisfies certain smoothness assumptions. In [9] we extended such smoothness conditions (so called source conditions) from the deterministic theory to a stochastic setup and obtained convergence rate results in the Prokhorov metric. The main ingredient are decay conditions, formulated via a function $\varphi_{\mathrm{de}}(\tau)$. This function will control the probability that $x^\dagger(\omega)$ carries a certain smoothness.

As in [9] we consider abstract source conditions, i. e., conditions of the form $x^\dagger = f(A^*A)v$. Under appropriate assumptions on f, convergence rates (corresponding to h in (22)) have been shown e. g., in [18] and [22]. Such assumptions can be e. g., monotonicity and concavity of f.

For f with $\lim_{\lambda\to 0} f(\lambda) = 0$ we define the smoothness set $X_{f,\tau}(\omega)$ as

$$X_{f,\tau}(\omega) := \{z \in X \mid z = f\big(A(\omega)^* A(\omega)\big)v,\ \|v\| \le \tau\}\,, \tag{20}$$

i. e., all elements $x \in X_{f,\tau}(\omega)$ satisfy a source condition with f, $A(\omega)$ and a *source element* v, $\|v\| \le \tau$. In order to obtain convergence rates in a stochastic setting we assume that the probability of $x^\dagger(\omega) \notin X_{f,\tau}(\omega)$ is small for large τ, more precisely, with probability $\varphi_{\mathrm{de}}(\tau)$, $x^\dagger(\omega)$ satisfies a source condition with source element v, $\|v\| \le \tau$. The decay of this function $\varphi_{\mathrm{de}}(\cdot)$, influences the resulting convergence rates.

Definition 4.5. *Let $X_{f,\tau}(\omega)$ be a smoothness set as in* (20)*. We say that a stochastic source condition for $x^\dagger$ holds, if there exists a decreasing function $\varphi_{\mathrm{de}}(\tau)$ such that*

$$\mu\{\omega \in \Omega \mid x^\dagger(\omega) \in X_{f,\tau}(\omega)\} \ge 1 - \varphi_{\mathrm{de}}(\tau)\ . \tag{21}$$

Discussions of these source conditions are given in [9]. To obtain stochastic convergence rate results it is important which (deterministic) rate can be obtained for the approximation error when $x^\dagger$ satisfies a source condition as in (20) with a given function f. In the following, we introduce a function h, which measures the speed of convergence that can be obtained with the given regularization method.[3] In many cases, $f \equiv h$ (e. g., for Tikhonov regularization with $f(\lambda) = \lambda^\nu$, $\nu \le 1$), but this need not be the case, e. g., if *saturation* occurs (cf. e. g. [8]).

Definition 4.6. *For a decreasing function h and a regularization scheme* (13)*, we say that f allows the deterministic convergence rate h if for any continuous linear operator $A : X \to Y$, any $x^\dagger$ and $x_\alpha = g_\alpha(A^*A)A^*Ax^\dagger$,*

$$x^\dagger \in \{z \in X \mid z = f\big(A^*A\big)v,\ \|v\| \le \tau\} \implies \big\|x^\dagger - x_\alpha\big\| \le \tau h(\alpha). \tag{22}$$

[3]Note that the function $h(\alpha)$ corresponds to $w_\mu(\alpha)$ in [8, Theorem 4.3], but that our approach is more general in the sense that it also allows logarithmic convergence rates.

The most popular smoothness functions f are either of Hölder or of logarithmic type, i. e., $f(\lambda) = \lambda^\nu$ or $f(\lambda) = (-\ln\lambda)^{-\nu}$. Convergence rate results for these cases can be found in [8] and [15], respectively.

Observe that the resulting function h must be independent of properties of A. This is typically not an issue, e. g., $f(\lambda) = \lambda^\nu$ with $\nu \leq 1$ implies $h(\alpha) = \alpha^\nu$ for Tikhonov regularization without any restrictions on the linear operator A.

Let us now derive the result for the approximation error for general regularization methods. The appearance of the corresponding theorem for Tikhonov regularization, Theorem 3.3 in [9], remains essentially unchanged. The reason is that also there we used a general setup by introducing the function f for the source condition and h for the convergence behavior.

Theorem 4.7 (Approximation Error)**.** *Let $x^\dagger(\omega)$ satisfy a stochastic source condition* (21)*. Suppose f allows the deterministic convergence rate h as in* (22)*. Then the distance of $\mu_{x^\dagger}$ and μ_{x_α} in the Prokhorov metric is bounded by*

$$\rho_{\mathrm{P}}(\mu_{x^\dagger},\, \mu_{x_\alpha}) \leq \inf_{\tau\geq 0} \max\left\{\tau h(\alpha),\, \varphi_{\mathrm{de}}(\tau)\right\}\,. \tag{23}$$

The analogous estimate holds for the Ky Fan metric (cf. Remark 3.2).

Proof. For the Prokhorov metric and Tikhonov regularization this result was derived in [9]. For general methods the first part of the proof follows analogously, defining $G_\alpha(\omega) := g_\alpha(A(\omega)^*A(\omega))A(\omega)^*A(\omega)$.

To deduce this result for the Ky Fan metric fix $\tau > 0$. With probability $1 - \varphi_{\mathrm{de}}(\tau)$ we have $x^\dagger(\omega) \in X_{f,\tau}(\omega)$, and therefore via (22)

$$\left\|x^\dagger(\omega) - x_\alpha(\omega)\right\| \leq \tau h(\alpha)\,,$$

i. e., $\mu(\|x^\dagger(\omega) - x_\alpha(\omega)\| \leq \tau h(\alpha)) \geq 1 - \varphi_{\mathrm{de}}(\tau)$. Taking the infimum over $\tau > 0$ gives the desired bound for the Ky Fan metric. □

A combination of Lemma 4.2 and Lemma 4.3 from the previous section with Theorem 4.7 yields our main result: A convergence rate result in the Prokhorov and Ky Fan metric for the total error and for general regularization methods. In Theorem 5.1, we show how this result simplifies for Landweber's method.

Theorem 4.8. *Let the assumptions of Lemma 4.2 and Theorem 4.7 be fulfilled. Let the source function $f(\lambda)$ allow the deterministic convergence rate $h(\alpha)$ for the regularization scheme* (13)*. Then with $\tilde{h}_\alpha$ as in* (17) *and*

$\rho_A := \rho_{\mathrm{P}}(\mu_{(A,y)},\, \mu_{(A^\xi,y)})$

$$\begin{aligned}
&\rho_{\mathrm{P}}(\mu_{x^\dagger},\, \mu_{x^\delta_{\alpha,A^\xi}}) \le \\
&\quad \le \inf_{\tau\ge 0} \max\{\tau h(\alpha),\, \varphi_{\mathrm{de}}(\tau)\} + \\
&\qquad + \inf_{\vartheta>0} \max\{\rho_A\, \tilde{h}_\alpha\big(\max\{\|A_0\| + \rho_A,\, \|y_0\|\} + \vartheta\big),\, \rho_A + \varphi_{\mathrm{op}}(\vartheta)\} \\
&\qquad + \max\left\{\sqrt{CG_\alpha},\, 1\right\} \rho_y .
\end{aligned}$$

The analogous bound holds for the Ky Fan metric (cf. Remark 3.2).

Proof. The proof follows combining Lemma 4.2, Lemma 4.3 and Theorem 4.7. □

If the operator A is known exactly (i. e., $A^\xi = A$, A may still be random), the decay condition (16), which is used in Lemma 4.2 only, is superfluous. This situation occurs for instance when the model depends in a deterministic way on some external parameters that can be observed, but are random. In this case the theorem above simplifies to the following result:

Corollary 4.9. *Let the assumptions of Theorem 4.7 be fulfilled and the (possibly random) operator A be given exactly, i. e., $A = A^\xi$. Then the error is bounded as*

$$\rho_{\mathrm{P}}(\mu_{x^\dagger},\, \mu_{x^\delta_\alpha}) \le \inf_{\tau\ge 0} \max\{\tau h(\alpha),\, \varphi_{\mathrm{de}}(\tau)\} + \max\left\{\sqrt{CG_\alpha},\, 1\right\} \rho_y\, .$$

The analogous bound holds for the Ky Fan metric (cf. Remark 3.2).

5 Landweber's Method

The following corollary demonstrates the results of the previous section for Landweber's method. In this iterative method, we define $x_0^\delta = 0$ and

$$x_k^\delta = x_{k-1}^\delta - A(\omega)^*(A(\omega)x_{k-1}^\delta - y^\delta(\omega))\,, \qquad k = 1, 2, \ldots \tag{24}$$

The resulting filter function $g_k(\lambda)$ is shown in Table 1. For the approximations x_k^δ, we get the following bounds in the Prokhorov and Ky Fan metric.

Theorem 5.1 (Landweber's method)**.** *Let $x^\dagger(\omega)$ satisfy a stochastic source condition* (21) *with $f(\lambda) = \lambda^\nu$. Furthermore, let $A(\omega)$, $A^\xi(\omega)$ and $y^\delta(\omega)$ be such, that decay condition* (16) *is satisfied, and that $\|A(\omega)^*A(\omega)\| \le 1$ as* *well as* $\|A^\xi(\omega)^* A^\xi(\omega)\| \le 1$ *for almost all* ω*. Then the approximation* $x_k^\delta(\omega)$ *obtained after* k *steps of Landweber iteration satisfies*

$$\begin{aligned} &\rho_{\mathrm{P}}(\mu_{x^\dagger}, \mu_{x_k^\delta}) \le \\ &\quad \le \inf_{\tau \ge 0} \max\left\{\tau \max\{\nu^\nu, 1\}(k+1)^{-\nu}, \varphi_{\mathrm{de}}(\tau)\right\} + 2k\rho_A + \sqrt{k}\rho_y. \end{aligned}$$

The analogous estimate holds for the Ky Fan distance $\rho_{\mathrm{K}}(x^\dagger, x_k^\delta)$ *(cf. Remark 3.2).*

Proof. The expressions for the first and the third term can be derived from results in [8]: For Landweber's method we have the stability estimate $\sqrt{CG_k} = \sqrt{k}$ (cf. Lemma 6.2 in [8]). When a Hölder-source condition is satisfied, the deterministic convergence rate is given as $h(\alpha) = \max\{\nu^\nu, 1\}(k+1)^{-\nu}$ (equation (6.8) in [8]).

A little more effort is necessary to obtain a bound on the instability caused by the noise in the operator. Following, e. g., the proof of Lemma 2.9 in [17] we find that

$$\begin{aligned} &g_k(A^*A)A^* - g_k(B^*B)B^* \\ &\quad = (I - A^*A)^k - (I - B^*B)^k \\ &\quad = \sum_{j=0}^{k} (I - A^*A)^{k-j-1}(A^*(B-A) + (B^* - A^*)B)(I - B^*B)^j \\ &\quad = \sum_{j=0}^{k} (I - A^*A)^{k-j-1} A^*(B-A)(I - B^*B)^j \\ &\quad\ + \sum_{j=0}^{k} (I - A^*A)^{k-j-1}(B^* - A^*)B(I - B^*B)^j\,. \end{aligned}$$

So $\|g_k(A^*A)A^* - g_k(B^*B)B^*\| \le 2k\|B - A\|$ and (17) is satisfied with

$$\tilde{h}_k(\lambda) = 2k. \tag{25}$$

Since this function is independent of λ we can replace

$$\inf_{\vartheta > 0} \max\{\rho_A\, \tilde{h}_k\big(\max\{\|A_0\| + \rho_A, \|y_0\|\} + \vartheta\big), \rho_A + \varphi_{\mathrm{op}}(\vartheta)\}$$

in Theorem 4.8 by $2k\rho_A$, which concludes the proof. □

The requirement $\|A(\omega)^* A(\omega)\| \le 1$ can be met by scaling the problem properly (in dependence of the realization $A(\omega)$). This scaling factor can as well be seen as a relaxation factor in the update rule (24), cf. [8, Ch. 6.1].

Remark 5.2 (Parameter choice rule)**.** Let ρ_A and $\rho_y \to 0$. Then a choice of k with

$$k \to \infty, \qquad \sqrt{k}\rho_y \to 0, \qquad k\rho_A \to 0 \tag{26}$$

ensures convergence of Landweber's method for noise level ρ_A and ρ_y tending to 0. To deduce results on convergence *rates*, more information is necessary. Besides the source condition (i. e., the parameter ν and the function $\varphi_{\mathrm{de}}(\cdot)$ in Theorem 5.1) also the interplay of ρ_A and ρ_y influences the parameter choice rule (see [9, Remark 4.5]).

Observe that (26) is an a-priori rule for choosing the regularization parameter as in the deterministic theory, but now referring to the data error in the Prokhorov or Ky Fan metric instead of a norm bound. At this point, our theory changes also the computations via a different parameter choice rule.

Furthermore note that it is necessary that the parameter choice rule (26) utilizes a bound on the noise measured in the Prokhorov or Ky Fan metric to deduce respective convergence results. As shown in Theorem 2.1, also in this stochastic setup "error-free methods", i. e., methods where α depends only on y^δ, but not on ρ_y, do not lead to convergence for ill-posed problems.

Acknowledgments

The author thanks Prof. H.W. Engl and S. Kindermann for useful and stimulating discussions.

This work has been supported by the Austrian National Science Foundation FWF under project grant SFB F013/F1308.

References

[1] A. Bakushinsky, *Remarks on the choice of the regularization parameter using quasioptimality and a quotient criterion.*, U.S.S.R. Comput. Math. Math. Phys., 24 (1984), pp. 181–182.

[2] A. Bakushinsky and A. Goncharsky, *Ill-Posed Problems: Theory and Applications*, Kluwer Academic Publishers, 1989.

[3] A. Bakushinsky and M. Kokurin, *Iterative methods for approximate solution of inverse problems.*, Mathematics and its Applications (Springer) 577. Dordrecht: Springer., 2004.

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., 1968.

[5] R. M. Dudley, *Distances of probability measures and random variables*, Ann. Math. Stat., 39 (1968), pp. 1563–1572.

[6] ——, *Real Analysis and Probability*, Wadsworth, second ed., 2003.

[7] H. W. Engl and W. Grever, *Using the L-curve for determining optimal regularization parameters.*, Numer. Math., 69 (1994), pp. 25–31.

[8] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[9] H. W. Engl, A. Hofinger, and S. Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems, 21 (2005), pp. 399–412.

[10] H. W. Engl and A. Wakolbinger, *Continuity properties of the extension of a locally Lipschitz continuous map to the space of probability measures*, Monatsh. Math., 100, 2 (1995), pp. 85–103.

[11] K. Fan, *Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit.*, Mathematische Zeitschrift, 49 (1944), pp. 681–683.

[12] P. C. Hansen, *Numerical tools for analysis and solution of Fredholm integral equations of the first kind.*, Inverse Problems, 8 (1992), pp. 849–872.

[13] A. Hofinger, *Ill-Posed Problems: Extending the Deterministic Theory to a Stochastic Setup*, vol. 51 of Schriften der Johannes-Kepler-Universität Linz. Reihe C: Technik und Naturwissenschaften, Trauner Verlag, Linz, 2006. (Thesis).

[14] ——, *The metrics of Prokhorov and Ky Fan for assessing uncertainty in inverse problems*, Sitzungsber., Abt. II, Österr. Akad. Wiss., Math.-Naturwiss. Kl., to appear (2006).

[15] T. Hohage, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim., 21 (2000), pp. 439–464.

[16] P. Huber, *Robust Statistics*, John Wiley & Sons, Inc., 1981.

[17] B. Kaltenbacher, *Some Newton-type methods for the regularization of nonlinear ill-posed problems*, Inverse Problems, 13 (1997), pp. 729–753.

[18] P. Mathé and S. V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), pp. 789–803.

[19] V. Morozov, *On the solution of functional equations by the method of regularization*, Sov. Math., Dokl., 7 (1966), pp. 414–417.

[20] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1967.

[21] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Stat., 36 (1965), pp. 423–439.

[22] U. Tautenhahn, *Optimality for ill-posed problems under general source conditions*, Numer. Funct. Anal. Optim, 19 (3 & 4) (1998), pp. 377–398.

[23] G. M. Vainikko, *The discrepancy principle for a class of regularization methods*, U.S.S.R. Comput. Maths. Math. Phys., 23 (1982), pp. 1–19.

[24] C. R. Vogel, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.