Convergence Rates in the Prokhorov Metric for Assessing Uncertainty in Ill-Posed Problems*

Heinz W. Engl[†‡] Andreas Hofinger[‡] Stefan Kindermann[§]

Abstract

In the convergence theory of regularization methods for ill-posed problems, so far deterministic error concepts were dominating, which leads to worst-case error estimates. Since this is sometimes not desirable, we aim at providing a framework for proving convergence rates in the Prokhorov metric for the regularization of ill-posed problems with stochastic noise. This allows to assess uncertainty in the sense of a confidence region for the probability that the deviation between exact and regularized solution stays below a given bound with given probability. We exemplify this method for the special case of Tikhonov regularization for linear ill-posed problems and apply the result to the problem of deblurring an image contaminated by random blurring.

1 Introduction

Inverse problems are usually ill-posed in the sense of Hadamard and hence especially sensitive to noise. In analyzing stable methods for their solution, so-called *regularization methods*, one therefore has to lay special emphasis on estimating the error propagation and, more general, the rate with which such methods converge as the noise level tends to 0. This is especially important since for ill-posed problems, convergence of regularization methods can be arbitrarily slow ([19]) and convergence *rates* can only be proven under a-priori conditions concerning the smoothness of the (unknown) exact solution. Even under such assumptions, convergence rates can be rather slow

*This work has been supported by the Austrian National Science Foundation FWF under project grant SFB F 13/08.

[†]Industrial Mathematics Institute, Johannes Kepler Universität, Linz, Austria.

[‡]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Linz, Austria.

[§]Department of Mathematics, University of California at Los Angeles, USA.

depending on the *degree of ill-posedness* ([8, p.40]) of the problem. In order to assess the efficiency of regularization methods and in order to choose e. g. regularization parameters in such a way that the best possible convergence rate for a given situation (which may be slow anyway) can at least be achieved, precise estimates of rates of convergence are necessary. See [8] for a detailed account of these aspects; there, as in most of the "functional analysis based" literature on the numerical solution of ill-posed problems, the error (both in the data and in the numerical approximations) is treated in a deterministic way, one assumes bounds on the error in the data and aims at bounds for the error between regularized and exact solutions in some (usually Hilbert space) norms. In this way, stochastic aspects like error *distribution* or confidence regions for regularized solutions in a probabilistic sense are neglected: It would be advantageous to know that with a given probability, the deviation between the regularized and exact solutions for an inverse problem with noisy data stays below a certain bound. Instead, error estimates in the traditional functional analytic setting are usually of a worst-case type in the sense that they give an estimate for this deviation which is valid for *all* noisy data compatible with the (norm) bound for the data noise. This is, in many situations, a much too stringent approach for assessing *uncertainty*. The appropriate notion of convergence which leads to confidence regions seems to be *convergence in distribution* of random variables, which is metrizable by the *Prokhorov metric*. This is our motivation for studying convergence (rates) for regularization methods in the Prokhorov metric. The first paper where this has been done (to our knowledge) is [11], where this was only a by-product of other investigations. More recently, the Prokhorov metric has been used in connection with parameter identification problems for differential equations from biology, where stochasticity is arising naturally in the model not only through noise (see [1]).

While our approach is based on convergence in distribution, there is as lot of literature on using other stochastic convergence concepts in connection with inverse problems, see e. g. [22, 16, 17, 4, 2, 12, 9, 10]. Motivated by statistical estimation theory, one usually aims there at estimates for the expected value of the squared norm of the deviation. Whether we use such a concept or convergence in distribution, we have to treat input and output variables of our inverse problem as random variables, not just as elements in some Hilbert space as in the usual functional analytic theory. Randomness may enter an inverse problem in at least two ways, which we exemplify for a linear ill-posed problem

$$Ax = y\,,$$

where A is a linear operator between Hilbert spaces X, Y:

Data Noise. Instead of the exact data y, noisy data y^δ are given. Furthermore, instead of the operator A only an approximate version A^ξ might be available, e. g., in the case of an integral operator A for which the kernel is not known exactly ("modeling error", see [15] and Section 5 for examples). If this noise is modeled via Y- and $L(X,Y)$-valued random variables, then consequently, also the solution x is an X-valued random variable.

Stochastic Model Parameters. Another source for randomness lies in the fact that parameters in the underlying model are of a stochastic nature themselves, even without noise. This is frequently the case in biological applications, where one is usually not interested in a single individual, but in an ensemble of individuals, and the parameters entering into a model and also the output from the model are then *distributions* of these parameters over this ensemble (see, e. g., [1]). Similar situations occur also in physics when one is studying an experiment which is repeated many times and the quantity of interest is not the outcome of a specific realization of the experiment but again the distribution of outcomes over the collection of experiments; see [15] for an example from time-resolved fluorescence. In such situations, it might be advantageous to use a stochastic model from the outset.

In the approach we present, it does not matter from which source the randomness arises, both aspects can be analyzed simultaneously in the same framework; moreover, the results will contain the well-known results from the deterministic theory by specialization to "constant" random variables.

In the following we assume that the original linear problem is influenced by some external random parameter ω, element of a probability space $(\Omega, \mathcal{A}, \mu)$, i. e., for fixed ω we have the equation

$$A(\omega)x(\omega) = y(\omega)\,. \tag{1.1}$$

Here, A and y are $L(X,Y)$- and Y- valued measurable functions, which implies by using measurable selection theorems (cf. e. g. [6]) that also x is measurable (or, in case of multiple solutions, that measurable selections exist). If $A(\omega)$ is a. s. a compact operator, then equation (1.1) is ill posed and we have to use, e. g., Tikhonov regularization to solve it in a stable way. The regularized solution $x^\delta_{\alpha,A^\xi}(\omega)$ obtained from noisy data y^δ using an approximate operator A^ξ with Tikhonov-regularization is given by

$$x^\delta_{\alpha,A^\xi}(\omega) := (A^\xi(\omega)^* A^\xi(\omega) + \alpha I)^{-1} A^\xi(\omega)^* y^\delta(\omega) =: R^\delta_\alpha(\omega) y^\delta(\omega)\,, \tag{1.2}$$

which is now a random variable. The aim is to estimate the deviation between the least-squares minimum-norm solution $x^\dagger(\omega)$ of (1.1), which is a random variable (see [9, 10]) and the approximation $x^\delta_{\alpha,A^\varepsilon}(\omega)$ in dependence of the noise in the data and the operator. There are convergence results available for stochastic noise for the situation that the noise level is defined via an expected value $E(\|y - y_\delta\|^2) = \delta^2$ and deviation and convergence are then also measured via expected values as $E(\|x^\delta_\alpha - x\|^2)$ (see [16, 22, 17, 4]).

As motivated above, we use convergence in distribution for measuring convergence, and quantify this by using the metric which generates convergence in distribution, the *Prokhorov metric*. Based on the probability measure μ we denote the distribution of an X-valued random variable x by μ_x, i. e., μ_x is a probability measure on X with

$$\mu_x(B) = \mu(\{\omega \mid x(\omega) \in B\}) =: \mu\left(x^{-1}(B)\right) \quad (B \subset X, \text{Borel-set}). \tag{1.3}$$

The distributions μ_y and μ_A of y and A are defined analogously.

Convergence in distribution of random variables is now defined as weak (-star) convergence of their distributions, which is generated by the following metric: The Prokhorov distance between two measures μ and $\tilde{\mu}$ is defined as (see, e. g., [3, 14, 18, 5])

$$\rho(\mu, \tilde{\mu}) = \inf\{\varepsilon > 0 \mid \mu(B) \leq \tilde{\mu}(B^\varepsilon) + \varepsilon, \,\forall\, \text{Borel-sets}\, B\}, \tag{1.4}$$

where $B^\varepsilon := \{z \in Z \mid d(z, B) \leq \varepsilon\}$. Hence, if the random variables x and z have the distributions μ_x and μ_z, respectively, then

$$\rho(\mu_x, \mu_z) = \inf\{\varepsilon > 0 \mid \mu\{\|x - z\| > \varepsilon\} < \varepsilon\}. \tag{1.5}$$

If a sequence of random variables x_n converges to x in distribution, then for any bounded uniformly continuous real-valued function f, also the expectations of the random variables obtained by applying f to the original sequence converge, i. e., $E(f(x_n)) \to E(f(x))$ (see, e. g., [3]). This result is not quantitative. Since we aim at quantitative estimates, i. e., convergence rates, the following estimate between the Prokhorov distance of two distributions and the difference between expectations of nonlinear functionals of the underlying random variables is important:

Let f be a bounded Lipschitz continuous real valued function on X with Lipschitz constant L and bound C; then for X-valued random variables x and z with distributions μ_x and μ_z,

$$\mid E(f(x)) - E(f(z)) \mid \leq (L + 2C)\rho(\mu_x, \mu_z). \tag{1.6}$$

This follows quite easily from (1.5) [23] and allows to deduce quantitative convergence results for expectations of certain nonlinear functionals of random variables from the convergence rates in the Prokhorov metric we prove in this paper.

If now, $x^\dagger$ is the least-squares minimum-norm solution of (1.1) with distribution $\mu_{x^\dagger}$, our aim is to estimate $\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}})$ in terms of the Prokhorov distances of input quantities, namely of $\rho(\mu_A, \mu_{A^\xi})$ and $\rho(\mu_y, \mu_{y^\delta})$. We do this by splitting the error term into several parts and estimating them separately by methods developed in [11]:

The general principle for obtaining these estimates is to first construct point-wise estimates for fixed ω, which can be obtained from traditional (deterministic) regularization theory, and then to lift them to the space of probability measures equipped with the Prokhorov metric. For the latter step, estimates about the tail behaviour of certain distribution, i. e., about the probability of large deviations, have to be used.

From such estimates for $\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}})$, one can finally obtain estimates for $\mid E(f(x^\dagger)) - E(f(x^\delta_{\alpha,A^\xi})) \mid$ via (1.6) for any bounded Lipschitz continuous real valued function f. In addition, by (1.5), the estimate for $\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}})$ gives information about a confidence region for $x^\delta_{\alpha,A^\xi}(\omega)$ in the sense described above. Such estimates are reminiscent of estimates used in function estimation from point values (cf. e. g. Corollary 2 in [20]).

We finally note that although we exemplify our method for a specific method, namely Tikhonov regularization, for *linear* problems, the principle is of much wider applicability. In this paper, we concentrate on outlining this principle, which does, on this level, not yet result in a new computational method, but in a new view on an existing method concerning error estimates. When this view is combined with parameter choice strategies like the discrepancy principle or similar strategies involving the noise level (see [8, 7]), new ways for computing regularization parameters will result since the concept of noise level is different. See the remark at the end of Section 4 for a first hint into this direction. All these aspect will be topics for future work.

2 Lifting of Hölder Estimates to Spaces of Probability Measures

Since deterministic error estimates for the regularization of mildly ill-posed problem are usually of Hölder type (as opposed to the severely ill-posed case, where one obtains logarithmic estimates, see e. g. [13]), the first task is to lift (local) *Hölder* estimates for mappings between Hilbert spaces (or, more general, metric spaces) to spaces of probability measures on these spaces, equipped with the Prokhorov metric. This can then be used for lifting deterministic error estimated for all kinds of (regularization) methods to the stochastic situation. In this paper, where we will only deal with linear inverse problems, we will not yet use this result verbatim; the reason for this is that for linear problems, we can split the total error into several parts, and for each part, we use a variant of the general lifting result and of its proof. Nevertheless, we formulate and prove this result also for later use, e. g., for nonlinear problems. We proceed along the lines of [11] where this was done for local *Lipschitz* estimates.

If $F : X \to Y$, then 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 on X, $\tilde{F}(\mu)$ defines a probability measure on Y via

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

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

$$\tilde{d}(F(z), F(\tilde{z})) \leq h(\max\{d(z,0), d(\tilde{z},0)\})\, d(z,\tilde{z})^\gamma \quad (z, \tilde{z} \in X\,). \tag{2.1}$$

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, which is a requirement on the *tail behavior* of one of the measures involved. We assume that the measure $\tilde{\mu}$ fulfills, with some monotonically decreasing, right continuous function κ ($\kappa(\vartheta) \to 0$, $\vartheta \to \infty$), the following decay condition:

$$\tilde{\mu}(\mathcal{B}(z_0, \vartheta)') \leq \kappa(\vartheta) \quad (\vartheta > 0)\,. \tag{2.2}$$

Here $\mathcal{B}(z_0, \vartheta)$ denotes a ball with radius ϑ around some specific element z_0, by B' we denote the complement of a set B, i. e., for $B \subset X$, $B' := X - B$.

The following theorem is an extension of Theorem 1 in [11] and shows that the lifted operator $\tilde{F}$ fulfills a Hölder condition with the same exponent as the original operator F:

Theorem 2.1. *Let assumptions* (2.1) *and* (2.2) *be fulfilled. Then we have*

$$\begin{aligned} &\rho(\tilde{F}(\mu), \tilde{F}(\tilde{\mu})) \leq \qquad (2.3)\\ &\quad \leq \inf_{\vartheta > 0} \max\{\rho(\mu,\tilde{\mu})^\gamma h(d(z_0, 0) + \vartheta + \rho(\mu,\tilde{\mu})), \rho(\mu,\tilde{\mu}) + \kappa(\vartheta)\}\,. \end{aligned}$$

Especially, with $\kappa^{-1}(\tau) := \inf\{\vartheta \geq 0 \mid \kappa(\vartheta) \leq \tau\}$ *we obtain the Hölder estimate*

$$\begin{aligned} &\rho(\tilde{F}(\mu), \tilde{F}(\tilde{\mu})) \leq \\ &\quad \leq \rho(\mu,\tilde{\mu})^\gamma \max\{h\big(d(z_0,0) + \kappa^{-1}(\rho(\mu,\tilde{\mu})) + \rho(\mu,\tilde{\mu})\big), 2\,\rho(\mu,\tilde{\mu})^{1-\gamma}\}. \end{aligned} \tag{2.4}$$

Proof. For obtaining an estimate for the Prokhorov distance of $\tilde{F}(\mu)$ and $\tilde{F}(\tilde{\mu})$, according to the definition of $\tilde{F}$ and the Prokhorov metric, we have to estimate the infimum of all $\varepsilon > 0$ s. t.

$$\mu(F^{-1}(B)) \leq \tilde{\mu}(F^{-1}(B^\varepsilon)) + \varepsilon \qquad \text{for all Borel-sets } B\,. \tag{2.5}$$

For $\delta > \rho(\mu,\tilde{\mu})$ and $\vartheta > 0$ we have, due to (2.2),

$$\begin{aligned} \mu(F^{-1}(B)) &\leq \tilde{\mu}(F^{-1}(B)^\delta) + \delta \\ &\leq \tilde{\mu}(F^{-1}(B)^\delta \cap \mathcal{B}(z_0,\vartheta)) + \tilde{\mu}(\mathcal{B}(z_0,\vartheta)') + \delta \\ &\leq \tilde{\mu}(F^{-1}(B)^\delta \cap \mathcal{B}(z_0,\vartheta)) + \kappa(\vartheta) + \delta. \end{aligned} \tag{2.6}$$

We now construct B^ε with $F(F^{-1}(B)^\delta \cap \mathcal{B}(z_0,\vartheta)) \subseteq B^\varepsilon$ which is equivalent to $F^{-1}(B)^\delta \cap \mathcal{B}(z_0,\vartheta) \subseteq F^{-1}(B^\varepsilon)$.

For z in $F^{-1}(B)^\delta \cap \mathcal{B}(z_0,\vartheta)$ the distance of $F(z)$ to B can be bounded as follows: Choose $\tilde{z} \in F^{-1}(B)$ with $d(z,\tilde{z}) \leq \delta$. Since $d(z,z_0) \leq \vartheta$, we have

$$\begin{aligned} \tilde{d}(F(z), F(\tilde{z})) &\leq h(\max\{d(z,0), d(\tilde{z},0)\})\, d(z,\tilde{z})^\gamma \\ &\leq h(d(z_0,0) + \max\{d(z,z_0), d(\tilde{z},z_0)\})\, \delta^\gamma \\ &\leq h(d(z_0,0) + \vartheta + \delta)\, \delta^\gamma\,. \end{aligned} \tag{2.7}$$

Combining (2.6) and (2.7) we obtain

$$\mu(F^{-1}(B)) \leq \tilde{\mu}\left(F^{-1}(B^{h(d(z_0,0)+\vartheta+\delta)\,\delta^\gamma})\right) + \kappa(\vartheta) + \delta\,.$$

Using the continuity of h from the right, (2.3) now follows by taking the infimum over all $\delta > \rho(\mu,\tilde{\mu})$ and $\vartheta > 0$. For the choice $\vartheta = \kappa^{-1}(\rho(\mu,\tilde{\mu}))$, (2.4) is obtained. □

The following special case is important if we consider the approximation of a deterministic quantity $\tilde{\mu}$ by a probability distribution μ:

Corollary 2.2. *Let* $\tilde{\mu}$ *be a point-measure* δ_{z_0} *and let assumption* (2.1) *be fulfilled. Then*

$$\rho(\tilde{F}(\mu), \tilde{F}(\tilde{\mu})) \leq \rho(\mu,\tilde{\mu})^\gamma \max\{h(d(z_0,0) + \rho(\mu,\tilde{\mu})), \rho(\mu,\tilde{\mu})^{1-\gamma}\}\,.$$

Proof. This follows from the fact that for a point-measure $\tilde{\mu}$ we have (2.2) with $\kappa(\vartheta) = 0$ for $\vartheta > 0$. Since h is monotonically increasing, the infimum in (2.3) is attained for the limit $\vartheta \to 0$. □

3 Lifting of Convergence Rates

It is a general result that for an ill-posed problem convergence rates for regularized solutions of the form $\|x^\dagger - x^\delta_{\alpha(\delta)}\| = \mathcal{O}(f(\delta))$ can only be achieved if the solutions $x^\dagger$ (or, equivalently, the exact right hand sides) satisfy certain abstract smoothness conditions (see, e. g. [8]). In these "source conditions" it is required that the exact solution $x^\dagger$ is in the range of a certain operator. For instance if $x^\dagger$ has the representation $x^\dagger = (A^*A)^\nu v$ with $\nu \leq 1$, then it can be shown that the approximation error for Tikhonov regularization (and many other methods) behaves as $\left\|x^\dagger - x_\alpha\right\| \leq \|v\|\,\alpha^\nu$.

In order to obtain convergence rates for (1.1), we have to transfer these smoothness conditions to a stochastic setting. We do this point-wise with respect to ω in combination with a requirement that the probability of the norm of the *source function* v being large decays appropriately.

First of all we have to introduce appropriate smoothness sets. Convergence rates under rather general source conditions were proven in [21]. Following the notation in [21] we consider sets defined by functions f with the following properties:

$$\left.\begin{array}{ll} 1. & \lim_{\lambda\to 0} f(\lambda) = 0 \\ 2. & f(\lambda) \text{ is strictly monotonically increasing} \\ 3. & \lambda f^{-1}(\lambda) \text{ is convex on } \mathbb{R}^+. \end{array}\right\} \tag{3.1}$$

If f satisfies conditions (3.1) we define the smoothness set $X_{f,\tau}(\omega)$ as

$$X_{f,\tau}(\omega) := \{z \in X \mid z = f\big(A(\omega)^*A(\omega)\big)v,\ \|v\| \leq \tau\}\,. \tag{3.2}$$

In order to obtain convergence rates in a stochastic setting we assume that $x^\dagger(\omega) \in \bigcup_{\tau>0} X_{f,\tau}(\omega)$ almost surely, and moreover, that the probability of $x^\dagger(\omega) \notin X_{f,\tau}(\omega)$ is small for large τ. This probability is measured by a function g, which will influence the resulting convergence rates.

Definition 3.1. *Let $X_{f,\tau}(\omega)$ be a smoothness set as in* (3.2) *with a function f satisfying* (3.1)*. We say that a stochastic source condition for $x^\dagger$ holds, if there exists a decreasing function $g(\tau)$ such that*

$$\mu\{\omega \mid x^\dagger(\omega) \in X_{f,\tau}(\omega)'\} \leq g(\tau)\,. \tag{3.3}$$

Section 5 is devoted to a detailed discussion of this stochastic source condition for the example of an integral operator.

We want to lift convergence rate results from the deterministic case to the stochastic case. In order to keep our results general, we introduce the function h, which measures the speed of convergence, that can be obtained with Tikhonov-regularization, when $x^\dagger$ is in the range of $f(A^*A)$. In many cases (e. g., for $f(\lambda) = \lambda^\nu$, $\nu \le 1$) we have $f \equiv h$, but this need not be the case, e. g., if *saturation* occurs (cf. e. g. [8]).

Definition 3.2. *Let f satisfy* (3.1)*. For a decreasing function h, we say that f allows the deterministic convergence rate h, if for any continuous linear operator $A : X \to Y$, any $x^\dagger$ and $x_\alpha = (A^*A + \alpha I)^{-1}A^*Ax^\dagger$ we have*

$$x^\dagger \in \{z \in X \mid z = f\big(A^*A\big)v,\ \|v\| \le \tau\} \Rightarrow \big\|x^\dagger - x_\alpha\big\| \le \tau h(\alpha). \tag{3.4}$$

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 [13], respectively.

In the next theorem we show that deterministic convergence rates together with the stochastic source condition (3.3) yield convergence rates in the Prokhorov metric:

Theorem 3.3 (Approximation Error). *Let $x^\dagger(\omega)$ satisfy a stochastic source condition* (3.3)*. Assume that f allows the deterministic convergence rate h as in (3.4). Then the distance of $\mu_{x^\dagger}$ and μ_{x_α} in the Prokhorov metric is bounded by*

$$\rho(\mu_{x^\dagger}, \mu_{x_\alpha}) \le \inf_{\tau \ge 0} \max\left\{\tau h(\alpha),\ g(\tau)\right\}\,. \tag{3.5}$$

Proof. In the following let τ be arbitrary, but fixed. For any Borel-set B,

$$\begin{aligned}\mu_{x^\dagger}(B) &= \mu\{\omega \mid x^\dagger(\omega) \in B\} = \\ &= \mu\{\omega \mid x^\dagger(\omega) \in B \cap X_{f,\tau}(\omega)\} + \mu\{\omega \mid x^\dagger(\omega) \in B \cap X_{f,\tau}(\omega)'\}.\end{aligned} \tag{3.6}$$

These two terms can now be estimated separately. Using the stochastic source condition (3.3) we obtain for the second term that

$$\mu\{\omega \mid x^\dagger(\omega) \in B \cap X_{f,\tau}(\omega)'\} \le g(\tau)\,.$$

In order to estimate the first term we use results from the deterministic theory. First of all we observe that, using the notation $G_\alpha(\omega) := (A(\omega)^*A(\omega) + \alpha I)^{-1}(A(\omega)^*A(\omega))$ we have (with a potentially set-valued $G_\alpha^{-1}(\omega)$)

$$\begin{aligned}&\mu\{\omega \mid x^\dagger(\omega) \in B \cap X_{f,\tau}(\omega)\} \le \\ &\quad \le \mu\{\omega \mid x^\dagger(\omega) \in G_\alpha^{-1}(\omega)G_\alpha(\omega)(B \cap X_{f,\tau}(\omega))\} \\ &\quad = \mu\{\omega \mid x_\alpha(\omega) \in G_\alpha(\omega)(B \cap X_{f,\tau}(\omega))\}\,.\end{aligned}$$

Since f allows the deterministic convergence rate h we have for $z \in X_{f,\tau}(\omega)$

$$\begin{aligned}\|G_\alpha(\omega)z - z\| &= \left\| \left[\left(A(\omega)^* A(\omega) + \alpha I \right)^{-1} \left(A(\omega)^* A(\omega) \right) - I \right] z \right\| \\ &= \left\| \alpha (A(\omega)^* A(\omega) + \alpha I)^{-1} z \right\| \le \tau h(\alpha)\,,\end{aligned}$$

where the last term is independent of ω. Hence, if $z \in (B \cap X_{f,\tau}(\omega))$ then the distance of $G_\alpha(\omega)z$ to B is at most $\tau h(\alpha)$, i. e., $G_\alpha(\omega)(B \cap X_{f,\tau}(\omega)) \subset B^{\tau h(\alpha)}$. Hence, altogether we find that

$$\mu_{x^\dagger}(B) \le \mu_{x_\alpha}(B^{\tau h(\alpha)}) + g(\tau)\,,$$

which holds for arbitrary choices of $\tau \ge 0$. By taking the infimum over τ we obtain (3.5). □

Since one of the functions in 3.5 is decreasing in τ and the other one is increasing, the point where the infimum is attained can be found by equating the two terms.

Remark 3.4 (Hölder source conditions). Let us exemplify the above convergence rate result for the case of Hölder source conditions. It is well known for Tikhonov regularization (see, e. g., [8, Example 4.15]) that $f = \lambda^\nu$, $\nu \le 1$ allows deterministic convergence rates with $h(\alpha) = \alpha^\nu$. Assume that for this choice of f, the stochastic source condition (3.3) is satisfied with $g(\tau) = c_1 \tau^{-e(\nu)}$, with some exponent $e(\nu)$ monotonically decreasing with ν and some constant c_1. We then obtain from (3.5) that

$$\rho(\mu_{x^\dagger}, \mu_{x_\alpha}) \le c_2 \alpha^\gamma \quad \text{with} \quad \gamma = \frac{\nu e(\nu)}{e(\nu) + 1}$$

holds with some appropriate constant c_2. In particular, if the source condition is fulfilled with an exponential decay rate, e. g., $g(\tau) = e^{-\tau}$, we obtain the rate $\mathcal{O}(\alpha^\gamma)$ for any $\gamma < \nu$.

Hence, for these types of source condition we can expect Hölder rates in (3.5), and the Hölder exponent approaches the one from the deterministic setting when the probability of large τ, i. e. of source functions with large norm, gets smaller.

Remark 3.5 (Deterministic Framework). The estimate (3.5) is sharp in the sense that for the deterministic setting, the well-known convergence rates for that case (see, e. g. [8]) are obtained as special case: If A and x are deterministic then the associated measure μ_x is a point measure. Suppose that $x^\dagger$ satisfies a source condition of the type $x^\dagger = f(A^*A)v$ with some v; then (3.3) becomes

$$\mu\{\omega \mid x^\dagger(\omega) \in X_{f,\tau}{}'\} = \begin{cases} 1 & \tau < \|v\| \\ 0 & \tau \geq \|v\| \ . \end{cases}$$

Since we are interested in asymptotic convergence rates, we may choose α small enough such that $\|v\|\, h(\alpha) \leq 1$. In this case the infimum in (3.5) is attained at $\tau = \|v\|$. Hence,

$$\rho(\mu_{x^\dagger}, \mu_{x_\alpha}) \leq \|v\|\, h(\alpha),$$

which is the well-known convergence rate for the deterministic case.

4 Convergence Rates for Noisy Data

In order to estimate the total error $\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}})$, we split this term into three parts: the approximation error (Theorem 3.3), the modelling error (Lemma 4.1), and the propagated data error (Lemma 4.2). The resulting estimate is presented in Theorem 4.3 and interpreted for the case of stochastic noise in Remark 4.5.

If the exact data and operator y and A are themselves random variables and perturbed by additional noise resulting in noisy data y^δ and an inexact operator A^ξ, we have to impose an additional decay condition controlling the probability that y and A differ from some deterministic quantities (e. g., their means) A_0 and y_0 : Let β be some monotonically decreasing, right continuous function with $\beta(\vartheta) \to 0$ for $\vartheta \to \infty$ such that

$$\mu\{\omega \mid \max\{\|A(\omega) - A_0\|\, , \|y(\omega) - y_0\|\} > \vartheta\} \leq \beta(\vartheta) \tag{4.1}$$

holds. If A does not depend on ω or $A(\omega)$ is uniformly bounded, then this condition is implied by the stochastic source condition (3.3).

The following two Lemmata are modifications of results in [11]. In contrast to [11], where the error in y and the error in A were combined into one overall error, we now treat these two terms separately, which is advantageous for obtaining optimal convergence rates.

Lemma 4.1. *Let $x_\alpha(\omega)$ be as in Definition 3.2 and $x_{\alpha,A^\xi}(\omega)$ denote the Tikhonov-regularized solution with exact data y and noisy operator A^ξ, i. e., $x_{\alpha,A^\xi}(\omega) := (A^\xi(\omega)^* A^\xi(\omega) + \alpha I)^{-1} A^\xi(\omega)^* y(\omega)$. Furthermore let $A(\omega)$ and*

$y(\omega)$ fulfill the decay condition (4.1)*. Then we have, with $\rho_A := \rho(\mu_A, \mu_{A^\xi})$,*

$$\begin{aligned}
&\rho(\mu_{x_\alpha}, \mu_{x_{\alpha,A^\xi}}) \le \\
&\quad \le \inf_{\vartheta>0} \max\{\rho_A\, \tilde{h}_\alpha\big(\max\{\|A_0\| + \rho_A,\ \|y_0\|\} + \vartheta\big),\ \rho_A + \beta(\vartheta)\} \\
&\quad \le \rho_A \max\{\tilde{h}_\alpha\big(\max\{\|A_0\| + \rho_A,\ \|y_0\|\} + \beta^{-1}(\rho_A)\big),\ 2\}\,,
\end{aligned}$$

where $\tilde{h}_\alpha$ is defined as

$$\tilde{h}_\alpha(\lambda) := \frac{\lambda}{\alpha} + 2\frac{\lambda^2}{\alpha^{3/2}}\,. \tag{4.2}$$

Proof. According to [11] we have

$$\begin{aligned}
&\big\|(A(\omega)^* A(\omega) + \alpha I)^{-1} A(\omega)^* - (A^\xi(\omega)^* A^\xi(\omega) + \alpha I)^{-1} A^\xi(\omega)^*\big\| \le \\
&\quad \le \frac{\big\|A(\omega) - A^\xi(\omega)\big\|}{\alpha} + 2\max\{\|A(\omega)\|\,, \big\|A^\xi(\omega)\big\|\}\frac{\big\|A(\omega) - A^\xi(\omega)\big\|}{\alpha^{3/2}}\,.
\end{aligned}$$

This yields the Lipschitz estimate

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

for the operator $F(A, y) := (A^*A + \alpha I)^{-1} A^* y$ and $\tilde{h}_\alpha$ as in (4.2). The proof now follows with Theorem 2.1 and the observation that for joint distributions (A, y), (B, y) we have $\rho(\mu_{(A,y)}, \mu_{(B,y)}) = \rho(\mu_A, \mu_B)$. □

Lemma 4.2. *Let $x_{\alpha,A^\xi}(\omega)$ be defined as in Lemma 4.1 and $x^\delta_{\alpha,A^\xi}(\omega)$ be as in* (1.2)*. Then we have*

$$\rho(\mu_{x_{\alpha,A^\xi}}, \mu_{x^\delta_{\alpha,A^\xi}}) \le \max\left\{\frac{1}{2\sqrt{\alpha}},\ 1\right\}\rho(\mu_y, \mu_{y^\delta})\,.$$

Proof. Spectral theory yields that for any bounded linear operator A,

$$\big\|(A^*A + \alpha I)^{-1} A^*\big\| \le \frac{1}{2\sqrt{\alpha}}\,, \tag{4.3}$$

which yields

$$\big\|F(A^\xi(\omega), y(\omega)) - F(A^\xi(\omega), y^\delta(\omega))\big\| \le \frac{1}{2\sqrt{\alpha}}\,\big\|y(\omega) - y^\delta(\omega)\big\|\,,$$

with F as in Lemma 4.1. The proof now follows as above with (4.3). □

A combination of Theorem 3.3, Lemma 4.1 and Lemma 4.2 yields our main result. Applications of this theorem are presented in Remark 4.5 and Section 5.

Theorem 4.3. *Let the assumptions of Theorem 3.3 and Lemma 4.1 be fulfilled. Then we have*

$$\begin{aligned}
&\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}}) \leq \\
&\quad \leq \inf_{\tau\geq 0} \max\left\{\tau h(\alpha),\, g(\tau)\right\} + \\
&\qquad + \inf_{\vartheta>0} \max\{\rho_A\, \tilde{h}_\alpha\big(\max\{\|A_0\| + \rho_A,\, \|y_0\|\} + \vartheta\big),\, \rho_A + \beta(\vartheta)\} \\
&\qquad + \max\left\{\frac{1}{2\sqrt{\alpha}},\, 1\right\} \rho(\mu_y, \mu_{y^\delta}).
\end{aligned}$$

If the operator A is known exactly (i. e., $A^\xi = A$, A may still be random!), then the decay condition (4.1), which is used in Lemma 4.1 only, is superfluous, and the theorem above simplifies to the following result:

Corollary 4.4. *Let the assumptions of Theorem 3.3 be fulfilled and the (possibly random) operator A be given without noise. Then we have the convergence rate*

$$\rho(\mu_{x^\dagger}, \mu_{x^\delta_\alpha}) \leq \inf_{\tau\geq 0} \max\left\{\tau h(\alpha),\, g(\tau)\right\} + \max\left\{\frac{1}{2\sqrt{\alpha}},\, 1\right\} \rho(\mu_y, \mu_{y^\delta})\,.$$

Remark 4.5 (Stochastic noise). We now consider the special case where A, x and y are deterministic, but contaminated by stochastic noise, which renders A^ξ and y^δ being stochastic. Conditions (3.3) and (4.1) are therefore fulfilled with g and β being characteristic functions. The regularized solutions $x^\delta_{\alpha,A^\xi}(\omega)$ are random variables. Assume that $x^\dagger$ is such that $h(\alpha) = \alpha^\nu$ with $\nu \leq 1$ (cf. Remark 3.4). Combining Remark 3.5 and Theorem 4.3 yields the error estimate (with v as in Remark 3.5, some $C > 0$, $\rho_A := \rho(\mu_A, \mu_{A^\xi})$ and $\rho_y := \rho(\mu_y, \mu_{y^\delta})$)

$$\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}}) \leq \|v\|\, \alpha^\nu + \frac{C}{\alpha^{3/2}}\rho_A + \frac{1}{2\sqrt{\alpha}}\rho_y\,.$$

An appropriate choice for the regularization parameter α to obtain convergence as $\rho_A \to 0$ and $\rho_y \to 0$ depends on the relation between these two "input" errors. If $\rho_y = \mathcal{O}\big(\rho_A^{(2\nu+1)/(2\nu+3)}\big)$ we take $\alpha \sim \rho_A^{2/(2\nu+3)}$ which yields the convergence rate $\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}}) = \mathcal{O}\big(\rho_A^{2\nu/(2\nu+3)}\big)$. If $\rho_A = \mathcal{O}\big(\rho_y^{(2\nu+3)/(2\nu+1)}\big)$, especially if there is no noise in A, we may instead choose α as

$$\alpha \sim \rho_y^{\frac{2}{2\nu+1}} \tag{4.4}$$

which yields

$$\rho(\mu_{x^\dagger}, \mu_{x^\delta_{\alpha,A^\xi}}) = \mathcal{O}\left(\rho(\mu_y, \mu_{y^\delta})^{\frac{2\nu}{2\nu+1}}\right) ,$$

i. e., we obtain the same convergence rate as in the deterministic case.

Note that (4.4) is an a-priori rule for choosing the regularization parameter like in the deterministic theory (cf. [8]), but now referring to the data error in the Prokhorov metric instead of a norm bound. At this point, our theory changes also the computations via a different parameter choice rule. The same will be true for (more realistic) a posteriori parameter choice rules like a discrepancy principle using the Prokhorov metric.

5 An Application: Reconstruction of a Randomly Blurred Image

Now we consider, as an example, the deblurring of an image with stochastic influence on the blurring kernel: Let $x^\dagger \in L^2(\mathbb{R}^2)$ be a grey-scale image which undergoes a random blurring effect (e. g. atmospheric turbulence, lens defects, defocus); the outcome of this process, y, is measured.

We can model the blurring effect by an integral operator with kernel $k(s,t,\omega)$, $k \in L^2(\mathbb{R}^2 \times \mathbb{R}^2 \times P)$, where P is a probability space and ω describes the random effects. The observed image $y(s,\omega)$ is then given by

$$Kx^\dagger(s) := \int k(s,t,\omega)x^\dagger(t)\,dt = y(s,\omega), s \in \mathbb{R}^2.$$

Since in general the real blurring kernel is unknown, this model is replaced by one with some averaged blurring kernel $k_0 \in L^2(\mathbb{R}^2 \times \mathbb{R}^2)$:

$$K_0x := \int k_0(s,t)x(t)\,dt = y(s,\omega). \tag{5.1}$$

An approximation to the exact image can now be obtained by solving equation (5.1) using Tikhonov regularization

$$x_\alpha(\omega) := (K_0^*K_0 + \alpha I)^{-1}K_0^*y(\cdot,\omega).$$

Note that $x_\alpha(\omega)$ depends on the random variable ω. In the presence of measurement errors, only a noisy version y^δ of y is available, and hence we obtain

$$x_\alpha^\delta(\omega) := (K_0^*K_0 + \alpha I)^{-1}K_0^*y^\delta(\cdot,\omega)\,. \tag{5.2}$$

Using the results of the previous sections, we can estimate the Prokhorov distance between x_α^δ and $x^\dagger$ in terms of the Prokhorov distance between the random kernel $k(s,t,\omega)$ and the estimated kernel $k_0(s,t)$ and the noise level $\rho(y,y^\delta)$.

The stochastic source condition (3.3) can be interpreted for Gaußian kernels:

Assumption 5.1. Assume that the blurring operator K has an isotropic blurring kernel $k \in L^2(\mathbb{R}^2 \times \mathbb{R}^2 \times P)$ in the form of a Gaußian

$$k(s,t,\omega) = \gamma(s-t,\sigma(\omega)) = \frac{1}{2\pi\sigma(\omega)^2}\exp\left(\frac{-|s-t|^2}{2\sigma(\omega)^2}\right)$$

and let K_0 be the integral operator with kernel $k_0(s,t) := \gamma(s-t,\sigma_0)$.

Under this assumption the stochastic source condition (3.3) can be interpreted as a smoothness condition on $x^\dagger$ and a condition on the tail of the distribution of $\sigma(\omega)$, i. e., the width of the blurring kernel:

Lemma 5.2. *Let $x^\dagger \in H^{2\nu}(\mathbb{R}^2)$, the blurring kernel k as in Assumption 5.1 and assume*

$$\mu\{\omega \mid \sigma(\omega)^{2\nu} \geq \tau\} \leq \tilde{g}(\tau). \tag{5.3}$$

Then $x^\dagger$ fulfills a stochastic source condition (3.3) *with a function $f(\lambda) = (-\log(\frac{\lambda}{e}))^{-\nu}$ for $\lambda < 1$ satisfying* (3.1) *and $g(\tau) = \tilde{g}(\frac{\tau}{\pi^{2\nu}\|x\|_{H^{2\nu}}})$.*

Proof. In the Fourier domain, $f(K^*K)$ can be written as multiplication operator: $f(K^*K)x = f(K^2)x = \mathcal{F}^{-1}(f(g_\sigma(\cdot))^2\mathcal{F}x)$, where

$$g_\sigma(\xi) = \mathcal{F}\gamma(\cdot,\sigma(\omega)) = \exp(-\tfrac{1}{2}\xi^2\pi^2\sigma(\omega)^2)$$

and $\mathcal{F}$ denotes the Fourier transform. Therefore, (3.3) is equivalent to

$$\mu\{\omega| \|\frac{(\mathcal{F}x^\dagger)(\xi)}{f(\exp(-\xi^2\pi^2\sigma^2))}\| \geq \tau\} \leq g(\tau)\,.$$

We choose $f(\lambda) = (-\log(\frac{\lambda}{e}))^{-\nu}$ for $\lambda \leq 1$ and extend f to $\mathbb{R}^+$ such that (3.1) holds. Since $g_\sigma(\xi) \leq 1$ this extension is irrelevant for the source condition yielding $f(\exp(-\xi^2\pi^2\sigma(\omega)^2)) = (\pi^2|\xi|^2\sigma^2+1)^{-\nu}$. The above estimate now reads

$$\mu\{\omega| \left\|(\mathcal{F}x^\dagger)(\xi)(|\xi|^2+(\pi\sigma)^{-2})^\nu\right\| \sigma(\omega)^{2\nu}\pi^{2\nu} \geq \tau\} \leq g(\tau)\,.$$

Since for $x^\dagger \in H^{2\nu}(\mathbb{R}^2)$, $\|\mathcal{F}x^\dagger(|\xi|^2+(\pi\sigma)^{-2})^\nu\| \leq \|x^\dagger\|_{H^{2\nu}}$, we obtain that a stochastic source condition (3.3) is fulfilled if

$$\mu\{\omega \mid \sigma(\omega)^{2\nu} \geq \frac{\tau}{\pi^{2\nu}\left\|x\right\|_{H^{2\nu}}}\} \leq g(\tau)\,. \qquad \square$$

So if x is smooth enough and the probability of $\sigma(\omega)$ being large decays sufficiently fast, we can obtain convergence rates for $\rho(\mu_{x^\dagger}, \mu_{x_\alpha})$ using the results from the previous sections:

We estimate the distance of the operators K and K_0 in terms of the parameter σ. The proof uses again a lifting argument like in Section 2:

Lemma 5.3. *Let the blurring operators K and K_0 be as in Assumption 5.1. Then, if $\sigma_0 > \rho(\sigma, \sigma_0)$, we have the estimate*

$$\rho(K, K_0) \leq 2e^{-1} \frac{1}{\sigma_0 - \rho(\sigma, \sigma_0)} \rho(\sigma, \sigma_0) .$$

Proof. First of all we construct an estimate for $\|K - K_0\|$ in dependence of $\sigma(\omega)$, where we fix ω and write σ for $\sigma(\omega)$. The operator norm of K can be estimated via the L^2-norm of the kernel γ defined in Assumption 5.1. The probability measure associated to K_0 is the point measure at σ_0. Using the Fourier transform and the Plancherel theorem we conclude that the operator norm of the convolution operator $\|K\|_{L^2 \to L^2}$ can be written as

$$\|K\|_{L^2 \to L^2} := \sup_{\|x\|_{L^2} \leq 1} \|Kx\| = \sup_{\|\mathcal{F}x\|_{L^2} \leq 1} \|\mathcal{F}\gamma \mathcal{F}x\| = \|\mathcal{F}\gamma\|_\infty .$$

Hence, for the difference of K and K_0 we obtain

$$\begin{aligned} \|K - K_0\|_{L^2 \to L^2} &= \|\mathcal{F}\gamma(\cdot, \sigma) - \mathcal{F}\gamma(\cdot, \sigma_0)\|_{L^\infty} = \\ &= \left\|\exp(-\tfrac{1}{2}\xi^2\pi^2\sigma^2) - \exp(-\tfrac{1}{2}\xi^2\pi^2\sigma_0^2)\right\|_{L^\infty} \\ &\leq 2e^{-1} \max\{\frac{1}{|\sigma|}, \frac{1}{|\sigma_0|}\} |\sigma - \sigma_0|. \end{aligned} \tag{5.4}$$

This estimate can now be lifted to the stochastic case by a slight modification of the arguments in [11] or, more general, of Section 2. Since here, the Lipschitz constant does not grow for $\sigma \to \infty$, but for $\sigma \to 0$, we cannot apply Theorem 2.1 directly to lift this estimate, but have to modify the corresponding proof slightly: Instead of intersecting with the set $\mathcal{B}(z_0, \vartheta)$, we now use the complement of this set, centered around zero. Furthermore, as in the proof of Corollary 2.2 we can take advantage of the fact that μ_{K_0} is a point-measure, and choose the particular radius $\vartheta = \sigma_0$. The proof now follows with similar arguments as (2.7), (5.4) implies the claimed estimate. □

We now obtain the main result of this section:

Theorem 5.4. *Let $x^\dagger \in H^{2\nu}(\mathbb{R}^2)$, the blurring kernel k be as in Assumption 5.1 and $\sigma_0 > 0$. Furthermore suppose that $\sigma(\omega)$ satisfies* (5.3) *with $\tilde{g}(\tau) = \tau^{-m}$. Then we have, for x_α^δ as in* (5.2)*, the estimate*

$$\rho(\mu_{x^\dagger}, \mu_{x_\alpha^\delta}) = \mathcal{O}\left((-\log(\tfrac{\alpha}{e}))^{-\nu\frac{m}{m+1}}\right) + \mathcal{O}\left(\frac{\rho(\sigma,\sigma_0)}{\alpha^{3/2}}\right) + \mathcal{O}\left(\frac{\rho(y,y^\delta)}{\sqrt{\alpha}}\right)$$

for $\alpha \to 0$, $\rho(y, y^\delta) \to 0$ and $\rho(\sigma, \sigma_0) \to 0$.

Proof. According to Lemma 5.2, $x^\dagger$ satisfies a stochastic source condition with $f(\lambda) = (-\log(\frac{\lambda}{e}))^{-\nu}$ for $\lambda \le 1$. By [13] we know that $f(\lambda)$ allows a deterministic convergence rate (3.4) with $h(\alpha) = C(-\log(\frac{\alpha}{e}))^{-\nu}$ for α sufficiently small, hence the assumptions of Theorem 3.3 are satisfied. Using the assumed decay condition on σ we find that

$$\inf_{\tau \ge 0} \max\{\tau C(-\log(\tfrac{\alpha}{e}))^{-\nu},\, \tilde{g}(\tau)\} = \mathcal{O}(-\log(\tfrac{\alpha}{e}))^{-\nu\frac{m}{m+1}}.$$

For an application of Theorem 4.3 to our example we have to check condition (4.1). Since $\|K(\omega)\| = 1$ and $\|y(\omega)\| \le \|K(\omega)\|\|x^\dagger\| \le \|x^\dagger\|$, this condition holds with $A_0 = 0$, $y_0 = 0$, and $\beta(\vartheta) = 0$ for $\vartheta > \max\{1, \|x^\dagger\|\}$. Now the result follows from Theorem 4.3, Lemma 5.3 and the definition of $\tilde{h}_\alpha$ in (4.2):

$$\begin{aligned}
&\inf_{\vartheta > 0} \max\{\rho(K,K_0)\,\tilde{h}_\alpha\big(\rho(K,K_0)+\vartheta\big),\, \rho(K,K_0)+\beta(\vartheta)\} \\
&\quad = e^{-1}\frac{\rho(\sigma,\sigma_0)}{\sigma_0-\rho(\sigma,\sigma_0)} \max\{\tilde{h}_\alpha(e^{-1}\frac{\rho(\sigma,\sigma_0)}{\sigma_0-\rho(\sigma,\sigma_0)} + \max\{1, \left\|x^\dagger\right\|\}),\, 1\} \\
&\quad = \frac{\rho(\sigma,\sigma_0)}{\alpha^{3/2}}\mathcal{O}(1) = \mathcal{O}\left(\frac{\rho(\sigma,\sigma_0)}{\alpha^{3/2}}\right)
\end{aligned}$$

□

This estimate also gives rise to an a priori choice for the regularization parameter α.

Remark 5.5. If $\sigma(\omega)$ satisfies (5.3) with an exponentially decaying $\tilde{g}(\tau)$, the first term in the estimate of Theorem 5.4 can be replaced by $\mathcal{O}((-\log(\alpha))^\gamma)$ for any $\gamma < \nu$. This holds, for instance, in the important case that $\sigma(\omega)$ is normally distributed.

Acknowledgments

This work has been supported by the Austrian National Science Foundation FWF under project grant SFB F 13/08.

References

[1] H. T. Banks and K. L. Bihari, *Modelling and estimating uncertainty in parameter estimation*, Inverse Probl., 17, No.1 (2001), pp. 95–111.

[2] M. Bertero and G. A. Viano, *On probabilistic methods for the solution of improperly posed problems*, Bollettino U.M.I., (1978), pp. 453–508.

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

[4] N. Bissantz, T. Hohage, and A. Munk, *Nonlinear Tikhonov regularization for statistical inverse problems*, Inverse Problems, (accepted).

[5] R. M. Dudley, *Real Analysis and Probability*, Wadsworth, 1989.

[6] H. W. Engl, *A general stochastic fixed-point theorem for continuous random operators on stochastic domains*, Journal of Mathematical Analysis and Applications, 66 (1978), pp. 220–231.

[7] H. W. Engl and H. Gfrerer, *A posteriori parameter choice for general regularization methods for solving linear ill-posed problems*, Appl. Numer. Math, 4 (1988), pp. 395–417.

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

[9] H. W. Engl and M. Z. Nashed, *Random generalized inverses and approximate solutions of random operator equations*, in Approximate Solution of Random Equations, A. T. Bharucha-Reid, ed., North Holland, 1979, pp. 149–210.

[10] ——, *Generalized inverses of random linear operators in Banach spaces*, Journal of Mathematical Analysis and Applications, 83 (1981), pp. 582–610.

[11] 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.

[12] C. W. Groetsch and C. R. Vogel, *Asymptotic theory of filtering for linear operator equations with discrete noisy data*, Math. Comp., 49 (1987), pp. 499–506.

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

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

[15] G. Landl, T. Langthaler, H. W. Engl, and H. F. Kauffmann, *Distribution of event times in time-resolved fluorescence: the exponential series approach - algorithm, regularization, analysis*, J. Comput. Phys., 95 (1991), pp. 1–28.

[16] M. A. Lukas, *Comparison of parameter choice methods for regularization with discrete noisy data*, Inverse Probl., 14 (1998), pp. 161–184.

[17] B. A. Mair and F. H. Ruymgaart, *Statistcal inverse estimation in Hilbert scales*, SIAM Journal of Applied Mathematics, 56 (1996), pp. 1424–1444.

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

[19] E. Schock, *Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence*, in Constructive Methods for the Practical Treatment of Integral Equations, G. Hämmerlin and K. H. Hoffmann, eds., Birkhäuser, Basel, 1985, pp. 234–243.

[20] S. Smale and D.-X. Zhou, *Shannon sampling and function reconstruction from point values*, Bulletin of the American Mathematical Society, (2004), pp. 279–305.

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

[22] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., 14 (1977), pp. 651–667.

[23] A. Wakolbinger. private communication, 2004.