ON THE LIFTING OF DETERMINISTIC CONVERGENCE RATES FOR INVERSE PROBLEMS WITH STOCHASTIC NOISE

DANIEL GERTH
Technische Universität Chemnitz
Fakultät für Mathematik
D-09107 Chemnitz, Germany

ANDREAS HOFINGER
Radon Institute for Computational and Applied Mathematics (RICAM)
Altenbergerstraße 69
A-4040 Linz, Austria

RONNY RAMLAU
Radon Institute for Computational and Applied Mathematics (RICAM)
(also Industrial Mathematics Institute, Johannes Kepler University Linz)
Altenbergerstraße 69
A-4040 Linz, Austria

(Communicated by Samuli Siltanen)

Abstract. Both for the theoretical and practical treatment of Inverse Problems, the modeling of the noise is crucial. One either models the measurement via a deterministic worst-case error assumption or assumes a certain stochastic behavior of the noise. Although some connections between both models are known, the communities develop rather independently. In this paper we seek to bridge the gap between the deterministic and the stochastic approach and show convergence and convergence rates for Inverse Problems with stochastic noise by lifting the theory established in the deterministic setting into the stochastic one. This opens the wide field of deterministic regularization methods for stochastic problems without having to do an individual stochastic analysis for each problem.

1. Introduction. In Inverse Problems, the model of the inevitable data noise is of utmost importance. In most cases, an additive noise model

\begin{equation}
\label{eq:noisy_data}
y_{\text{noisy}} = y + \epsilon
\end{equation}

is assumed. In \( \mathcal{Y} \), \( y \in \mathcal{Y} \) is the true data of the unknown \( x \in \mathcal{X} \) under the action of the (in general) nonlinear operator \( F : \mathcal{X} \to \mathcal{Y} \),

\begin{equation}
\label{eq:nonlinear_operator}
F(x) = y,
\end{equation}

and \( \epsilon \) in \( \mathcal{Y} \) corresponds to the noise. The spaces \( \mathcal{X} \), \( \mathcal{Y} \) are assumed to be metric spaces with metrics \( d_X \) and \( d_Y \), respectively. In most cases \( \mathcal{X} \) and \( \mathcal{Y} \) will be Banach
spaces or Hilbert spaces such that the metric is induced by a norm. However, the technique used in this paper is not restricted to these two classes of spaces.

When speaking of Inverse Problems, we assume that (2) is ill-posed. In particular this means that solving (2) for \( x \) with noisy data (1) is unstable in the sense that “small” errors in the data may lead to arbitrarily large errors in the solution. Hence, (1) is not a sufficient description of the noise. More information is needed in order to compute solutions from the data in a stable way. In the deterministic setting, one assumes

\[ d_\mathcal{Y}(y, y^\delta) \leq \delta \]

for some \( \delta > 0 \). In a Banach space this would read \( ||y - y^\delta||_\mathcal{Y} \leq \delta \). Here and further on we use the superscript \( ^\delta \) to indicate the deterministic setting. Solutions of (2) under the assumption (1),(3) are often computed via a Tikhonov-type variational approach

\[ x^\delta_\alpha = \arg\min_{x \in \mathcal{D}(F)} d_\mathcal{Y}(F(x), y^\delta) + \alpha \Phi(x), \]

where again \( d_\mathcal{Y} \) is a distance functional and \( \Phi(\cdot) \) is the penalty term used to stabilize the problem and to incorporate a-priori knowledge into the solution. The regularization parameter \( \alpha \) is used to balance between data misfit and the penalty and has to be chosen appropriately. The literature in the deterministic setting is rich, at this point we only refer to the monographs [12, 26, 32] for an overview. Throughout this paper we assume that there is already a deterministic regularization method for (2), see Section 4.

The deterministic worst-case error stands in contrast to stochastic noise models where a certain distribution of the noise \( \epsilon \) in (1) is assumed. We shall indicate the stochastic setting by the superscript \( ^\eta \). In this paper, \( \eta \) will be the parameter controlling the variance of the noise. Depending on the actual distribution of \( \epsilon \), \( d_\mathcal{Y}(y, y^\eta) \) may be arbitrarily large, but with low probability. A popular approach to find an approximate solution to (2) is the Bayesian method. For more detailed information, we refer to [7, 28, 33, 35, 36]. In the Bayesian setting, the solution of the Inverse Problem is given as a distribution of the random variable of interest, the posterior distribution \( \pi_{\text{post}} \), determined by Bayes formula

\[ \pi_{\text{post}}(x|y^\eta) = \frac{\pi(x|y^\eta)\pi_{\text{pr}}(x)}{\pi_{y^\eta}(y^\eta)}. \]

That is, roughly spoken, all values \( x \) are assigned a probability of being a solution to (2) given the noisy data \( y^\eta \). In (5), the likelihood function \( \pi(x|y^\eta) \) represents the model for the measurement noise, i.e., its distribution. In case the operator is random (see Section 4.2) also this randomness will be encoded in \( \pi_{\text{pr}}(x) \). The prior distribution \( \pi_{\text{pr}}(x) \) represents a-priori information about the unknown or its distribution, whether the exact solution itself carries randomness or is deterministic. The data distribution \( \pi_{y^\eta}(y^\eta) \) as well as the normalization constants are usually neglected since they only influence the normalization of the posterior distribution. In practice one is often more interested in finding a single solution instead of the distribution itself. Popular point estimates are the conditional expectation (conditional mean, CM)

\[ \mathbb{E}(\pi_{\text{post}}(x|y^\eta)) = \int x\pi_{\text{post}}(x|y^\eta)dx \]
and the maximum a-posteriori (MAP) solution

\[ x_{\text{MAP}} = \arg\max_x \pi_{\text{post}}(x|y^n), \]

i.e., the most likely value for \( x \) under the prior distribution and given the data \( y^n \). Both point estimators are widely used. The computation of the CM-solution is often slow since it requires repeated sampling of stochastic quantities and the evaluation of high-dimensional integrals. The MAP-solution, however, essentially leads to a Tikhonov-type problem. Namely, assuming \( \pi(y^n|x) \propto \exp(-d_Y(F(x), y^n)) \) and \( \pi_{pr}(x) \propto \exp(-\alpha \Phi(x)) \), one has

\[ x_{\text{MAP}} = \arg\max_x \exp(-d_Y(F(x), y^n)) \exp(-\alpha \Phi(x)) \]

\[ = \arg\min_x d_Y(F(x), y^n) + \alpha \Phi(x), \]

analogously to (4). The MAP is usually defined in finite dimensions as in an infinite dimensional setting the existence of (6) is not clear. In general a weaker formulation is required in infinite dimensions, see [21].

Also non-Bayesian stochastic approaches for Inverse Problems often seek to minimize a functional (4), see, e.g., [3, 27] or use techniques known from deterministic theory such as filter methods [4, 5]. Finally, Inverse Problems appear in the context of statistics. Hence, the statistics community has developed methods to solve (2), partly again based on the minimization of (4). We refer to [13] for an overview.

In summary, Tikhonov-type functionals (4) and other deterministic methods frequently appear also in the stochastic setting. From a practical point of view, one would expect to be able to use deterministic regularization methods for (2) even when the noise is stochastic. Indeed, the main question for the actual computation of the solution, given a particular sample of noisy data \( y^n \), is the choice of the regularization parameter. A second question, mostly coming from the deterministic point of view, is the one of convergence of the solutions when the noise approaches zero. In the stochastic setting these questions are answered often by a full stochastic analysis of the problem. In this paper we present a framework that allows to find appropriate regularization parameters, prove convergence of regularization methods and find convergence rates for Inverse Problems with a stochastic noise model by directly using existing results from the deterministic theory.

We take several ideas from the dissertation [22], which is only publicly available as book [23] and not published elsewhere. The paper is organized as follows. The Ky Fan metric, which will be the main ingredient of our analysis, and its relation to the expectation, will be introduced in Section 2. In Section 3 we discuss an issue occurring in the transition from deterministic to stochastic noise for infinite dimensional problems. We present our framework to lift convergence results from the deterministic setting into the stochastic setting in Section 4. In the first part we generalize results from [22] such that they can be applied to any deterministic regularization method when only the noise is assumed to be random. We also show that not only the Hilbert space case from [22] can be dealt with but that the lifting is in principle possible for problems in metric spaces. Additionally we discuss the expectation as quantification of the magnitude of the noise rather than only the Ky Fan metric. The relation between Ky Fan metric and the expectation also shows a way to prove convergence in expectation of the regularized solutions. We also include a discussion on the case that \( F \) is not injective. In the second part we demonstrate the lifting strategy when also the unknown and/or the operator are
allowed to carry randomness. Since a general lifting theory seems no longer possible we demonstrate the technique via two examples, taken from [22]. In Section 5 some examples for the lifting technique are given.

2. The Ky Fan metric. Throughout the the paper we assume \((\Omega, \mathcal{F}, \mathbb{P})\) to be a complete probability space with a set \(\Omega\) of outcomes of the stochastic event, \(\mathcal{F}\) the corresponding \(\sigma\)-algebra and \(\mathbb{P}\) a probability measure, \(\mathbb{P}: (\Omega, \mathcal{F}) \rightarrow [0,1]\). We restrict ourselves here to probability measures for the sake of simplicity. Extensions to more general measures are straightforward.

In order to measure the magnitude of the stochastic noise and the quality of the reconstructions, we need metrics that incorporate the stochastic nature of the problem. One such metric, which will be the the main tool for our stochastic convergence analysis, is the Ky Fan metric (cf. [14]). It is defined as follows.

**Definition 2.1.** Let \(X_1\) and \(X_2\) be random variables in a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with values in a metric space \((\chi, d)\). The distance between \(X_1\) and \(X_2\) in the Ky Fan metric is defined as

\[
\rho_K(X_1, X_2) := \inf_{\varepsilon > 0} \{\mathbb{P}(\{\omega \in \Omega : d_\chi(X_1(\omega), X_2(\omega)) > \varepsilon\}) < \varepsilon\}.
\]

A comprehensive summary on the Ky Fan metric can be found in [24]. We will often drop the explicit reference to \(\omega\) in (7). This metric essentially allows to lift results from a metric space to the space of random variables as the connection to the deterministic setting is inherent via the metric \(d_\chi\) used in its definition. We will implicitly assume that equation (2) is scaled appropriately since \(\rho_K(X_1, X_2) \leq 1\) for all \(X_1, X_2\).

An immediate consequence of (7) is that \(\rho_K(X_1, X_2) = 0\) if and only if \(X_1 = X_2\) almost surely. Convergence in the Ky Fan metric is equivalent to convergence in probability, i.e., for a sequence \(\{X_k\}_{k \in \mathbb{N}}\) of random variables taking values in \(\mathcal{X}\) and \(X \in \mathcal{X}\) one has

\[
\rho_K(X_k, X) \xrightarrow{k \to \infty} 0 \quad \Leftrightarrow \quad \forall \varepsilon > 0: \quad \mathbb{P}(d_\mathcal{X}(X_k, X) > \varepsilon) \xrightarrow{k \to \infty} 0.
\]

Hence convergence in the Ky Fan metric also leads to pointwise (almost sure) convergence of certain subsequences in the metric \(d_\chi\).

A somewhat more intuitive and more frequently used stochastic metric is the expectation, or more general, a (stochastic) \(L_p\) metric. For random variables \(Y_1\) and \(Y_2\) with values in a metric space \((\chi, d_\chi)\),

\[
\mathbb{E}(d_\chi(Y_1, Y_2)^p) = \int_{\Omega} d_\chi(Y_1, Y_2)^p d\mathbb{P}(\omega)
\]
defines the \(p\)-th moment of \(d_\chi(Y_1, Y_2)\) for \(p \geq 1\), assuming the existence of the integral. We will use \(p = 1\) and refer to it as convergence in expectation. Note that since the variance is defined as

\[
\text{Var}(d_\chi(Y_1, Y_2)) = \mathbb{E}(d_\chi(Y_1, Y_2)^2) - \mathbb{E}(d_\chi(Y_1, Y_2))^2 \geq 0
\]
one always has

\[
\mathbb{E}(d_\chi(Y_1, Y_2)) \leq \sqrt{\mathbb{E}(d_\chi(Y_1, Y_2)^2)}.
\]
We will show later that for parameter choice rules the expectation of the noise has to be slightly overestimated, hence estimating $E(d_Y(y,y^\eta))$ via the popular and often easier to compute $L_2$-norm $E((d_Y(y,y^\eta))^2)$ with \([5]\) is not problematic.

While the main part of our analysis is based on the description of the noise and the reconstruction quality in the Ky Fan metric, we will also allow the expectation as measure of the stochastic noise. We also discuss convergence of the reconstructed solutions in expectation under an additional condition in Remark 2. To this end, we comment in the following on the connection between those two metrics.

It is well-known that convergence in expectation implies convergence in probability, see for example \([10]\). Hence, convergence in the Ky Fan metric is implied by convergence in expectation (and also by convergence of higher moments). Namely, with Markov’s inequality one has, for an arbitrary nonnegative random variable $X$ with $E(X) < \infty$ and $C > 0$

\[
\mathbb{P}(X > C) \leq \frac{E(X)}{C}.
\]

The converse does not hold in general, see the example in \([24, \text{Section 3.1}].\) Under the additional assumption that a sequence is uniformly integrable (see for example \([6, \text{Definition A.3.1}].\)\), however, one can show conversely that convergence in probability implies convergence in expectation. Uniform integrability in principle means that a sequence is bounded as a whole.

**Theorem 2.2** (\([6], \text{Theorem A.3.2}.\). Let \(\{x_k\}_{k \in \mathbb{N}} \subset L_1(\mathbb{P})\) be a sequence convergent almost everywhere (or in probability) to a function $x$. If the sequence \(\{x_k\}_{k \in \mathbb{N}}\) is uniformly integrable, then it converges to $x$ in the norm of $L_1(\mathbb{P})$.

In \([24, \text{Section 3.1}].\) one can see in an example that indeed unboundedness of the random variables is the reason that the Ky Fan metric does not imply convergence in expectation. We will use this observation in Remark 2 to demonstrate how one could move from convergence in the Ky Fan metric to convergence in expectation. To close this section, let us remark on the computation of the Ky Fan distance. In general it can be estimated via the moments of the noise.

**Theorem 2.3** (\([24], \text{Theorem 3.2}.\). Let $Y_1, Y_2$ be random variables in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $E((d_Y(Y_1,Y_2)^s)) < \infty$ for some $s \in \mathbb{N}$. Then

\[
\rho_{K}(Y_1,Y_2) \leq \sqrt{s} E((d_Y(Y_1,Y_2)^s)).
\]

Note that even if the moments exist for all $s \in \mathbb{N}$

\[
\lim_{s \to \infty} \sqrt{s} E((d_Y(Y_1,Y_2)^s)) \neq E((d_Y(Y_1,Y_2))),
\]

see \([10, 22]\), due to the tail of the distributions. In the Gaussian case, the following direct estimate has been derived in \([25, 34]\).

**Proposition 1.** Let $\epsilon$ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of $\epsilon$ is $N(0,\eta^2I_m)$ with $\eta > 0$. Then it holds in $(\mathbb{R}^m, ||\cdot||_2)$ that

\[
\rho_{K}(\epsilon,0) \leq \min \left\{ 1, \sqrt{2\eta} \sqrt{m} - \min \left\{ \ln \left( \eta^2 2\pi m^2 \left( \frac{e}{2} \right)^m \right), 0 \right\} \right\}.
\]

It is easy to see that for Gaussian noise as in the previous proposition

\[
E(||\epsilon||_2) \leq \eta \sqrt{m},
\]

see, e.g., \([10]\). Comparing \((11)\) and \((12)\), one sees that $E(||\epsilon||_2) < \rho_{K}(\epsilon,0)$ and in particular the decay of $\rho_{K}(\epsilon,0)$ slows down with decreasing $\eta$. 

---

**Inverse Problems and Imaging** Volume 11, No. 4 (2017), 663–687
3. On the noise model. Before addressing the convergence theory, we would like to discuss stochastic noise modeling and its intrinsic conflict with the deterministic model. In the Hilbert space setting, the noise is typically modeled as follows, see for example [4, 29, 31, 32]. Let \( \xi : \Omega \to Y \) be a stochastic process. Then for \( y \in Y \)

\[
\langle y, \xi \rangle
\]

defines a real-valued random variable. Assuming that

\[
E((\hat{y}, \xi)^2) < \infty
\]

for all \( \hat{y} \in Y \) and that this expectation is continuous in \( \hat{y} \),

\[
E(\langle \hat{y}, \xi \rangle \langle y, \xi \rangle)
\]

defines a continuous, symmetric nonnegative bilinear form. In particular, there exists the covariance operator

\[
C : Y \to Y
\]

with

\[
\langle C \hat{y}, y \rangle = E(\langle \hat{y}, \xi \rangle \langle y, \xi \rangle).
\]

For the stochastic analysis of infinite dimensional problems via deterministic results, (13) is problematic. Namely, if \( \{u_n\}_{n \in \mathbb{N}} \) is an orthonormal basis in \( Y \), the set \( \{\langle u_n, \xi \rangle\}_{n \in \mathbb{N}} \) consists of infinitely many identically distributed random variables with \( 0 < E|\langle u_n, \xi \rangle|^2 = \text{const} < \infty \) [32]. Thus

\[
E \left( \sum_{n=1}^{\infty} |\langle u_n, \xi \rangle|^2 \right) = \infty \quad \text{a.s.}
\]

and a realization of the noise is an element of the Hilbert space \( Y \) with probability zero. Let us take the common example of Gaussian white noise which can be modeled via the above construction. Namely, with

\[
E(\langle y, \xi \rangle) = 0 \quad \forall y \in Y
\]

and the covariance operator

\[
C = \eta^2 I,
\]

where \( I \) is the identity and \( \eta > 0 \) the variance parameter, the Gaussian white noise is described [29, 32]. As consequence of (14) and explained for example in [29], a realization of such a Gaussian random variable is with probability zero an element of an infinite dimensional \( L_2 \)-space. It is therefore inappropriate to use an \( L_2 \)-norm for the residual in case of an infinite dimensional problem. Since in this case a realization of Gaussian white noise only lies (almost surely) in any Sobolev space \( H^s \) with \( s < -d/2 \) where \( d \) is the dimension of the domain, one should adjust the norm for the residual accordingly. Except for the paper [29] this issue seems not to have been addressed in the literature. A main reason for this might be that for the practical solution of the Inverse Problem this is not a severe issue since in reality the measurements are finite dimensional and, in order to use a computer to solve the problem, a finite dimensional approximation of the unknown object has to be used. In this case the sum in (14) is finite and the noise lies almost surely in the finite dimensional space. However, difficulties arise whenever one seeks to investigate convergence of the discretized problem to its underlying infinite dimensional problem. We will not address this issue and assume throughout the whole work that \( E(d_Y(\epsilon, 0)) < \infty \) or use the slightly weaker bound on the Ky-Fan metric (see Section 2). In order to handle the Ky Fan metric we need to be able to evaluate probabilities \( P(d_Y(y, y^0) > \epsilon), 0 \leq \epsilon \leq 1 \), which is only meaningful if...
y - y' =: \epsilon \in \mathcal{Y}. Assuming that \mathcal{Y} is finite dimensional, then this is clear. For infinite dimensional problems, however, we have to assume that the noise is smooth enough for the sum in (14) to converge. Examples for this are Brownian noise (1/f^2-noise) or pink noise (1/f-noise), see e.g. [15, 30]. At this point we would also like to mention that as a consequence of our rather generic noise model we might not make use of some specific properties of the noise as would be possible when focusing on a particular distribution of the noise. However, we are able to show convergence for a large variety of regularization methods.

4. Convergence in the stochastic setting.

4.1. Deterministic Inverse Problems with stochastic noise. As mentioned previously, the intention of this paper is to show convergence for Inverse Problems under a stochastic noise model using results from the deterministic setting. Assume we have at hand a deterministic regularization method of our liking for the solution of (2) under the noisy data (1) where now \( d\mathcal{Y}(y, y^\delta) \leq \delta \) for some \( \delta > 0 \). By regularization method we understand (possibly nonlinear) mappings 

\[ R_\alpha : D(R_\alpha) \subset \mathcal{Y} \to \mathcal{X}, \quad y^\delta \mapsto x^\delta \]

where \( x^\delta = R_\alpha(y^\delta) \) is the regularized solution to the regularization parameter \( \alpha \) given the data \( y^\delta \). Often, \( x^\delta \) is obtained via the minimization of functionals of the type (4). In order to deserve the name regularization we require \( R_\alpha \) to fulfill

\[ \lim_{\delta \to 0} d_\mathcal{X}(R_\alpha(y^\delta), x^\dagger) = 0 \]  

under a certain choice of the regularization parameter \( \alpha \) chosen either a priori \( \alpha = \alpha(\delta) \) or a posteriori \( \alpha = \alpha(\delta, y^\delta) \). In our notation \( x^\dagger \) is the true solution, usually the minimum norm solution with respect to the penalty \( \Phi \) in (4), i.e.,

\[ \Phi(x^\dagger) \leq \Phi(\bar{x}) \quad \text{for all} \quad \bar{x} : F(\bar{x}) = y. \]

Note that, in particular for nonlinear problems, \( x^\dagger \) does not need to be unique. In [22, 23] it was pointed out that this is problematic for the lifting arguments. A standard argument in the deterministic theory is to prove convergence of subsequences to the desired solution, and then deduces convergence of the whole series of regularized solutions, if possible. In the stochastic setting, this is not possible in general since subsequences for different \( \omega \) do not have to be related. A constructed example for this behavior can be found in Section 4.1. of [22, 23]. In order to lift general deterministic regularization methods into the stochastic setting we must therefore require that \( x^\dagger \) is unique for now. We will later demonstrate that in principle our approach can also deal with non-injective operators. Our convergence results are formulated assuming the noise to be bounded with respect to the Ky Fan metric or in expectation. For the analysis we mainly use a lifting argument using deterministic theory. In [22, 23, Theorem 4.1], it was proved how by means of the Ky Fan metric deterministic results can be lifted to the space of random variables for nonlinear Tikhonov regularization. Since the theorem is based solely on the fact that there is a deterministic regularization theory and that the probability space \( \Omega \) can be decomposed into a part where the deterministic theory holds and a small part where it does not, it is easily generalized. Before we state the Theorem, we need the following Lemmata.
Lemma 4.1. ([11], see also [10]) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Let $x_k$ and $x$ be measurable functions from $\Omega$ into a metric space $\chi$ with metric $d_\chi$. Suppose $x_k(\omega) \xrightarrow{d} x(\omega)$ for $\mathbb{P}$-almost all $\omega \in \Omega$. Then for any $\varepsilon > 0$ there is a set $\Omega'$ with $\mathbb{P}(\Omega \setminus \Omega') < \varepsilon$ such that $x_k \xrightarrow{d_\chi} x(\omega)$ uniformly on $\Omega'$, that is
\[
\lim_{k \to \infty} \sup\{d_\chi(x_k(\omega), x(\omega)) : \omega \in \Omega'\} = 0.
\]

Lemma 4.2 ([22] [23], Proposition 1.10). Let $\{x_k\}_{k \in \mathbb{N}}$ be a sequence of random variables that converges to $x$ in the Ky Fan metric. Then for any $\nu > 0$ and $\varepsilon > 0$ there exist $\Omega \subset \Omega'$, $\mathbb{P}(\Omega) \geq 1 - \varepsilon$, and a subsequence $x_{k_j}$ with
\[
d_\chi(x_{k_j}(\omega), x(\omega)) \leq (1 + \nu)\rho_K(x_{k_j}, x) \quad \forall \omega \in \Omega.
\]
Furthermore there exists a subsequence that converges to $x$ almost surely.

The proof of this Lemma can also be found in [24].

With this, we are ready for the convergence theorem which we shall split in two parts, one for the Ky Fan metric as error measure and one for the expectation. We just need to remark on a formulation used in the following theorems.

Remark 1. In the following theorems we will assume that “all required assumptions for the deterministic theory (except the bound on the noise) hold with probability one”. This formulation is used to retain the general character of our theory that is applicable to any deterministic regularization method as long as only the noise is stochastic. In general convergence and even more convergence rates require certain assumptions on both the operator $F$ and the exact solution $x^\dagger$. However these depend on the specific situation and we can not provide a list of possible conditions without losing generality. For classical Tikhonov-regularization where $X$ and $Y$ are Hilbert spaces and [11] is given with $d_2(F(x), y^\delta) = ||F(x) - y^\delta||_2$ and $\Phi(x) = ||x||_2^2$, convergence in the sense of [15] follows alone by an appropriate choice of the regularization parameter. The parameter choice is included in our lifting theory, consequently we do not need other assumptions for convergence in the stochastic setting. In order to derive convergence rates of the form $d_\chi(R_\alpha(y^\delta), x^\dagger) \leq \varphi(d_2(y, y^\delta))$ with some appropriate function $\varphi : [0, \infty) \to [0, \infty)$ additional smoothness properties are required. For classical Tikhonov regularization it is known that $||R_\alpha(y^\delta) - x^\dagger||_2 = \mathcal{O}(\delta^{\frac{\nu}{\nu + 1}})$ whenever $x^\dagger \in \text{range}(\mathcal{A}^*\mathcal{A})^\nu/2$ for some $0 < \nu \leq 2$, see, e.g., [12] [32]. In order to lift the convergence rate into the stochastic setting we have to adopt this condition. Different choices of $d_2$ or $\Phi(x)$ however would possibly require different assumptions. For nonlinear Tikhonov regularization (see Section [1.2]) for example we would require that the conditions a) to d) in Theorem 4.8 hold with probability one. Please be again reminded that in this Section we consider only the noise to be random. We will address the case of randomness in the operator and/or the solution in Section 4.2.

Theorem 4.3. Let $R_\alpha$ be a regularization method for the solution of [2] in the deterministic setting under a suitable choice of the regularization parameter. Let now $y^\delta = y + \epsilon(\eta)$ where $\epsilon(\eta)$ is a stochastic error such that $\rho_K(y, y^\delta) \to 0$ as $\eta \to 0$. Then, assuming [2] has a unique solution $x^\dagger$ and all required assumptions for the deterministic theory (except the bound on the noise) hold in the sense of Remark 7 the regularization method $R_\alpha$ fulfills
\[
\lim_{\eta \to 0} \rho_K(R_\alpha(y^\delta), x^\dagger) = 0
\]
under the same parameter choice rule as in the deterministic setting with \( \delta \) replaced by \( \rho_K(y, y^\dagger) \).

**Proof.** Denote \( x_\alpha(\eta) := R_\alpha(y^\eta) \). Define \( \theta := \limsup_{k \to \infty} \rho_K(x^\dagger, x_{\alpha(\eta_k)}) \). (Note that \( 0 \leq \theta \leq 1 \) due to the properties of the Ky Fan metric). We show in the following that for arbitrary \( \varepsilon > 0 \) we have \( \theta/2 \leq \varepsilon \) and hence

\[
\limsup_{k \to \infty} \rho_K(x^\dagger, x_{\alpha(\eta_k)}) = \lim_{k \to \infty} \rho_K(x^\dagger, x_{\alpha(\eta_k)}) = 0.
\]

As a first step we pick a “worst case” subsequence \( \{y^{\eta_k}\} \) of \( \{y^\eta\} \), a subsequence for which the corresponding solutions satisfy \( \rho_K(x^\dagger, x_{\alpha(\eta_k)}) \geq \theta/2 \). We now show that even from this “worst case” sequence we can pick a subsequence \( \{y^{\eta_k}\} \) for which we have \( \limsup \rho_K(x^\dagger, x_{\alpha(\eta_k)}) \leq \varepsilon \) for arbitrary \( \varepsilon > 0 \).

Let \( \varepsilon > 0 \). According to Lemma 4.2, we can pick a subsequence \( \{y^{\eta_k}\} \) and a set \( \tilde{\Omega} \) with \( \mathbb{P}(\tilde{\Omega}) \geq 1 - \frac{\varepsilon}{2} \) as well as \( d_Y(y(\omega), y^{\eta_k}(\omega)) \leq (1 + \nu)\rho_K(y, y^{\eta_k}), \nu > 0 \) arbitrarily small, on \( \tilde{\Omega} \). For all \( \omega \in \tilde{\Omega} \), the noise tends to zero. We can therefore use the deterministic result with \( \delta = \rho_K(y(\omega), y^{\eta_k}) \) and deduce that \( x_{\alpha(\eta_k)}(\omega) \) converges to the unique solution \( x^\dagger(\omega) \) for \( \eta_k \to 0, \omega \in \tilde{\Omega} \) where in the choice of the regularization parameter \( \delta \) is substituted by \( \rho_K(y(\omega), y^{\eta_k}) \). The convergence is not uniform in \( \omega \); nevertheless, pointwise convergence implies uniform convergence except on sets of small measure according to Lemma 4.1. Therefore there exist \( \tilde{\Omega}' \subset \tilde{\Omega}, \mathbb{P}(\tilde{\Omega}') < \frac{\varepsilon}{2} \) and \( j_0 \in \mathbb{N} \) such that \( d_X(x_{\alpha(\eta_{k_j})}(\omega), x^\dagger(\omega)) < \varepsilon \forall \omega \in \tilde{\Omega} \cap \tilde{\Omega}' \) and \( j \geq j_0 \). We thus have

\[
\mathbb{P}\left( \left\{ \omega \in \tilde{\Omega} : d_X(x_{\alpha(\eta_{k_j})}(\omega), x^\dagger(\omega)) > \varepsilon \right\} \right) \leq \mathbb{P}(\tilde{\Omega}') \leq \frac{\varepsilon}{2}.
\]

Since we split \( \Omega = \tilde{\Omega} \cup \tilde{\Omega} \cap \tilde{\Omega}' \cup \tilde{\Omega}' \) with \( \mathbb{P}(\Omega \setminus \tilde{\Omega}) < \frac{\varepsilon}{2}, \mathbb{P}(\Omega \setminus \tilde{\Omega}) + \mathbb{P}(\tilde{\Omega}') \leq \varepsilon \) we have shown existence of a subsequence \( \eta_{k_j} \) such that

\[
\mathbb{P}\left( \left\{ \omega \in \tilde{\Omega} : d_X(x_{\alpha(\eta_{k_j})}(\omega), x^\dagger(\omega)) > \varepsilon \right\} \right) \leq \varepsilon
\]

for \( \eta_{k_j} \) sufficiently small. This \( \varepsilon \) is, by definition of the Ky Fan metric, an upper bound for the distance between \( x_{\alpha(\eta_{k_j})} \) and \( x^\dagger \). Therefore we have

\[
\limsup_{j \to \infty} \rho_K(x_{\alpha(\eta_{k_j})}, x^\dagger) \leq \varepsilon.
\]

On the other hand, the original sequence satisfies \( \liminf_{j \to \infty} \rho_K(x^\dagger, x_{\alpha(\eta_k)}) \geq \theta/2 \). Since \( \liminf_{j \to \infty} \rho_K(x^\dagger, x_{\alpha(\eta_k)}) \leq \limsup_{j \to \infty} \rho_K(x_{\alpha(\eta_k)}, x^\dagger) \) it follows \( \theta/2 \leq \varepsilon \). Because \( \varepsilon > 0 \) was arbitrary, this implies \( \theta = 0 \), which concludes the proof.

As we have shown in the previous theorem convergence properties can directly be lifted into the stochastic setting by replacing \( \delta \) with the estimate of the noise in the Ky Fan metric. For the expectation this is not the case as we will see in Theorem 4.4. We have to gradually inflate the expectation by a parameter \( \tau \) in order to obtain convergence (and later convergence rates). The reason for this is that, as already demonstrated in Theorem 4.3, our strategy is to show that there is \( \Omega_\varepsilon \subset \Omega \) (in Theorem 4.3 \( \Omega_\varepsilon = \Omega \setminus \tilde{\Omega} \)) such that on \( \Omega_\varepsilon \) we have a worst case (deterministic)
upper bound on the noise level and $\mathbb{P}(\Omega_\epsilon) \to 1$ as $\eta \to 0$. For general stochastic noise $\epsilon$ in (1) it follows from Markov’s inequality (9) that for any $\tau > 0$

$$\mathbb{P}(|\epsilon|_Y \geq \tau \mathbb{E}(|\epsilon|_Y)) \leq \frac{1}{\tau}.$$

In [10] it was shown that even when $\epsilon$ is a finite dimensional Gaussian random variable and $\mathbb{P}(|\epsilon|_Y \geq \tau \mathbb{E}(|\epsilon|_Y))$ can be calculated analytically this probability is, just as in (16), independent of the variance $\eta$. Hence both in general and for the specific case of Gaussian noise one has $\mathbb{P}(\Omega_\epsilon) = c$ with a constant $c = c(\tau)$ between zero and one, independent of $\eta$. Our approach to circumvent this effect is to link $\tau$ with the variance and increase $\tau$ with decreasing $\eta$. Let $\tau = \tau(\eta)$ and $\tau \to \infty$ as $\eta \to 0$. We then have

$$\mathbb{P}(|\epsilon|_Y \geq \tau(\eta)\mathbb{E}(|\epsilon|_Y)) \leq \frac{1}{\tau(\eta)} \to 0,$$

i.e., $\mathbb{P}(\Omega_\epsilon) \to 1$ for $\eta \to 0$. For Gaussian noise we see, comparing (11) and (12) that the artificial inflation we impose on the expectation via $\eta$-dependent $\tau$ in the following theorems is automatically included in the Ky Fan distance. We suppose that this is the reason why the convergence theory carries over in such a direct fashion for the Ky Fan metric. We now give the theorem for the lifting strategy using the expectation to quantify the noise level.

**Theorem 4.4.** Let $R_\alpha$ be a regularization method for the solution of (2) in the deterministic setting under a suitable choice of the regularization parameter. Let now $y^\eta = y + \epsilon(\eta)$ where $\epsilon(\eta)$ is a stochastic error such that $\mathbb{E}(d_\tau(y, y^\eta)) \to 0$ as $\eta \to 0$. Then, assuming (2) has a unique solution $x^1$ and all required assumptions for the deterministic theory (except the bound on the noise) hold in the sense of Remark 7 the regularization method $R_\alpha$ fulfills

$$\lim_{\eta \to 0} \rho_K(R_\alpha(y^\eta), x^1) = 0$$

under the same parameter choice rule as in the deterministic setting with $\delta$ replaced by $\tau(\eta)\mathbb{E}(d_\tau(y, y^\eta))$ where $\tau(\eta)$ fulfills

$$\tau(\eta) \to \infty \quad \text{and} \quad \lim_{\eta \to 0} \tau(\eta)\mathbb{E}(d_\tau(y, y^\eta)) = 0.$$  

**Proof.** As previously we pick a “worst case” subsequence $\{y^\eta_{\omega}\}$ of $\{y^\eta\}$, a subsequence for which the corresponding solutions satisfy $\rho_K(x^1, x_{\alpha(\eta_{\omega})}) \geq \theta/2$.

Let $\epsilon > 0$. We can now pick a subsequence which we again denote by $\{y^\eta_{k_l}\}$ fulfilling

$$\frac{2}{\tau(\eta_{k_l})} \leq \epsilon,$$

where without loss of generality $\tau(\eta_{k_l}) > 1$, such that

$$\mathbb{P}(\omega : d_\tau(y(\omega), y^\eta_{k_l}(\omega)) > \tau(\eta_{k_l})\mathbb{E}(d_\tau(y, y^\eta_{k_l}))) \leq \frac{1}{\tau(\eta_{k_l})} \leq \frac{\epsilon}{2}.$$

This again defines, via the complement in $\Omega_1$, $\Omega$ with $\mathbb{P}(\Omega) \geq 1 - \frac{\epsilon}{2}$ on which $d_\tau(y(\omega), y^\eta_{k_l}(\omega)) \leq \tau(\eta_{k_l})\mathbb{E}(d_\tau(y, y^\eta_{k_l})).$ As before, we can now apply the deterministic theory by substituting $\delta$ with $\tau(\eta_{k_l})\mathbb{E}(d_\tau(y, y^\eta_{k_l})).$ The remainder of the proof is identical to the one of Theorem 4.3.

The theorems justify the use of deterministic algorithms under a stochastic noise model. Since the proof is solely based on relating the stochastic noise to a deterministic one on subsets of $\Omega$ and does not use any specific properties of the
regularization methods or the underlying spaces, it opens most of the deterministic methods for the stochastic noise model. In particular, the parameter choice rules from the deterministic setting are easily adapted once \( \rho \) methods for the stochastic noise model. In particular, the parameter choice rules or the underlying spaces, it opens most of the deterministic solutions. For example if \( X \) is a Banach space of functions \( x : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}, n \in \mathbb{N}, \) with norm \( || \cdot ||_X \) then it is enough to enforce \( ||R_\alpha(y^n)||_X \leq K \) and \( |x(t)| \leq M \) for all \( t \in D \) where \( K \) and \( M \) are positive constants. Often one can assume an a-priori bound \( ||x^\dagger||_X \leq \tilde{K} \) and a bound on the function values \( |x^\dagger(t)| \leq \tilde{M} \) from the application under consideration. If one discards any regularized solution \( R_\alpha(y^n) \) for which \( ||R_\alpha(y^n)||_X > CK \) and/or \( |R_\alpha(y^n)[t]| > CM \) with some \( 1 < C < \infty, \) then the whole sequence of regularized solutions is uniformly bounded with constants \( K = CK \) and \( M = CM. \) Without going into further detail we would like to note that in practical applications and in particular finite dimensional problems such box constraints are usually not a restriction when \( C \) is chosen large enough, thus yielding convergence of the regularized solutions in expectation.

The general convergence theorem is followed by convergence rates which are obtained under additional assumptions. Often these conditions ensure at least local uniqueness of the true solution. If not, we have to require such a property for the same reason as previously.

**Theorem 4.5.** Let \( R_\alpha \) be a regularization method for the solution of [2] in the deterministic setting such that, under a set of assumptions on the operator \( F \) and the solutions \( x^\dagger \) and a suitable choice of the regularization parameter,

\[
d_X(R_\alpha(y^\delta), x^\dagger) \leq \varphi(d_Y(y^\delta))
\]

with a monotonically increasing right-continuous function \( \varphi. \)

Let now \( y^n = y + \epsilon(\eta) \) where \( \epsilon(\eta) \) is a stochastic error such that

a) \( \rho_K(y, y^n) \rightarrow 0 \) or

b) \( \mathbb{E}(d_Y(y, y^n)) \rightarrow 0 \)

as \( \eta \rightarrow 0. \) Then, assuming all required assumptions for the deterministic theory (except the bound on the noise) hold in the sense of Remark 2 and that there is (either by the deterministic conditions or by additional assumption) a (locally) unique solution \( x^\dagger \) to [2], the regularization method \( R_\alpha \) fulfills

\[
\rho_K(R_\alpha(y^\delta), x^\dagger) = O(\max\{\varphi(\rho_K(y, y^n)), \rho_K(y, y^n)\})
\]
in case a) or, respectively, in case b),
\[ \rho_K(R_\alpha(y^\delta), x^1) = \mathcal{O}(\max(\varphi(\tau(\eta))\mathbb{E}(d_Y(y, y^n))), \mathbb{P}(d_Y(y, y^n) \geq \tau(\eta)\mathbb{E}(d_Y(y, y^n)))) \}
under the same parameter choice rule as in the deterministic setting with \( \delta \) replaced by \( \rho_K(y, y^n) \) (case a) or \( \tau(\eta)\mathbb{E}(d_Y(y, y^n)) \) where \( \tau(\eta) \) fulfills \( \tau(\eta) \leq \frac{1}{\tau(\eta)} \).

Proof. We start again with the Ky Fan distance as noise measure. Since we have
\[ d_Y(y, y^n) \leq \rho_K(y, y^n) \]
by Markovs inequality. Hence, with probability \( 1 - \frac{1}{\tau(\eta)} \) we are in the deterministic setting with \( \delta = \tau(\eta)\mathbb{E}(d_Y(y, y^n)) \) and
\[ \mathbb{P}(d_Y(y, y^n) \geq \tau(\eta)\mathbb{E}(d_Y(y, y^n))) \leq \frac{1}{\tau(\eta)}. \]
The convergence rate follows by the definition of the Ky Fan metric.

For Inverse Problems, the convergence rates are most often given by functions which decay at most linearly, i.e.,
\[ \max\{\varphi(\rho_K(y, y^n)), \rho_K(y, y^n)\} = \varphi(\rho_K(y, y^n)). \]
Hence in this case the convergence rates are preserved in the Ky Fan metric. For the expectation this is not the case as we need to introduce the inflation parameter \( \tau \), see the discussion before Theorem 4.3.

For many nonlinear Inverse Problems the requirement of a unique solution is too strong. Often one has several solutions of the same quality, in particular there exists more than one minimum norm solution. In this case, Theorem 4.3 is not applicable. In the example \( \{22, 23\} \) Example 4.3 and 4.5] with two minimum norm solutions the noise was constructed such that, while the error in the data converges to zero, for each fixed \( \omega \in \Omega \) the regularized solutions jump between both solutions such that no converging subsequence can be found. The main problem there is that the Ky Fan distance cannot incorporate the concept that all minimum norm solutions are equally acceptable. We will now define a pseudo metric that resolves this issue.

**Definition 4.6.** Let \((\mathcal{X}, d_\mathcal{X})\) be a metric space. Denote by \( \mathcal{L} \) the set of minimum-norm solutions to \( \mathcal{L} \). Then
\[ \rho_K^\mathcal{L}(x) := \inf_{\varepsilon > 0} \left\{ \mathbb{P} \left( \inf_{x^1 \in \mathcal{L}} d_\mathcal{X}(x, x^1) > \varepsilon \right) \leq \varepsilon \right\} \]
measures the distance between an element \( x \in \mathcal{X} \) and the set \( \mathcal{L} \), in particular it is
\[ \rho_K^\mathcal{L}(x) = 0 \iff x \in \mathcal{L} \] almost surely.
With this, one can define a pseudometric on \((\Omega, \mathcal{F}, \mathbb{P})\) via
\[
\rho_{KL}^L(x_1, x_2) := \max\{\rho_{KL}^L(x_1), \rho_{KL}^L(x_2)\}.
\]
Obviously \((18)\) is positive, symmetric and fulfills the triangle inequality. However, 
\(\rho_{KL}^L(x_1, x_2) = 0\) does not imply \(x_1 = x_2\) a.e. but instead \(x_1 \land x_2 \in \mathcal{L}\) which fixes the 
aforementioned issue of the Ky Fan metric and allows the following theorems.

**Theorem 4.7.** Let \(R_\alpha\) be a regularization method for the solution of \((2)\) in the 
deterministic setting under a suitable choice of the regularization parameter. Let 
now \(y_\eta = y + \epsilon(\eta)\) where \(\epsilon(\eta)\) is a stochastic error such that 
\(a)\) \(\rho_K(y, y_\eta) \to 0\) or 
\(b)\) \(E(d_Y(y, y_\eta)) \to 0\) as \(\eta \to 0\). Then, assuming all necessary assumptions for the deterministic theory 
(except the bound on the noise) hold with probability one, the regularization method 
\(R_\alpha\) fulfills
\[
\lim_{\eta \to 0} \rho_{KL}^L(R_\alpha(y_\eta)) = 0
\]
under the same parameter choice rule as in the deterministic setting with \(\delta\) replaced 
by \(\rho_K(y, y_\eta)\) (case \(a)\)) or \(\tau(\eta)E(d_Y(y, y_\eta))\) where \(\tau(\eta)\) fulfills \((17)\) (case \(b)\)). In 
particular, the sequence of regularized solutions fulfills
\[
\lim_{\eta_1, \eta_2 \to 0} \rho_{KL}^L(R_\alpha(y_\eta_1), R_\alpha(y_\eta_2)) = 0
\]

**Proof.** The proof follows the lines of the one of Theorem 4.3 with \(\rho_K(\cdot, x^\dagger)\) replaced 
by \(\rho_{KL}(\cdot)\). Also Lemma 4.1 is easily adjusted to incorporate multiple solutions. □

### 4.2. Fully stochastic Inverse Problems.

So far we assumed that only the noise is stochastic whereas the operator \(F\) and the unknown \(x\) were assumed to be determin- 
istic. We will now also allow these objects to be stochastic. Hence \((2)\) reads
\[
F(x(\omega), \omega) = y(\omega).
\]
The task in this section is again to find an approximation \(x_\alpha^\delta\) to an exact solution 
\(x^\dagger\) satisfying \((19)\) where as before only noisy data \((1)\) is available. Now the un- 
known and or the operator might contain randomness themselves. In particular 
for each \(\omega \in \Omega\) we might have a different exact solution \(x^\dagger(\omega)\). Hence the regular- 
ized solutions now depend on the realization \(\omega\) in a several possible ways and 
each regularized solution is obtained for a specific \(\omega\). It was shown in [22, 23] how 
deterministic conditions such as source conditions can be incorporated into the sto- 
chastic setting by assuming that the deterministic conditions hold with a certain 
probability. However, additional conditions may occur when lifting these in order 
to ensure the deterministic requirements again up to a certain probability. Due to 
the possible multiplicity of stochastic conditions which might appear in this con- 
text it seems impossible to develop a lifting strategy in such a general fashion as 
in the previous section. We will therefore consider two classical examples, namely 
nonlinear Tikhonov regularization and Landweber’s method for nonlinear Inverse 
Problems. The theory is taken completely from [22, 23]. In this section we are in 
a Hilbert space setting. In particular the metrics \(d_X\) and \(d_Y\) are now norms in the 
Hilbert spaces \(X\) and \(Y\), respectively.
4.2.1. Nonlinear Tikhonov Regularization. We seek the solution of a nonlinear ill-posed problem \( \text{(19)} \) via the variational problem

\[
x^*_n = \arg\min_{x} ||F(x) - y^n||^2 + \alpha ||x - x^*||^2
\]

with a reference point \( x^* \in X \) and given noisy data \( y^n \) according to \( \text{(1)} \), where the stochastic distribution of the noise is assumed to be known. We shall skip the general convergence theorem (which follows as in the previous section) and move to convergence rates directly. In the deterministic theory, i.e., when \( y^\delta \) is the noisy data with \( ||y - y^\delta|| \leq \delta \), we have the following theorem from [12].

**Theorem 4.8.** Let \( D(F) \) be convex, \( y^\delta \in Y \) such that \( ||y - y^\delta|| \leq \delta \) and \( x^\dagger \) denote the \( x^\ast \)-minimum norm solution of \( \text{(19)} \). Furthermore let the following conditions hold.

a) \( F \) is Fréchet-differentiable

b) There exists \( \gamma \geq 0 \) such that \( ||F'(x^\dagger) - F'(x)|| \leq \gamma ||x^\dagger - x|| \) in a sufficiently large ball \( B_\theta(x^\dagger) \cap D(F) \)

c) \( x^\dagger - x^* \) satisfies the source condition \( x^\dagger - x^* = F'(x^\dagger) v \) for some \( v \in Y \).

d) The source element satisfies \( \gamma ||v|| < 1 \).

Then for the choice \( \alpha = c\delta \) with some fixed \( c > 0 \) we obtain

\[
||x^\dagger - x_n^\dagger|| \leq \frac{\delta + \alpha ||v||}{\sqrt{\alpha \sqrt{1 - \gamma ||v||}}} = \mathcal{O}(\sqrt{\delta}) \text{ and } ||F(x_n^\dagger) - y^\delta|| = \mathcal{O}(\delta).
\]

As given in Theorem 4.6 of [22], the following stochastic formulation of Theorem 4.8 holds.

**Theorem 4.9.** Let \( D(F) \) be convex, let \( y^n \) be such that \( 0 \leq \rho_K(y, y^n) < \infty \) and \( x^\dagger \) denote the \( x^\ast \)-minimum norm solution of \( \text{(19)} \) for almost all \( \omega \). Furthermore let the following conditions hold.

a) \( F(\cdot, \omega) \) is Fréchet-differentiable for almost all \( \omega \)

b) \( F'(\cdot, \omega) \) satisfies, with some \( \gamma : \Omega \rightarrow \mathbb{R}_+ \),

\[
||F'(x^\dagger(\omega), \omega) - F'(x, \omega)|| \leq \gamma(\omega)||x^\dagger(\omega) - x||
\]

in the ball \( B_\theta(\omega)(x^\dagger(\omega)) \cap D(F) \), where \( \theta(\omega) \geq 2||x^\dagger(\omega) - x^*(\omega)|| + \varepsilon \) for some \( \varepsilon > 0 \) independent of \( \omega \).

c) (smoothness) \( \mathbb{P}(\Omega_{ac}) = 1 \) where

\[
\Omega_{ac} := \{ \omega : \exists v(\omega), x^\dagger(\omega) - x^*(\omega) = F'(x^\dagger(\omega), \omega)^* v(\omega) \}.
\]

d) (closeness) There exists \( \varphi_{cl} : [0, 1] \rightarrow [0, 1], \lim_{\xi \rightarrow 1^-} \varphi_{cl}(\xi) = 0 \) such that

\[
\mathbb{P}(\omega \in \Omega_{ac} : \gamma(\omega)||v(\omega)|| > \xi) < \varphi_{cl}(\xi).
\]

e) (decay) There exists \( \varphi_{de} : [0, \infty) \rightarrow [0, 1], \lim_{\tau \rightarrow \infty} \varphi_{de}(\tau) = 0 \) such that

\[
\mathbb{P}(\omega \in \Omega_{ac} : ||v(\omega)|| > \tau) < \varphi_{de}(\tau) \quad \forall \tau \geq 0.
\]

Then for the choice \( \alpha = c\rho_K(y, y^n) \), \( c > 0 \), we obtain

(20) \[
\rho_K(x^\dagger, x_n^\dagger) \leq \inf_{\xi \in [0, 1]} \max_{\xi \in (0, 1)} \left\{ \rho_K(y, y^n) + \varphi_{cl}(\xi) + \varphi_{de}(\tau), \sqrt{\rho_K(y, y^n)} \frac{1 + ct}{\sqrt{1 - \xi}} \right\}.
\]

**Proof.** We have \( ||y - y^n|| \leq \rho_K(y, y^n) \) with probability \( 1 - \rho_K(y, y^n) \). Now fix \( \xi < 1 \) and \( 0 < \tau < \infty \). Then we have due to d) and e)

(21) \[
\mathbb{P}(\omega \in \Omega : \gamma(\omega)||v(\omega)|| < \xi \land ||v(\omega)|| < \tau) \leq 1 - (\varphi_{cl}(\xi) + \varphi_{de}(\tau)).
\]
For the $\omega$ in (21) we can apply Theorem 4.8 and obtain
\[
\left\| x^\dagger(\omega) - x^n_\alpha(\omega) \right\| \leq \frac{\rho_K(y,y^n) + \alpha \tau}{\sqrt{\alpha \sqrt{1 - \xi}}}
\]
or, fixing the parameter $\alpha = c \rho_K(y,y^n), c > 0$,
\[
\left\| x^\dagger(\omega) - x^n_\alpha(\omega) \right\| \leq \sqrt{\rho_K(y,y^n)} \frac{1 + c \tau}{\sqrt{1 - \xi}}
\]
The above estimate holds on some set that has a probability greater or equal than $\epsilon$
\[
\subset \Omega \gamma
\]
conditions such as the decay condition in Theorem 4.9. Namely, since $\epsilon$
on $\Omega$
Inverse Problems and Imaging Volume 11, No. 4 (2017), 663–687
This estimate is valid for arbitrary choices of $\xi$ and $\tau$ above, therefore we may bound
the Ky fan distance of $x^\dagger$ and $x^n_\alpha$ by taking the infimum with respect to $\xi$ and $\tau$.

As before, the core principle of the lifting strategy is to ensure that there exists
a subset $\Omega_\epsilon \subset \Omega$ such that all deterministic assumptions hold with probability
one on $\Omega_\epsilon$ and $\mathbb{P}(\Omega_\epsilon) \to 1$ as $\eta \to 0$. This may lead to the introduction of new
conditions such as the decay condition in Theorem 4.9. Namely, since $\gamma(\omega)$ and
$\|v(\omega)\|$ may vary with $\omega$, it may be possible for a sequence $\{\omega_k\}_{k \in \mathbb{N}}$ that $\gamma(\omega_k) \to 0$
and $\|v(\omega_k)\| \to \infty$ such that still for all $k \in \mathbb{N} \gamma(\omega_k)\|v(\omega_k)\| < 1$. In this case
the parameter $\tau$ cannot be treated as a constant in the convergence rate, but it
influences it to a significant degree. The decay condition had to be imposed in
order to control the growth of $\tau$. It is, however, possible to avoid condition e) by
imposing other conditions. For example, one could require that $\gamma(\omega)$ is bounded
below by some $0 < c < 1$. In this case condition d) implies e). A more detailed
discussion is given in [22].

Accordingly, in order to lift other deterministic convergence rate results into
the fully stochastic setting, a careful examination of the conditions necessary for
convergence in the stochastic setting, understanding their cross-connections and
dependencies is important. However, once the conditions have been translated to the
stochastic setting, convergence rates follow immediately using the Ky Fan metric.
We will now examine how particular choices of the stochastic parameters in Theorem 4.9
influence the convergence rate. To this end, we cite Remark 4.8 of [22].

Let in the first examples the operator be deterministic, i.e., $F(\cdot,\omega) = F(\cdot)$ where
$\gamma(\omega) = \gamma = 1$. First consider the case that $\|v\| \in U[0,1]$, i.e., it is uniformly
distributed on the interval $[0,1]$. We therefore have $\varphi_{cl}(\xi) = 1 - \xi$, as well as
$\varphi_{de} = 0$ for $\tau > 1$. Thus Theorem 4.9 implies
\[
\rho_K(x^\dagger, x^n_\alpha) \leq \inf_{0 < \alpha < \infty} \inf_{\xi(0,1)} \max \left\{ \rho_K(y,y^n) + 1 - \xi, \sqrt{\rho_K(y,y^n)} \frac{\rho_K(y,y^n) + \alpha}{\alpha \sqrt{1 - \xi}} \right\}
\]
which gives for $\alpha = c \rho_K(y,y^n), c > 0$, the optimal rate
\[
\rho_K(x^\dagger, x^n_\alpha) = O(\rho_K(y,y^n)^{1/3}).
\]
For the second case suppose that $\varphi_{de}(\tau) = c \tau^{-e}$ for some exponent $e > 0$. Since
now we do not have $\varphi_{cl}(\xi) \to 0$, but $\varphi_{cl} \geq c > 0$ we obtain
\[
\rho_K(x^\dagger, x^n_\alpha) \leq \inf_{0 < \alpha < \infty} \inf_{\xi(0,1)} \max \left\{ \frac{1}{\sqrt{\alpha \sqrt{1 - \xi}}}, \sqrt{\rho_K(y,y^n)} \frac{\rho_K(y,y^n) + \alpha}{\alpha \sqrt{1 - \xi}} \right\}.
\]
Since the right hand side does not converge to zero we do not obtain a convergence rate anymore. However, convergence itself still follows from Theorem 4.3.

Finally, consider the case when both d) and e) from Theorem 4.9 influence the convergence behavior, because $F$ is stochastic with varying $\gamma(\omega)$. For instance in the case that for some $\omega \in U[0,1]$ we have $x^\dagger(\omega) = \omega x^\dagger$ and $\gamma(\omega) = 1 - \omega$, we find that $\varphi_{cl}(\xi) = 1 - \xi$ and $\varphi_{de}(\tau) = c/(1 + \tau)$ are compatible realizations of $\varphi_{cl}(\cdot)$ and $\varphi_{de}(\cdot)$. With this one can show

$$\rho_K(x^\dagger, x^\alpha_0) = O(\rho_K(y, y^\eta)^{1/4})$$

under the parameter choice $\alpha = c\rho_K(y, y^\eta)^{5/4}$, $c > 0$. From the given examples it is evident that the convergence speed is heavily influenced by the conditions d) and e) in Theorem 4.9. Therefore, although the general formula for the convergence rate (20) may suggest that the convergence rate is close to the deterministic one, it may be significantly slower due to the additional stochastic properties.

4.2.2. Nonlinear Landweber iteration. We shall discuss another classical regularization method to demonstrate the lifting approach. As before we seek the solution of a nonlinear ill-posed problem (19) given noisy data $y^\eta$ according to (1) where the stochastic distribution of the noise is assumed to be known. Landweber’s method can be seen as a descent algorithm for $||F(x) - y^\delta||^2$ and is defined via the iteration

$$x^\delta_{k+1} = x^\delta_k - \beta F'(x^\delta_k)^*(F(x^\delta_k) - y^\delta), \quad k = 1, 2, \ldots,$$

where $\beta > 0$ is an appropriately chosen stepsize and $x^\delta_0$ an initial guess. Landweber’s method constitutes a regularization method if it is stopped early enough [20]. In the deterministic theory, i.e., when $y^\delta$ is the noisy data with $||y - y^\delta|| \leq \delta$, we have the following theorem from [20] for convergence rates of the Landweber method.

**Theorem 4.10.** Let $\mathcal{D}(F)$ be convex, $y^\delta \in \mathcal{Y}$ such that $||y - y^\delta|| \leq \delta$ and $x^\dagger$ denote the $x^*$-minimum norm solution of (19). Assume (19) has a solution in $\mathcal{B}_\delta(x^*)$. Furthermore let the following conditions hold on $\mathcal{B}_{2\delta}(x^*)$.

a) There are bounded linear operators $R_x$ satisfying $||R_x - I|| \leq C||x - x^\dagger||$ for $0 < C < \infty$ such that

$$F'(x) = R_x F'(x^\dagger)$$

b) $x^\dagger - x^*$ satisfies the source condition $x^\dagger - x^* = (F'(x^\dagger)^* F'(x^\dagger))^\nu v$ for some $v \in \mathcal{Y}$ and $0 < \nu \leq 1/2$.

Let $||v||$ be sufficiently small. Then, if the regularization parameter is stopped according to the discrepancy principle, i.e., at the unique index $k_*$ for which for the first time

$$||F(x_k) - y^\delta|| \leq \tilde{\tau}\delta$$

with $\tilde{\tau} > 2^{1+\nu}/(1-2\nu) > 2$, we obtain

$$||x^\dagger - x^\delta_{k_*}|| \leq c||v||^{1/(2\nu+1)} \delta^{2\nu/(2\nu+1)}.$$  

We can obtain a stochastic version of Theorem 4.10 in the same way and with the same techniques used to show that Theorem 4.9 followed from Theorem 4.8.

**Theorem 4.11.** Let $\mathcal{D}(F)$ be convex, $y^\eta \in \mathcal{Y}$ be given with $\rho_K(y, y^\eta)$ and let $x^\dagger(\omega)$ denote the $x^*$-minimum norm solution of (2). Assume (2) has a solution in $\mathcal{B}_\delta(x^*(\omega))$ for almost all $\omega$. Furthermore let the following conditions hold on $\mathcal{B}_{2\delta}(x^*)$. 

**Inverse Problems and Imaging** Volume 11, No. 4 (2017), 663–687
a) \( F'(x, \omega) = R_{x, \omega} F'(x^\dagger(\omega), \omega) \) where for almost all \( \omega \) \( \{ R_{x, \omega} : x \in B_{\rho}(x^*) \} \) describes a family of bounded linear operators \( R_{x, \omega} : \mathcal{Y} \times \Omega \to \mathcal{Y} \) with
\[
||R_{x, \omega} - I|| \leq C(\omega)||x - x^\dagger(\omega)||
\]
b) \( x^\dagger - x^* \) satisfies the source condition
\[
x^\dagger(\omega) - x^*(\omega) = (F'(x^\dagger(\omega), \omega)^* F'(x^\dagger(\omega), \omega))^{\nu}(\omega)
\]
for some \( v(\omega) \in \mathcal{Y} \) and \( 0 < \nu \leq \frac{1}{2} \).
c) There exists \( \varphi_{cl} : [0, \infty) \to [0, 1], \lim_{\xi \to 1^-} \varphi_{cl}(\xi) = 0 \) such that
\[
\mathbb{P}(\omega \in \Omega : C(\omega)||v(\omega)|| > \xi) < \varphi_{cl}(\xi)
\]
d) There exists \( \varphi_{de} : [0, \infty) \to [0, 1], \lim_{\tau \to \infty} \varphi_{de}(\tau) = 0 \) such that
\[
\mathbb{P}(\omega \in \Omega_{de} : ||v(\omega)|| > \tau) < \varphi_{de}(\tau) \quad \forall \tau \geq 0.
\]
Then, if the regularization parameter is stopped according to the discrepancy principle, i.e., at the unique index \( k_* \) for which for the first time
\[
||F(x_k) - y^\eta|| \leq \hat{\tau}_{\rho_K}(y, y^\eta)
\]
with \( \hat{\tau} > 2 \), we obtain for \( c_0 > 0 \) sufficiently small the rate
\[
\rho_{K}(x^\dagger_0, x^\eta_{k_*}) \leq 
\inf_{0 < \tau \leq \infty} \max \left\{ \rho_K(y, y^\eta_{\tau}) + \varphi_{cl}(c_0) + \varphi_{de}(\tau), c \tau^{1/(2\nu+1)} \rho_K(y, y^\eta_{2\nu/(2\nu+1)}) \right\}
\]
where the constant \( c \) depends on \( \nu \) only.

To further demonstrate the capabilities of the lifting technique we allow randomness in another condition. Namely, in the fully stochastic setting, the source condition b) from Theorem 4.11 need not hold with constant exponent \( \nu \) for all \( \omega \in \Omega \). There are at least two situations which lead to the power \( \nu \) being a stochastic quantity as well, i.e., it holds
\[
(22) \quad x^\dagger(\omega) = (F'(x^\dagger(\omega), \omega)^* F'(x^\dagger(\omega), \omega))^{\nu}(\omega)v(\omega)
\]
with \( 0 < \nu(\omega) \leq \frac{1}{2} \). In the first case all solutions \( x^\dagger(\omega) \) come from some initial element \( v(\omega) = v \in \mathcal{Y}, \) with small \( \mathcal{Y} \)-norm. Some randomly smoothing operator is acting on this element and generates \( x^\dagger(\omega) \). (One could for instance think of some kind of evolution process, e.g., a diffusion process that is applied to some initial value \( v \).) The smoothness of \( x^\dagger(\omega) \) is therefore random. Secondly, \( x^\dagger \) may be a deterministic element satisfying a certain smoothness condition. The data \( y(\omega) \) is generated by applying a forward operator \( F(\cdot, \omega) \) with random smoothing properties. If the realization of \( F(\cdot, \omega) \) is strongly smoothing, this corresponds to a source condition with small \( \nu(\omega) \), if \( F(\cdot, \omega) \) is weakly smoothing we have the source condition with larger \( \nu(\omega) \).

The following proposition shows the convergence rate that results from the source condition \( (22) \) for the case that \( \nu(\omega) \) is uniformly distributed on the interval \([0, \frac{1}{2}]\).

**Proposition 2.** Let all conditions of Theorem 4.11 hold except for b) and d). Let \( x^\dagger(\omega) \) satisfy \( (22) \) where \( ||v(\omega)|| \) is uniformly bounded and sufficiently small. Let \( \nu(\omega) \) be uniformly distributed on the interval \([0, \frac{1}{2}]\), i.e.,
\[
\mathbb{P}(\omega \in \Omega : 0 \leq \nu(\omega) < \nu \leq \frac{1}{2}) = 2\nu.
\]
Then the approximations $x_{k}^{n}$ obtained by Landweber’s method satisfy the convergence rate

$$\rho_{K}(x^{\dagger}, x_{k}^{n}) = O \left( \frac{W(-\log(\rho_{K}(y, y^{\delta})))}{-\log(\rho_{K}(y, y^{\delta}))} \right)$$

where $W$ denotes the Lambert W-function, defined by $W(z)e^{W(z)} = z$, see [8].

Proof. As can be seen from the proof of Theorem 3.1 in [20], the requirement "$||v||$ sufficiently small", becomes stronger, the larger $\nu$ is. Supposing that $||v||$ in (22) is sufficiently small for the case $\nu = \frac{1}{2}$, implies therefore that also the convergence conditions for $\nu \leq \frac{1}{2}$ are satisfied.

Secondly we observe that the convergence rate in Theorem 4.11 contains a constant $\tilde{c}$ that depends on $\nu$. Although it is difficult to state an explicit formula for $\tilde{c}$, investigation of [20] shows that $\tilde{c}(\nu)$ attains its maximum value when $\nu = \frac{1}{2}$.

After these observations we start with the actual derivation of the convergence rate. For the sake of simplicity we assume that all appearing constants are just equal to 1. Furthermore we may assume that $\varphi_{cl}(\cdot)$ and $\varphi_{de}(\cdot)$ both vanish. Asymptotically, for given $\omega$ we therefore have the estimate

$$||x^{\dagger}(\omega) - x_{k}^{n}(\omega)|| \leq \rho_{K}(y, y^{\delta})^{2\nu(\omega)}.$$ 

Measuring the distance in the Ky Fan metric we must, since we assumed that $\nu(\omega)$ is as in (22), solve the equation

$$\rho_{K}(y, y^{\delta})^{2\tilde{\nu}} = 2\nu \tag{24}$$

for $\nu$. We first consider the simplified equation

$$\rho_{K}(y, y^{\delta})^{2\tilde{\nu}} = 2\tilde{\nu}$$

which is solved by

$$\tilde{\nu}(\rho_{K}(y, y^{\delta})) = \frac{W(-\log(\rho_{K}(y, y^{\delta})))}{-2\log(\rho_{K}(y, y^{\delta}))}.$$ 

In the following we show that the above approximate solution is sufficiently accurate. Therefore we can replace the term $2\tilde{\nu} + 3\tilde{\nu} = \tilde{\nu}(\rho_{K}(y, y^{\delta}))(1 + \varepsilon(\tilde{\nu}(\rho_{K}(y, y^{\delta}))))$. The original equation from (24) then contains the term $2\tilde{\nu} + 3\varepsilon + 1$. Neglecting the quadratic part, we can replace this term with $2\tilde{\nu} + 1$, and obtain an equation that MATHEMATICA can solve for $\varepsilon(\tilde{\nu})$. The solution for the correction term is given as

$$\varepsilon(\tilde{\nu}) = \log(\tilde{\nu}) + (2\tilde{\nu} + 1)W \left( \frac{\log(\tilde{\nu})}{2\tilde{\nu} + 1} \right) - \log(\tilde{\nu})$$

and tends to zero approximately linearly in $\tilde{\nu}$. Thus this correction becomes small rather quickly, and we can consider the asymptotic bound in (23) as sufficiently accurate due to the asymptotics of the Lambert W-function.

5. Examples. We will now apply the theory developed in the previous sections to selected deterministic regularization methods.
5.1. Filter-based regularization methods. Let \( A \) be a linear compact operator between Hilbert spaces \( \mathcal{X} \) and \( \mathcal{Y} \) with singular system \( \{ \sigma_n, u_n, v_n \}_{n \in \mathbb{N}} \), see, e.g. [12]. Then, for \( y \in \mathcal{D}(A) \), the generalized inverse \( A^\dagger \) to \( A \) is given by

\[
A^\dagger y = \sum_{\sigma_n > 0} \sigma_n^{-1} (y, u_n) v_n.
\]

Since for compact operators the singular values approach zero, their inverse blows up and the generalized inverse yields a meaningless solution to (2) for noisy data. A popular class of regularization methods is based on the filtering of the generalized inverse. Introducing an appropriate filter function \( F_\alpha(\sigma) \) depending on the regularization parameter \( \alpha \) that controls the growth of \( \sigma^{-1} \), the regularized solutions are defined by

\[
R_\alpha(y) = \sum_{\sigma_n > 0} F_\alpha(\sigma_n^{-1} (y, u_n)) v_n.
\]

Examples for filter based methods are for example the classical Tikhonov regularization, truncated singular value decomposition or Landwebers method [12, 32]. The regularization properties are fully determined by the filter functions. In the deterministic setting, the conditions can be found in, e.g., [32, Theorem 3.3.3.].

Convergence rates can be obtained for a priori and a posteriori parameter choice rules under stricter conditions on the filter functions. We will only comment on an a priori choice here in order to keep this section short. An example of the discrepancy principle as a posteriori parameter choice is given in the next section in a different context. Using the smoothness condition

\[
x^\dagger \in \text{range}(A^* A)^{\nu/2}, \quad \| x^\dagger \|_\nu := \{ \| z \|_\chi : x^\dagger = (A^* A)^{\nu/2} z, z \in \mathcal{N}(A)^{-1} \} \leq \rho
\]

the following theorem can be obtained.

**Theorem 5.1.** [32, Theorem 3.4.3] Let \( y \in \text{range}(A) \) and \( \| y - y^\delta \|_\gamma \leq \delta \). Assume that it holds \( \| x^\dagger \|_\nu \leq \rho \) and for \( 0 \leq \nu \leq \nu^* \),

\[
\sup_{0 < \sigma \leq \sigma_1} \sigma^{-1} |F_\alpha(\sigma)| \leq c \alpha^{-\beta}
\]

(25)

\[
\sup_{0 < \sigma \leq \sigma_1} |1 - F_\alpha(\sigma)| \sigma^{\nu^*} \leq c_{\nu^*} \alpha^{\beta \nu^*},
\]

(26)

where \( \beta > 0 \) and \( c, c_{\nu^*} \) are constants independent of \( \delta \). Then with the a priori parameter choice

\[
\alpha = C \left( \frac{\delta}{\rho} \right)^{1/(\nu+1)}, \quad C > 0 \quad \text{fixed},
\]

the method induced by the filter \( F_\alpha \) is order optimal for all \( 0 \leq \nu \leq \nu^* \), i.e.,

\[
\| x^\dagger - R_\alpha y^\delta \| \leq c \delta^{1/\gamma} \rho^{1/\gamma}
\]

for some constant \( c \) independent of \( \delta \) and \( \rho \).

Now we use Theorem 4.5 and obtain convergence rates in the Ky Fan metric.

**Theorem 5.2.** Let \( y \in \text{range}(A) \) and \( \rho_K(y, y^\eta) \) be known. Assume that it holds \( \| x^\dagger \|_\nu \leq \rho \) and for \( 0 \leq \nu \leq \nu^* \), (25) and (26) hold. Then with the a priori parameter choice

\[
\alpha = C \left( \frac{\rho_K(y, y^\eta)}{\rho} \right)^{1/(\nu+1)}, \quad C > 0 \quad \text{fixed},
\]
the method induced by the filter \( F_\alpha \) fulfills
\[
\rho_K(x^\dagger, R_\alpha y^\eta) \leq c \rho_K(y, y^\eta)^{\frac{1}{\nu+1}}
\]
for some constant \( c \) independent of \( \delta \) and \( \eta \).

More about filter methods in the stochastic setting including numerical examples can be found in [16].

5.2. Sparsity-regularization for an autoconvolution problem. We consider an autoconvolution equation
\[
(F(x))(s) = \int_0^s x(s-t)x(t) \, dt, \quad 0 \leq s \leq 1
\]
between Hilbert spaces \( X = L_2[0,1] \) and \( Y = L_2[0,1] \) where \( x \in \mathcal{D}(F) \). Such an equation is of great interest in, for example, stochastics or spectroscopy and has been analyzed in detail in [19]. Recently, a more complicated autoconvolution problem has emerged from a novel method to characterize ultra-short laser pulses \([2, 17]\). Here, we want to show the transition from the deterministic setting to the stochastic setting in a numerical example. We base our results on the deterministic paper [1].

Using the Haar-wavelet basis, the authors of [1] reformulate (27) as an equation from \( \ell_2 \) to \( \ell_2 \) by switching to the space of coefficients in the Haar basis. In order to stabilize the inversion, an \( \ell_1 \) penalty term is used such that the task is to minimize the functional
\[
J_\alpha(x) = ||F(x) - y^\delta||_2^2 + \alpha ||x||_1.
\]
The regularization parameter \( \alpha \) in (28) is chosen according to the discrepancy principle. In [1], the following formulation is used: For \( 1 < \tau_1 \leq \tau_2 \) choose \( \alpha = \alpha(\delta, y^\delta) \) such that
\[
\tau_1 \delta \leq ||F(x^\alpha_\delta) - y^\delta||_2 \leq \tau_2 \delta
\]
holds. The authors show that this leads to a convergence of the regularized solutions to a solution of (27) with minimal \( \ell_1 \)-norm of its coefficients. It was also shown that the regularization parameter fulfills
\[
\alpha(\delta, y^\delta) \to 0, \quad \frac{\delta^2}{\alpha(\delta, y^\delta)} \to 0 \quad \text{as} \quad \delta \to 0.
\]

By courtesy of Stephan Anzengruber we were allowed to use the original code for the numerical simulation in [1]. We only changed the parts directly connected to the data noise. Namely, we replaced the deterministic error \( ||y - y^\delta||_2 \leq \delta \) with i.i.d Gaussian noise,
\[
y^\eta = y + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \eta^2 I).
\]
The discretization is due to the truncation of the expansion of the functions in the Haar-basis after \( m \) elements. The parameter choice \([25]\) was realized with \( \delta \) replaced by \( \tau(\eta) \mathbb{E}(||\epsilon||_2) \) in accordance with Theorem 4.3. Instead of the correct expectation
\[
\mathbb{E}(||\epsilon||_2) = \frac{\eta}{\sqrt{2}} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})},
\]
see [16], we used the upper bound
\[
\mathbb{E}(||\epsilon||_2) \leq \eta \sqrt{m}
Figure 1. $\mathbb{E}(||\epsilon||)^2/\alpha$ (dashed) and regularization parameter $\alpha$ (solid) versus variance $\eta$. Left: A constant value of $\tau$ in the discrepancy principle with the expectation of the noise leads to the regularization parameter decreasing too fast, thus the deterministic condition $\delta^2/\alpha \to 0$ is violated (dashed line) Right: increasing $\tau$ appropriately with decreasing variance resolves this issue.

since, as shown in Section 4, the expectation has to be “blown up” anyway. In a first experiment we let $\tau(\eta) = 1.3 = \text{const}$. In this case, the numerical results suggest that the regularization parameter decreases too fast, i.e., $\frac{\tau(\eta)}{\mathbb{E}(||\epsilon||)^2}$ does not converge to zero as the requirement in (30) states; see Figure 1. For comparison, in a second run we chose $\tau(\eta) = \sqrt{1 - \log(\eta^2 2\pi m^2(\frac{3}{2})^m)}$ where $m$ is the amount of data points. This way, $\tau(\eta)\mathbb{E}(||\epsilon||)^2 \propto \rho_K(y, y^n)$. Now $\frac{\tau(\eta)}{\mathbb{E}(||\epsilon||)^2}$ converges to zero as it should be the case from (30), see Figure 1.

At this point we would like to mention that the discrepancy principle using the Ky Fan distance and the deterministic one are not completely equivalent since a different way of measuring the noise is used. Typically the stochastic noise level will be smaller (it need not bound 100% of the possible realizations) and the iteration will be stopped later than in the deterministic setup.

5.3. Linear Inverse Problems with Besov-space prior. In [18] the lifting strategy was used in a slightly different way. In particular, the Ky Fan metric was used to obtain a novel parameter choice rule. The convergence rates obtained there, however, can also be viewed in the framework of this work. The scope of that paper was to transfer the deterministic convergence results from [9] into the stochastic setting. The seminal paper [9] initiated the investigation of sparsity-promoting regularization for Inverse Problems. Looking for the solution of the linear ill-posed problem

$$Ax = y$$

between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ with given noisy data $y^\delta = y + \epsilon$, the regularization strategy was to obtain an approximation $x^\delta_\alpha$ to $x^\dagger$ via

$$x^\delta_\alpha = \min_x ||Ax - y^\delta||_2^2 + \sum_{\lambda \in \Lambda} w_{\lambda} |\langle x, \psi_\lambda \rangle|^p \psi_\lambda,$$

where $\Lambda$ is an appropriate index set, $w_{\lambda} > 0 \ \forall \lambda \in \Lambda$, $\{\psi_\lambda\}_{\lambda \in \Lambda}$ a dictionary (typically an orthonormal basis or frame) in $\mathcal{X}$ and $1 \leq p \leq 2$. Choosing a sufficiently smooth wavelet basis for $\{\psi_\lambda\}_{\lambda \in \Lambda}$ and setting $w_{\lambda} = 2^{s\|\lambda\|}$ with $\zeta = s - d(\frac{1}{2} + \frac{1}{p}) > 0$, the penalty term in (31) corresponds to a norm in the Besov space $B^{s}_{p,q}(\mathbb{R}^d)$. In [18]
the problem was formulated in a discretized version in order to formulate the MAP estimate as in [9]. We will also consider a discretized version here such that both results can be compared directly. It is assumed that the data \( y^\delta \) is given as an \( m \)-dimensional vector and that a finite index set \( \Lambda \) is used for the wavelet basis, for example by truncating the wavelet expansion after the first \( n \) components. For more information we refer to [18]. Therefore, formulating the problem of determining \( x \) from noisy data \( y^\eta = y + \epsilon \), \( \epsilon \sim \mathcal{N}(0, \eta^2 I_m) \), in the Bayesian setting with the distributions \( \pi_\epsilon(y^\delta | x) \propto \exp(-\frac{1}{2\eta^2}||Ax - y^\eta||_2^2) \) and \( \pi_{pr}(x) \propto \exp(-\frac{\tilde{\alpha}}{2}||x||_{B^p_{p,\rho}(\mathbb{R}^d)}^p) \) and using the maximum a-posteriori solution leads to the formulation

\[
\begin{align*}
x^{MAP} = \min_x ||Ax - y^\eta||_2^2 + \tilde{\alpha}\eta^2||x||_{B^p_{p,\rho}(\mathbb{R}^d)}^p,
\end{align*}
\]

where \( \eta \) is the variance of the noise and \( \tilde{\alpha} \) can roughly be described as the inverse variance of the prior. The product \( \alpha := \tilde{\alpha}\eta^2 \) gives the actual regularization parameter. In the context of Section 4.2 a deterministic operator \( A \) is considered but due to the Bayesian view the unknown \( x \) is modeled as a random variable.

It follows in direct application of Theorem 4.3 that the deterministic condition

\[
\alpha \to 0, \quad \frac{\delta^2}{\alpha} \to 0 \quad \text{as} \quad \delta \to 0,
\]

with \( \delta \) replaced by \( \rho_K(y, y^0) \) from [11] translates to the conditions

\[
\tilde{\alpha}\eta^2 \to 0, \quad \frac{\log(\eta)}{\tilde{\alpha}} \to 0 \quad \text{as} \quad \eta \to 0,
\]

leading to convergence of \( x^{MAP} \) to the unique (in case \( p = 1 \) the operator is assumed to be injective) solution \( x^\dagger \) of minimal norm \( ||\cdot||_{B^p_{p,\rho}(\mathbb{R}^d)} \) in the Ky Fan metric. The proof of convergence rates is based on two assumptions:

\[
\begin{align*}
C_l \sum_{\lambda \in \Lambda} 2^{-2|\lambda|\beta}||\langle x, \psi_\lambda \rangle||^p \leq ||Ax|| \leq C_u \sum_{\lambda \in \Lambda} 2^{-2|\lambda|\beta}||\langle x, \psi_\lambda \rangle||^p
\end{align*}
\]

where \( \beta, C_l, C_u > 0 \) and

\[
||x^\dagger||_{B^p_{p,\rho}(\mathbb{R}^d)} \leq \rho
\]

for some \( \rho > 0 \). Combining Proposition 4.5, Proposition 4.6, Proposition 4.7 from [9] it is

\[
\begin{align*}
&\left\|x_{\alpha}^\delta - x^\dagger\right\| \leq C \left( \delta + \sqrt{\delta^2 + \alpha\rho^p}\right)^{\frac{\beta}{2+\beta}} \left( \rho + \left(\frac{\rho^p + \delta^2}{\alpha}\right)^{1/p}\right)^{\frac{\beta}{2+\beta}}.
\end{align*}
\]

Translated into the stochastic setting, the right hand side of (32) reads

\[
\begin{align*}
&\mathcal{E}(\eta, m, \tilde{\alpha}) \leq \rho^{\frac{\beta}{2+\beta}}
\end{align*}
\]

where with \( L_m(\eta) = \min\{0, \eta^2 2\pi m^2 (\frac{\gamma}{2})^m\} \),

\[
\begin{align*}
\mathcal{E}(\eta, m, \tilde{\alpha}) := \eta \left( \sqrt{m - L_m(\eta)} + \sqrt{m - L_m(\eta) + \frac{\tilde{\alpha}\rho^p}{2}}\right)
\end{align*}
\]

and

\[
\tilde{\rho} = \rho + \left(\frac{\rho^p + \frac{2m - L_m(\eta)}{\tilde{\alpha}}}{2}\right)^{1/p}.
\]
We know that the deterministic rate is an upper bound to the reconstruction error whenever $||y - y^0|| = ||e|| \leq \rho_K(y, y^0)$ and $||x^t||_{B^p_{\rho,\rho}(\mathbb{R}^d)} \leq \rho$. Hence, it is

$$P \left( ||x_{\text{MAP}}^\star - x^t|| \geq C\mathcal{E}(\eta, m, \tilde{\alpha}) \right) \leq \frac{\Gamma(m, m - L_m(\eta))}{\Gamma\left(\frac{m}{2}\right)} + \frac{\Gamma(m, \frac{\delta^2}{2})}{\Gamma\left(\frac{m}{2}\right)}$$

where

$$P(||y - y^0|| > \rho_K(y, y^0)) = \frac{\Gamma(m, m - L_m(\eta))}{\Gamma\left(\frac{m}{2}\right)}$$

and

$$P(||x^t||_{B^p_{\rho,\rho}(\mathbb{R}^d)} \geq \rho) = \frac{\Gamma(m, \frac{\delta^2}{2})}{\Gamma\left(\frac{m}{2}\right)}$$

where the Besov-space functions were truncated after the first $n$ basis functions. By Definition of the Ky Fan metric, it follows immediately from (35) that

$$\rho_K(x_{\text{MAP}}) = \max \left\{ C\mathcal{E}(\eta, m, \tilde{\alpha}) \frac{\hat{\rho}^{\frac{\alpha}{\tau_p}}}{\Gamma\left(\frac{m}{2}\right)} \right\}.$$

Since $\tilde{\alpha}$ is a free parameter, we can balance the terms in (36), i.e. solve the nonlinear equation

$$C\mathcal{E}(\eta, m, \tilde{\alpha}) \frac{\hat{\rho}^{\frac{\alpha}{\tau_p}}}{\Gamma\left(\frac{m}{2}\right)} = \frac{\Gamma(m, m - L_m(\eta))}{\Gamma\left(\frac{m}{2}\right)} + \frac{\Gamma(m, \frac{\delta^2}{2})}{\Gamma\left(\frac{m}{2}\right)}$$

for $\tilde{\alpha}$. With this parameter choice rule one obtains by construction

$$\rho_K(x_{\text{MAP}}) = \mathcal{O}(\mathcal{E}(\eta, m, \tilde{\alpha}) \frac{\hat{\rho}^{\frac{\alpha}{\tau_p}}}{\Gamma\left(\frac{m}{2}\right)}).$$

We can also apply the theory developed in this work to this problem. In the deterministic setting, see [9], it was proposed to chose the regularization parameter $\alpha = \delta^2 / p^2$. Combining [9, Proposition 4.5] and [9, Proposition 4.7] then yields the rate

$$||x_{\alpha}^\delta - x^t|| \leq C \left( \frac{\delta}{C_\delta} \right) \frac{\hat{\rho}^{\frac{\alpha}{\tau_p}}}{\Gamma\left(\frac{m}{2}\right)}$$

with $C_\delta$ from (32) and some $C > 0$. Theorem 4.5 then yields in the stochastic setting the parameter choice $\alpha \sim \rho_K(y^0, y^0)^2 / p^2$ and

$$\rho_K(x_{\alpha}^t, x^t) = \mathcal{O} \left( \frac{\rho_K(y^0, y)}{C_\delta} \frac{\hat{\rho}^{\frac{\alpha}{\tau_p}}}{\Gamma\left(\frac{m}{2}\right)} \right).$$

In the notation of (34) it is for Gaussian noise $\epsilon \sim \mathcal{N}(0, \eta^2 I_m)$

$$\rho_K(y^0, y) \leq \eta \sqrt{m - L_m(\eta)},$$

see Proposition 4.1. Since $\rho_K(y^0, y) < \mathcal{E}(\eta, m, \tilde{\alpha})$, compare (34) and (39), the convergence rate in (37) is slightly slower than the one in (38), but they share the same order of convergence.
Conclusions. Our goal was to demonstrate how convergence results for Inverse Problems in the deterministic setting can be carried over into the stochastic setting. Using the Ky Fan metric, we have shown that, when only the noise is assumed to be stochastic whereas the other quantities are deterministic, this is possible in a straightforward way. Namely, assuming the knowledge of an estimate of $\rho_K(y, y^d)$, the convergence results and parameter choice follows from the deterministic setting by replacing $\delta$, which originates from the basic deterministic assumption $||y - y^d|| \leq \delta$, with $\rho_K(y, y^d)$. We have shown that, under some slight modifications, it is possible to use the expectation as measure for the magnitude of the noise. In a fully stochastic situation, where additionally to the noise other objects might be of stochastic nature, the lifting of deterministic convergence results is possible as well. However, careful analysis is necessary in order to carry the deterministic conditions over into the stochastic setting.

REFERENCES

[1] S. W. Anzengruber and R. Ramlau, Morozov’s discrepancy principle for Tikhonov-type functionals with nonlinear operators. Inverse Problems, 26 (2010), 025001, 17pp.
[2] S. Birkholz, G. Steinmeyer, S. Koke, D. Gerth, S. Bürger and B. Hofmann, Phase retrieval via regularization in self-diffraction-based spectral interferometry. JOSA B, 32 (2015), 983–992.
[3] N. Bissantz, T. Hohage and A. Munk, Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. Inverse Problems, 20 (2004), 1773–1789.
[4] N. Bissantz, T. Hohage, A. Munk and F. Ruymgaart, Convergence rates of general regularization methods for statistical inverse problems and applications. SIAM J. Numer. Anal., 45 (2007), 2610–2636.
[5] G. Blanchard and P. Mathé, Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. Inverse Problems, 28 (2012), 115011, 23pp.
[6] V. I. Bogachev, Gaussian Measures. AMS, Providence RI, 1998.
[7] D. Calvetti and E. Somersalo, An introduction to Bayesian Scientific Computing. Ten Lectures on Subjective Computing. Springer, New York, 2007.
[8] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth, On the Lambert W function. Adv. Comput. Math., 5 (1996), 329–359.
[9] I. Daubechies, M. Defrise and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm. Pure Appl. Math., 57 (2004), 1413–1457.
[10] R. M. Dudley, Real Analysis and Probability. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.
[11] D. Th. Egoroff, Sur les suites de fonctions mesurables. CR Acad. Sci. Paris, 152 (1911), 244–246.
[12] H. W. Engl, M. Hanke and A. Neubauer, Regularization of Inverse Problems. Kluwer Academic Publishers Group, Dordrecht, 1996.
[13] S. N. Evans and P. B. Stark, Inverse problems as statistics. Inverse Problems, 18 (2002), R55–R97.
[14] K. Fan, Entfernung zweier zufälligen Grössen und die Konvergenz nach Wahrscheinlichkeit. Mathematische Zeitschrift, 49 (1944), 681–683.
[15] G. Gardner, White and brown music, fractal curves and one-over-f fluctuations. Scientific American, 238 (1978), 16–32.
[16] D. Gerth, Problem-adapted Regularization for Inverse Problems in the Deterministic and Stochastic Setting. Ph.D thesis, Johannes Kepler University Linz, 2015.
[17] D. Gerth, B. Hofmann, S. Birkholz, S. Koke and G. Steinmeyer, Regularization of an autoconvolution problem in ultrashort laser pulse characterization. Inverse Probl. Sci. Eng., 22 (2014), 245–266.
[18] D. Gerth and R. Ramlau, A stochastic convergence analysis for Tikhonov regularization with sparsity constraints. Inverse Problems, 30 (2014), 055009, 24pp.
[19] R. Gorenflo and B. Hofmann, On autoconvolution and regularization. Inverse Problems, 10 (1994), 353–373.
[20] M. Hanke, A. Neubauer and O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. Numer. Math., 72 (1995), 21–37.
Lifting convergence rates 687

[21] T. Helin and M. Burger, Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31 (2015), 085009, 22pp.

[22] A. Hofinger, Ill-posed problems: Extending the Deterministic Theory to a Stochastic Setup, Ph.D thesis, Johannes Kepler University Linz, 2006.

[23] A. Hofinger, Ill-posed problems: Extending the Deterministic Theory to a Stochastic Setup, Trauner-Verlag, Linz, 2006.

[24] A. Hofinger, The metrics of prokhorov and ky fan for assessing uncertainty in inverse problems, *FWF Sitzungsber. Abt. II*, 215 (2006), 107–125, Available online [http://www.planet-austria.at/0xc1aa500d_0x00239061.pdf](http://www.planet-austria.at/0xc1aa500d_0x00239061.pdf)

[25] A. Hofinger and H. K. Pikkarainen, Convergence rate for the Bayesian approach to linear inverse problems. *Inverse Problems*, 23 (2007), 2469–2484.

[26] B. Hofmann, *Regularization for Applied Inverse and Ill-posed Problems*, Teubner Verlagsgesellschaft, Leipzig, 1986.

[27] T. Hohage and F. Werner, Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, 28 (2012), 104004, 15pp.

[28] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, 2005.

[29] H. Kekkonen, M. Lassas and S. Siltanen, Analysis of regularized inversion of data corrupted by white Gaussian noise. *Inverse Problems*, 30 (2014), 045009, 18pp.

[30] S. Kogan, *Electronic Noise and Fluctuations in Solids*, Cambridge University Press, Cambridge, 1996.

[31] M. Lassas, E. Saksman and S. Siltanen, Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3 (2009), 87–122.

[32] A. K. Louis, *Inverse und Schlecht Gestellte Probleme*, B. G. Teubner, Stuttgart, 1989.

[33] K. Mosegaard and M. Sambridge, Monte Carlo analysis of inverse problems. *Inverse Problems*, 18 (2002), R29–R54.

[34] A. Neubauer and H. K. Pikkarainen, Convergence results for the Bayesian inversion theory. *J. Inverse Ill-Posed Probl.*, 16 (2008), 601–613.

[35] A. M. Stuart, *Inverse Problems: A Bayesian perspective*, Acta Numerica, 19 (2010), 451–559.

[36] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.

Received April 2016; revised April 2017.

*E-mail address:* daniel.gerth@mathematik.tu-chemnitz.de

*E-mail address:* ronny.ramlau@jku.at