# Assessing Uncertainty in Nonlinear Inverse Problems with the Metric of Ky Fan

Andreas Hofinger* Stefan Kindermann*

October 11, 2006

**Abstract**

Recently, the metric of Ky Fan was successfully applied to assess uncertainty in linear inverse problems. In this work, we present an extension of these results to nonlinear problems. More precisely, we derive results on convergence and convergence rates for Tikhonov regularization applied to nonlinear stochastic ill-posed problems, and discuss strategies to extend these local convergence results to global ones.

## 1 Introduction

In [6, 10] an approach was presented that allows extension of many results from the deterministic theory of inverse problems to a stochastic setup. A main ingredient in the new approach was the metric of Ky Fan, which has proven to be a powerful concept to quantify convergence and convergence speeds in stochastic inverse problems. In this work, we extend the results of [6] and [10] to the nonlinear case.

We consider stochastic nonlinear inverse problems and assume that the original problem is influenced by some external random parameter $\omega$, element of a probability space $(\Omega, \mathcal{A}, \mu)$. I.e., for fixed $\omega$ we consider the nonlinear equation

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

*Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria. Spezialforschungsberich SFB F013/F1308 email: [andreas.hofinger | stefan.kindermann]@oeaw.ac.at

where $x$ and $y$ are random variables with values in Hilbert spaces $X$, $Y$, respectively, and all involved operators (e. g., $F(\cdot,\omega)$, $F'(\cdot,\omega)$, $F'^{*}(\cdot,\omega)$, ...) are random operators (cf. [11]). Furthermore, we assume that (1) admits a measurable solution. For possible assumptions on $F$ that guarantee this condition we refer to [11, Chpt. 1.5] and the references therein.

We consider problems where the solution of (1) does not depend continuously on the right hand side $y(\omega)$. This is, for instance, the case when for almost all $\omega$ the operator $F(\cdot,\omega)$ is compact, locally injective, and acts between infinite dimensional spaces $X$ and $Y$ (cf. [5]). To obtain a stable solution when only noisy data $y^{\delta}(\omega)$ are given, regularization techniques must be applied. A well known regularization method for deterministic nonlinear inverse problems is Tikhonov regularization. In this work, we extend the stochastic convergence rate analysis, introduced in [6, 10] for linear problems, to Tikhonov regularization for stochastic nonlinear problems.

While for linear problems the metrics of Ky Fan and Prokhorov where considered, we now focus on the Ky Fan metric only (see [12] for a detailed comparison of these two metrics). The Prokhorov metric works with distributions, the Ky Fan metric uses random variables to define distances. This metric is defined as follows.

**Definition 1.1 (Ky Fan metric).** *The distance of two random variables $\xi_1$, $\xi_2$ in the* Ky Fan metric *is defined as ([8], also [3])*

$$\rho_{\kappa}(\xi_1, \xi_2) := \inf \{\varepsilon > 0 \mid \mu\{\omega \in \Omega \mid d(\xi_1(\omega), \xi_2(\omega)) > \varepsilon\} < \varepsilon\} . \tag{2}$$

Note that the Ky Fan metric represents a quantitative version of *convergence in probability* and naturally leads to (multi-dimensional) confidence intervals. For some additional background on the metric of Ky Fan and connections between the metrics of Ky Fan and Prokhorov see e. g., [3, 12].

# 2 Deterministic Convergence Results

Let us first of all briefly consider the *deterministic* theory of nonlinear inverse problems. We consider problems of the form

$$F(x) = y\,, \tag{3}$$

where $F$ maps between two Hilbert spaces $X$ and $Y$. For technical reasons, in the following analysis some weak assumptions on $F$ are needed, in particular we assume that (cf. [5, 7])

(i). $F$ is continuous,

(ii). $F$ is weakly closed, i. e., for $x_n \in \mathcal{D}(F)$, $x_n \rightharpoonup x$ and $F(x_n) \rightharpoonup y$ imply $F(x) = y$ and $x \in \mathcal{D}(F)$.

For problem (3), we are interested in the $x^*$*-minimum norm solution*, where $x^*$ represents some initial guess for the solution of (3).

**Definition 2.1 ($x^*$-minimum norm solution).** *We call $x^\dagger$ an $x^*$-minimum norm solution of* (3) *if*

$$F(x^\dagger) = y$$

*and*

$$\left\|x^\dagger - x^*\right\| = \min\left\{\|x - x^*\| \mid F(x) = y\right\} .$$

Observe that such a solution needs not exist, and even if it exists, it needs not be unique.

If the solution of (3) does not depend continuously on the data, it is necessary to regularize the problem. For nonlinear operators, Tikhonov regularization is defined via the functional

$$J(x) := \left\|F(x) - y^\delta\right\|^2 + \alpha \left\|x - x^*\right\|^2 , \qquad x \in \mathcal{D}(F) \; . \tag{4}$$

A minimizer[1] of this functional is called a regularized solution $x_\alpha^\delta$. As for linear problems, one now investigates convergence of regularized solutions as the noise level $\delta$, $\|y - y^\delta\| \leq \delta$, tends to 0. In contrast to linear inverse problems (cf. [5, Theorem 5.2]), at first convergence of subsequences only can be obtained. This is due to the fact that the $x^*$-minimum norm solution need not be unique (cf. also Example 3.3). The following theorem is derived in [19] (see also [7],[5, Theorem 10.3]).

**Theorem 2.2 (Convergence).** *Let $y^\delta \in Y$, $\|y - y^\delta\| \leq \delta$ and $\alpha(\delta)$ be such that $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$. Let $x_{\alpha_k}^{\delta_k}$ denote a sequence of minimizers of* (4)*. Then $x_{\alpha_k}^{\delta_k}$ has a convergent subsequence; the limit of every convergent subsequence is a $x^*$-minimum norm solution of* (3)*. If furthermore the $x^*$-minimum norm solution $x^\dagger$ is unique, the original sequence converges and*

$$\lim x_{\alpha_k}^{\delta_k} = x^\dagger \, .$$

As in the linear case, *source conditions* are needed to obtain convergence rates. These conditions are phrased in terms of derivatives[2] of $F$ at the solution $x^\dagger$, and also involve the initial guess $x^*$.

[1]Note that for nonlinear problems the minimizer of (4) need not be unique.

[2]For some problems also *derivative free* source conditions and methods are known [16].

**Definition 2.3 (Source Condition).** *We say that $x^\dagger$ satisfies a (Hölder-) source condition with parameter $\nu$ if there exists a $v \in Y$ such that*

$$x^\dagger - x^* = (F'(x^\dagger)^* F'(x^\dagger))^\nu v \,. \tag{5}$$

Under such a source condition, and the additional assumption that the source element $v$ is sufficiently small, one obtains a convergence rate results. First of all we consider the case where (5) is satisfied with $\nu = 1/2$. Theorem 5.1 treats the case of arbitrary $0 < \nu \leq 1/2$. The following result is derived in [7] (see also [5, Theorem 10.4]).

**Theorem 2.4 (Convergence Rates).** *Let $\mathcal{D}(F)$ be convex, $y^\delta \in Y$ such that $\|y^\delta - y\| \leq \delta$ and $x^\dagger$ denote an $x^*$-minimum norm solution of* (3). *Furthermore let the following conditions hold.*

*(i). $F$ is Frechet-differentiable.*

*(ii). There exists $\gamma \geq 0$ such that $\|F'(x^\dagger) - F'(x)\| \leq \gamma \|x^\dagger - x\|$ in a sufficiently large ball $\mathcal{B}_\vartheta(x^\dagger) \cap \mathcal{D}(F)$.*

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

*(iv). The source element satisfies $\gamma \|v\| < 1$.*

*Then for the choice $\alpha = c\delta$ with some fixed $c > 0$, we obtain*

$$\left\|x^\dagger - x_\alpha^\delta\right\| \leq \frac{\delta + \alpha\|v\|}{\sqrt{\alpha}\sqrt{1-\gamma\|v\|}} = \mathcal{O}\left(\sqrt{\delta}\right) \quad \text{and} \quad \left\|F(x_\alpha^\delta) - y^\delta\right\| = \mathcal{O}\left(\delta\right).$$

In the theorem above, it is required that condition (ii) is satisfied in a sufficiently large ball $\mathcal{B}_\vartheta(x^\dagger)$. In general, $\vartheta > 2\|x^\dagger - x^*\|$ is needed, nevertheless if the solution $x^\dagger$ is unique, this may as well be relaxed to any $\vartheta > 0$ (cf. [5]).

In Theorem 4.1 we present a stochastic version of the above theorem, where conditions (ii), (iii) and (iv) are coupled with probabilities. So while e. g., $F'(\cdot, \omega)$ will have to satisfy (ii) around $x^\dagger(\omega)$, the corresponding constant $\gamma(\omega)$ need not be bounded uniformly with respect to $\omega$. These relaxations of the deterministic conditions will lead to different convergence rate results.

# 3 Convergence and Non-Convergence

In this section we transfer the deterministic convergence result of Theorem 2.2 to the stochastic problem (1) and demonstrate with a counter example that non-unique solutions can lead to non-convergence.

In the following, suppose that for almost all $\omega \in \Omega$ the operator $F(\cdot, \omega)$ satisfies conditions (i) and (ii) in section 2. For equation (1), we define the Tikhonov functional as

$$\left\| F(x(\omega), \omega) - y^\delta(\omega) \right\|^2 + \alpha \left\| x(\omega) - x^*(\omega) \right\|^2; \quad x(\omega) \in \mathcal{D}\left(F(\cdot, \omega)\right).$$

We call any minimizer of this functional a regularized solution and denote it by $x_\alpha^\delta(\omega)$. Observe that we consider pointwise minimizers here (i.e. for each $x_\alpha^\delta(\omega)$ only one realization of $y^\delta$ is needed), and not a minimizer in the Ky Fan metric (i. e., not a minimizer of a functional of the form $\rho_\kappa(F(x), y^\delta) + \alpha \rho_\kappa(x, x^*)$, which can only be computed if the whole random variable $y^\delta$ is available).

The first theorem shows (local) convergence of Tikhonov regularization in the Ky Fan metric. In [6, 10] it was possible to derive stochastic convergence rates from deterministic results via a lifting technique. Nevertheless, for the following result some more effort is necessary, since such a lifting is possible only when a quantitative bound on $\|x^\dagger(\omega) - x_\alpha^\delta(\omega)\|$ is available. Theorem 2.2 gives only the qualitative information that $x_\alpha^\delta(\omega)$ converges to $x^\dagger$, but no information about the speed[3].

In contrast to Theorem 2.2, we require uniqueness of the $x^*$-minimum norm solution; Example 3.3 shows the necessity of this assumption.

**Theorem 3.1 (Convergence).** *Let $y^\delta$ be such that $\rho_\kappa(y, y^\delta) \leq \delta$. Let $\alpha(\delta)$ satisfy $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$. Furthermore, let the $x^*$-minimum norm solution $x^\dagger$ be unique for almost all $\omega$. Then*

$$\lim_{\delta \to 0} \rho_\kappa(x^\dagger, x_{\alpha(\delta)}^\delta) = 0.$$

*Proof.* Consider a sequence $y^{\delta_k}$ with $\rho_\kappa(y, y^{\delta_k}) \leq \delta_k$ and $\delta_k \to 0$. Let $\eta := \limsup \rho_\kappa(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k})$. (Note that $0 \leq \eta \leq 1$ due to properties of the Ky Fan metric.) We show in the following that for arbitrary $\varepsilon > 0$, we have $\eta/2 \leq \varepsilon$ and consequently $\limsup \rho_\kappa(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k}) = \lim \rho_\kappa(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k}) = 0$.

As a first step, we pick a "worst case" subsequence of $y^{\delta_k}$, a subsequence where the corresponding solutions satisfy $\rho_\kappa(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k}) \geq \eta/2$ and therefore $\liminf \rho_\kappa(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k}) \geq \eta/2$ (without loss of generality we may choose the

[3] It is well-known that it cannot give more, cf. Proposition 3.11 in [5].

sequence itself). We now show that even from this "worst case" sequence we can pick a subsequence with $\limsup \rho_{\kappa}(x^\dagger, x_{\alpha(\delta_{k^j})}^{\delta_{k^j}}) \le \varepsilon$.

Let $\varepsilon > 0$. Convergence in the Ky Fan metric implies pointwise convergence of subsequences [3, 4]. Moreover, by the quantitative statement in [12, Proposition 2.2], we can pick a subsequence $y^{\delta_{k^j}}$ and a set $\Omega_\varepsilon$ with $\mu(\Omega_\varepsilon) \ge 1 - \frac{\varepsilon}{2}$ as well as $\|y^{\delta_{k^j}}(\omega) - y(\omega)\| \le 2\delta_{k^j}$ on $\Omega_\varepsilon$. For all $\omega \in \Omega_\varepsilon$, the noise level $\delta_{k^j}$ tends to zero. Therefore, Theorem 2.2 implies convergence[4] of $x_\alpha^{\delta_{k^j}}(\omega)$ to the unique solution $x^\dagger(\omega)$ for $\delta_{k^j} \to 0$ and $\omega \in \Omega_\varepsilon$. This convergence need not be uniform in $\omega$; nevertheless, pointwise convergence implies uniform convergence when sets of small measure are omitted (Egoroff's theorem on almost uniform convergence; see e. g., [2, 3]). Thus for $\delta_{k^j}$ sufficiently small

$$\mu\left(\omega \in \Omega_\varepsilon \mid \left\|x_\alpha^{\delta_{k^j}}(\omega) - x^\dagger(\omega)\right\| > \varepsilon\right) \le \varepsilon/2\,.$$

So in total, we have shown existence of a subsequence $\delta_{k^j}$ such that

$$\mu\left(\omega \in \Omega \mid \left\|x_\alpha^{\delta_{k^j}}(\omega) - x^\dagger(\omega)\right\| > \varepsilon\right) \le \varepsilon$$

for $\delta_{k^j}$ sufficiently small. Rewriting this probability estimate using the Ky Fan metric, we obtain $\limsup_{j\to\infty} \rho_{\kappa}(x_\alpha^{\delta_{k^j}}, x^\dagger) \le \varepsilon$. On the other hand, the original sequence satisfied $\liminf_{k\to\infty} \rho_{\kappa}(x^\dagger, x_{\alpha(\delta_k)}^{\delta_k}) \ge \eta/2$.

Since $\liminf_{k\to\infty} \rho_{\kappa}(x_\alpha^{\delta_k}, x^\dagger) \le \limsup_{j\to\infty} \rho_{\kappa}(x_\alpha^{\delta_{k^j}}, x^\dagger)$, we obtain $\eta/2 \le \varepsilon$. Because $\varepsilon > 0$ was arbitrary, $\eta = 0$, which concludes the proof. $\square$

**Remark 3.2.** As opposed to Theorem 2.2, we assumed (local) uniqueness of the $x^*(\omega)$-minimum-norm solutions $x^\dagger(\omega)$ to deduce convergence in Theorem 3.1. In the deterministic theorem—even without assuming uniqueness—at least the convergence of subsequences can be shown. Nevertheless, this convergence of subsequences cannot be lifted to a stochastic setting without additional assumptions on $x^\dagger$, since the subsequences (the corresponding enumerations) for different $\omega$ do not have to be related. As soon as $\Omega$ is not finite, also diagonalization arguments fail.

In the following we construct a (counter-)example, where the noise tends to 0 in the Ky Fan metric, but $x_\alpha^\delta(\omega)$ has no convergent subsequence. The main ingredients for this construction are a non-unique $x^*$-minimum norm solution and an uncountable probability space $\Omega$.

[4] Uniqueness of $x^\dagger(\omega)$ is needed for this step; otherwise we find only convergent *subsequences*, and cannot perform the subsequent lifting argument.

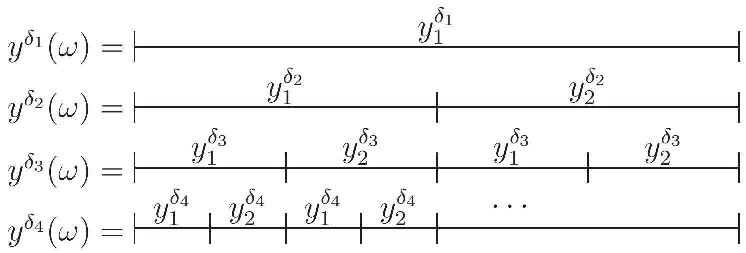


Figure 1: Construction of a sequence $y^{\delta_k}(\omega)$ with $\rho_{\kappa}(y, y^{\delta_k}) \leq \delta_k$ and $\rho_{\kappa}(x^\dagger, x_k^\delta) \geq 1/2$.

**Example 3.3 (Non-convergence).** Consider an equation $F(x) = y$ with two solutions $x_1^\dagger$, $x_2^\dagger$, and suppose that for some sequence $\delta_k \to 0$ we can construct $y_1^{\delta_k}$, $y_2^{\delta_k}$, with the following properties

- $\|y_i^{\delta_k} - y\| \leq \delta_k$ for $i = 1, 2$ and
- the regularized solutions $x_{\alpha,i}^{\delta_k}$ obtained from the sequence $y_i^{\delta_k}$ converge to the solution $x_i^\dagger$ as $\delta_k \to 0$.

For the sake of simplicity, we suppose that $\|x_1^\dagger - x_2^\dagger\| \geq 1$, and that the regularized solutions are sufficiently close to their limits, such that always $\|x_{\alpha,1}^{\delta_i} - x_{\alpha,2}^{\delta_j}\| \geq 1/2$. (An inverse problem that exhibits these characteristics is presented in Example 3.4 below.) Under these assumptions we can construct a sequence of noisy data $y^{\delta_k}(\omega)$ that converges to $y(\omega) = y$ in the Ky Fan metric, but where the corresponding solutions obtained via Tikhonov regularization do not.

In Figure 1, we consider $\Omega = [0, 1]$, and show an appropriate construction of noisy data $y^{\delta_k}$. For $k = 1$ we pick $y^{\delta_1}(\omega)$ such that all regularized solutions are close to $x_1^\dagger$; for $k = 2$, $x_\alpha^{\delta_2}(\omega)$ is close to $x_1^\dagger$ when $\omega \in [0, \frac{1}{2}]$, and close to $x_2^\dagger$ when $\omega \in (\frac{1}{2}, 1]$; for $k = 3$ the domain is split into 4 sub-domains and so forth. According to the assumptions above, this situation happens for noisy data of the form

$$
\begin{aligned}
y^{\delta_1}(\omega) &= y_1^{\delta_1} \\
y^{\delta_2}(\omega) &= y_1^{\delta_2}\,\chi_{[0,\frac{1}{2}]}(\omega) + y_2^{\delta_2}\,\chi_{(\frac{1}{2},1]}(\omega) \\
y^{\delta_3}(\omega) &= y_1^{\delta_3}\left(\chi_{[0,\frac{1}{4}]}(\omega) + \chi_{(\frac{1}{2},\frac{3}{4}]}(\omega)\right) + y_2^{\delta_3}\left(\chi_{(\frac{1}{4},\frac{1}{2}]}(\omega) + \chi_{(\frac{3}{4},1]}(\omega)\right) \\
&\ \ \vdots
\end{aligned}
$$

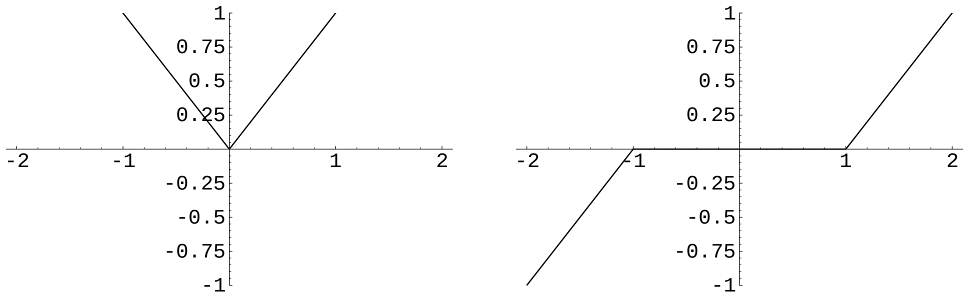


Figure 2: The two components $f_1$ (left) and $f_2$ (right) of the function $f$ defined in Example 3.4.

Now clearly, $y^{\delta_k}$ tends to $y$ in the Ky Fan metric, since with probability 1 we have $\|y^{\delta_k}(\omega) - y\| \leq \delta_k$, so

$$\rho_{\text{K}}(y^{\delta_k},\, y) \leq \delta_k\,.$$

To estimate the distance between the regularized solutions, we observe that under the assumptions above for any $i \neq j$, on a set of measure 1/2, $x_\alpha^{\delta_i}(\omega)$ is close to $x_1^\dagger$, while $x_\alpha^{\delta_j}(\omega)$ is close to $x_2^\dagger$. This implies

$$\rho_{\text{K}}(x_\alpha^{\delta_i},\, x_\alpha^{\delta_j}) \geq \frac{1}{2} \qquad \forall i \neq j\,.$$

The corresponding sequence of solutions has no convergent subsequence.

**Example 3.4 (Multiple Solutions).** Let the nonlinear function $f : \mathbb{R} \to \mathbb{R}^2$ be defined as

$$f(x) = (f_1(x),\, f_2(x))\,,$$

with

$$f_1(x) = |x| \quad \text{and} \quad f_2(x) = \begin{cases} x+1 & x \leq -1 \\ 0 & -1 < x < 1 \\ x-1 & 1 \leq x. \end{cases}$$

The graph of this function is shown in Figure 2. Observe that for any $y \in \mathbb{R}^2$ of the form $y = (y_1, 0)$ with $0 < y_1 \leq 1$ there are two solutions for the equation $f(x) = y$, namely $x_1 = y_1$ and $x_2 = -y_1$. As soon as $|y_1| > 1$ the solutions become unique, due to the influence of the second component of $f$.

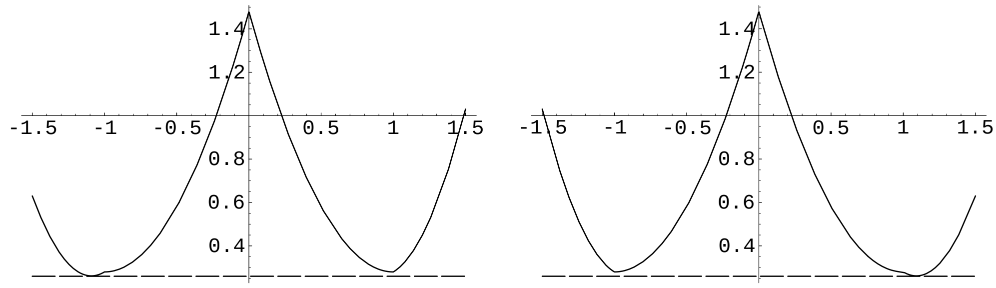


Figure 3: The Tikhonov functional for Example 3.4 with $\alpha = 0.2$ and the cases $d = -0.2$ (left) and $d = +0.2$ (right).

Suppose that the exact data for the equation $f(x) = y$ is given by $y = (1, 0)$, which yields the solutions $x_1^\dagger = -1$ and $x_2^\dagger = 1$. Suppose further that the noisy data $y^\delta$ are given as

$$y^\delta = (1 + |d|,\, d) \quad \left\|y - y^\delta\right\| = \sqrt{2}|d| =: \delta\,, \quad d \in \mathbb{R}.$$

For arbitrarily small noise level $\delta = \sqrt{2}|d|$, the sign of $d$ decides whether the regularized solution is close to $x_1^\dagger$ or close to $x_2^\dagger$ (cf. Figure 3): When $\alpha$ is chosen as $\alpha = |d| = \delta/\sqrt{2}$, the regularized solutions are

$$x_\alpha^\delta = \operatorname{sign}(d) \left(1 + \frac{|d|}{2 + |d|}\right) .$$

In particular $|x_\alpha^\delta| > 1$, so for arbitrarily small noise level $|x_\alpha^\delta - x_\alpha^{-\delta}| \geq 2$; the obtained construction satisfies the assumptions in Example 3.3 above.

So as we have seen, if no additional restrictions on solutions of (1) are imposed, even a convergent subsequence of regularized solutions need not exist. These problems disappear as soon as source conditions are assumed, because these ensure local uniqueness of solutions.

# 4 Convergence Rates

The next theorem gives a result on *convergence rates*. As in the linear case (see [6]), source conditions are needed to obtain rates. These conditions are typically of the form $x^\dagger \in \mathcal{R}((F'(x^\dagger)^* F'(x^\dagger))^\nu)$, which is an abstract smoothness requirement on the true solution $x^\dagger$.

The major difference in the results for linear and nonlinear problems is that we must require conditions on the smallness of the corresponding source element. A discussion of two possibilities to handle this issue is given in section 5.

In the following theorem, we treat the case $\nu = 1/2$. Some of the conditions in Theorem 2.4 are required to be satisfied almost surely; others are required with a certain probability only.

**Theorem 4.1 (Convergence Rate).** *Let $\mathcal{D}(F)$ be convex, let $y^\delta$ be such that $\rho_\kappa(y, y^\delta) \leq \delta$ and let $x^\dagger(\omega)$ be an $x^*$-minimum norm solution for almost all $\omega$. Moreover, let the following conditions hold*

*(i). $F(\cdot, \omega)$ is Frechet-differentiable for almost all $\omega$.*

*(ii). $F'(\cdot, \omega)$ satisfies*

$$\left\|F'(x^\dagger(\omega), \omega) - F'(x, \omega)\right\| \leq \gamma(\omega) \left\|x^\dagger(\omega) - x\right\|$$

*in a ball $\mathcal{B}_\vartheta(x^\dagger(\omega))$ with $\vartheta \geq 2\|x^\dagger(\omega) - x^*(\omega)\| + \varepsilon$ and $\varepsilon > 0$ (cf. Remark 4.2).*

*(iii). (source condition) $\mu(\Omega_{\mathrm{sc}}) = 1$ where*

$$\Omega_{\mathrm{sc}} := \left\{\omega \mid \exists v(\omega),\, x^\dagger(\omega) - x^*(\omega) = F'(x^\dagger(\omega), \omega)^* v(\omega)\right\}.$$

*(iv). (closedness condition) With $\gamma(\omega)$ as in (ii) and $v(\omega)$ as in (iii),*

$$\mu\left(\omega \in \Omega_{\mathrm{sc}} \mid \gamma(\omega)\, \|v(\omega)\| > \xi\right) < \varphi_{\mathrm{cl}}(\xi), \quad \lim_{\xi \to 1^-} \varphi_{\mathrm{cl}}(\xi) = 0.$$

*(v). (decay condition) With $v(\omega)$ as in (iii),*

$$\mu\left(\omega \in \Omega_{\mathrm{sc}} \mid \|v(\omega)\| > \tau\right) < \varphi_{\mathrm{de}}(\tau), \quad \lim_{\tau \to \infty} \varphi_{\mathrm{de}}(\tau) = 0.$$

*Then for $\alpha$ with $\alpha \to 0$ and $\delta^2/\alpha \to 0$ sufficiently small,*

$$\rho_\kappa(x^\dagger, x_\alpha^\delta) \leq \inf_{\substack{\tau < \infty \\ \xi \in (0,1)}} \max\left\{\delta + \varphi_{\mathrm{cl}}(\xi) + \varphi_{\mathrm{de}}(\tau), \frac{\delta + \alpha\tau}{\sqrt{\alpha}\sqrt{1-\xi}}\right\}.$$

*In particular, for the choice $\alpha(\delta) \sim \delta$ we obtain*[5]

$$\rho_\kappa(x^\dagger, x_{\alpha(\delta)}^\delta) \leq \inf_{\substack{\tau < \infty \\ \xi \in (0,1)}} \max\left\{\delta + \varphi_{\mathrm{cl}}(\xi) + \varphi_{\mathrm{de}}(\tau), \sqrt{\delta}\frac{\mathcal{O}(1+\tau)}{\sqrt{1-\xi}}\right\}. \tag{6}$$

[5]For an alternative parameter selection method $\alpha = \alpha(\delta, \varphi_{\mathrm{de}}(\cdot))$ cf. Remark 4.3.

*Proof.* Since Definition 1.1 of the Ky Fan metric involves an infimum, we must for technical reasons first of all suppose that $\delta > \rho_{\kappa}(y, y^\delta)$. With probability $1-\delta$ the noise level can be estimated by $\|y - y^\delta\| \leq \delta$. Now fix $\delta$ and choose $\xi < 1$ and $0 < \tau < \infty$. With probability $1 - (\varphi_{\mathrm{cl}}(\xi) + \varphi_{\mathrm{de}}(\tau))$ conditions (iv) and (v) are satisfied for these values of $\xi$ and $\tau$. For the corresponding values of $\omega$ we apply Theorem 2.4 and obtain the estimate

$$\left\|x^\dagger(\omega) - x^\delta_{\alpha(\delta)}(\omega)\right\| \leq \frac{\delta + \alpha\tau}{\sqrt{\alpha}\sqrt{1-\xi}}\,,$$

when $\delta^2/\alpha \leq \varepsilon$, with $\varepsilon$ as in (ii) (cf. the proof in [5, Thm. 10.4]). Fixing the parameter choice $\alpha \sim \delta$,

$$\left\|x^\dagger(\omega) - x^\delta_{\alpha(\delta)}(\omega)\right\| \leq \sqrt{\delta}\frac{c(1+\tau)}{\sqrt{1-\xi}}\,.$$

This estimate holds on a set with measure greater or equal to $1-(\delta+\varphi_{\mathrm{cl}}(\xi)+\varphi_{\mathrm{de}}(\tau))$, the Ky Fan metric can therefore be bounded as

$$\rho_{\kappa}(x^\dagger, x^\delta_{\alpha(\delta)}) \leq \max\left\{\delta + \varphi_{\mathrm{cl}}(\xi) + \varphi_{\mathrm{de}}(\tau), \frac{\delta + \alpha\tau}{\sqrt{\alpha}\sqrt{1-\xi}}\right\}\,.$$

This estimate is valid for arbitrary choices of $\xi$ and $\tau$ above, therefore we may bound the Ky Fan distance of $x^\dagger$ and $x^\delta_{\alpha(\delta)}$ by taking the infimum with respect to $\xi$ and $\tau$. Since all functions involving $\delta$ are continuous from the right, we obtain the same estimate also for $\delta = \rho_{\kappa}(y, y^\delta)$ $\square$

**Remark 4.2.** In the theorem above we assumed that $F'(\cdot, \omega)$ satisfies condition (ii) in a sufficiently large ball. In the deterministic Theorem 2.4, there was a distinction whether the solution is unique or not; for unique solutions the condition could be relaxed to any ball around $x^\dagger$. In contrast to the deterministic theory, here the condition cannot be relaxed even for unique solutions. In the deterministic setting the fact was used that after a certain, but finite fade-in phase the proposed rate is obtained. In the stochastic setting, this phase may have different lengths for different realizations of $x^\dagger(\omega)$; therefore such an argument fails.

We now consider some special cases for the form of $\varphi_{\mathrm{cl}}(\cdot)$ and $\varphi_{\mathrm{de}}(\cdot)$ above. An additional discussion of the smallness condition is given in section 5.2.

**Remark 4.3.** In the first two cases the operator is assumed to be deterministic, i. e., $F(\cdot, \omega) = F(\cdot)$, with $\gamma(\omega) = \gamma = 1$.

First of all suppose $\|v\| \in U[0,1]$, i. e., it is uniformly distributed on the interval $[0,1]$. We therefore have $\varphi_{\mathrm{cl}}(\xi) = 1-\xi$, as well as $\varphi_{\mathrm{de}}(\tau) = 0$ for $\tau > 1$. Thus Theorem 4.1 implies

$$\rho_{\kappa}(x^{\dagger},\, x^{\delta}_{\alpha(\delta)}) \leq \inf_{0<\alpha<\infty} \inf_{\xi\in(0,1)} \max\left\{\delta + 1 - \xi, \frac{\delta+\alpha}{\sqrt{\alpha}\sqrt{1-\xi}}\right\}.$$

Neglecting the $\delta$-term in the first part we obtain $1-\xi = (\delta+\alpha)/(\sqrt{\alpha}\sqrt{1-\xi})$, which gives for $\alpha \sim \delta$

$$\rho_{\kappa}(x^{\dagger},\, x^{\delta}_{\alpha(\delta)}) = \mathcal{O}\left(\delta^{1/3}\right) .$$

Observe that in this case $\alpha \sim \delta$ is the optimal choice, independent of the structure of $\varphi_{\mathrm{cl}}(\xi)$, since

$$\frac{\delta+\alpha}{\sqrt{\alpha}\sqrt{1-\xi}} = \frac{\sqrt{\delta}}{\sqrt{1-\xi}}\left(\sqrt{\frac{\delta}{\alpha}} + \sqrt{\frac{\alpha}{\delta}}\right)$$

shows the optimal rate when the two terms in parentheses are balanced.

For the second case suppose $\varphi_{\mathrm{de}}(\tau) = c\tau^{-e}$ (cf. [6, Remark 3.4]). Since $\varphi_{\mathrm{cl}}(\xi) \nrightarrow 0$, but $\varphi_{\mathrm{cl}}(\xi) \geq c$ we obtain

$$\rho_{\kappa}(x^{\dagger},\, x^{\delta}_{\alpha(\delta)}) \leq \inf_{0<\alpha<\infty} \inf_{\substack{\tau<\infty \\ \xi\in(0,1)}} \max\left\{c + c\tau^{-e}, \frac{\delta+\tau\alpha}{\sqrt{\alpha}\sqrt{1-\xi}}\right\}.$$

The right hand side does not tend to zero, we do not get any convergence rate (nevertheless, the method still converges according to Theorem 3.1).

Finally, consider the case when both (iv) and (v) 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 $\varphi_{\mathrm{cl}}(\xi) = 1-\xi$ and $\varphi_{\mathrm{de}}(\tau) = c/(1+\tau)$ are compatible realizations of $\varphi_{\mathrm{cl}}(\cdot)$ and $\varphi_{\mathrm{de}}(\cdot)$. To get the infimal bound for the Ky Fan distance, we clearly need $\varphi_{\mathrm{cl}}(\xi) \sim \varphi_{\mathrm{de}}(\tau)$. This allows us to eliminate $\xi$ and yields

$$\frac{\tilde{c}}{1+\tau} = \frac{\delta+\alpha\tau}{\sqrt{\alpha}\sqrt{c/(1+\tau)}} = \sqrt{\frac{1+\tau}{c}}\sqrt{\delta}\left(\sqrt{\frac{\delta}{\alpha}} + \sqrt{\frac{\alpha}{\delta}}\tau\right)$$

The term in parentheses is balanced when $\delta/\alpha \sim \tau$. Since for convergence we need $\tau \to \infty$ we can replace $1+\tau$ by $\tau$, and obtain $\tau \sim \delta^{-1/4}$. Altogether, this implies the rate

$$\rho_{\kappa}(x^{\dagger},\, x^{\delta}_{\alpha(\delta)}) = \mathcal{O}\left(\delta^{\frac{1}{4}}\right) ,$$

under the parameter choice $\alpha = \alpha(\delta, \varphi_{\mathrm{de}}(\cdot)) \sim \delta^{5/4}$.

**Remark 4.4.** In the remark above, we considered the case where $\gamma(\omega)$ can be arbitrarily small. Observe that, when $\gamma(\omega)$ is bounded from below, condition (iv) implies (v). In this case all realizations of $x^\dagger(\omega)$ and $x^*(\omega)$ satisfy a source condition with a similar source element (namely with $\|v(\omega)\| \leq 1/\min\gamma < \infty$). The resulting convergence rate is (besides the influence of $\xi$) the same as in the deterministic case.

Condition (v) in Theorem 4.1 seems more natural than condition (iv): While the first one ensures that all possible solutions carry some common smoothness (e. g., are twice differentiable), the latter one also requires quantitative bounds on this smoothness (e. g., the $H^2$-norm of $x^\dagger(\omega)$ is less than 1 a. s.). This is certainly a strong requirement. In the next section, we present two possibilities how to alter the assumptions of Theorem 4.1 to avoid condition (iv).

# 5 Globalization of Convergence

In nonlinear regularization theory, convergence of the approximate solution is in general only obtained in a neighborhood of the exact solution. Even for the deterministic theory, convergence results usually require some smallness condition (for instance the source element $v$ or $x^\dagger - x^*$ has to be sufficiently small). Although this is somehow unsatisfactory, it cannot be avoided due to the nonlinearity of the problem.

However, in this section we want to discuss remedies for this situation. There are at least two approaches possible: Since the difficulty arises from the nonlinearity of the operator one idea to get rid of the smallness conditions is to impose additional conditions on the nonlinear operator to obtain better estimates. This idea is exposed in section 5.1.

On the other hand if additional conditions are not possible for the problem of interest one can think of means to find a good (i. e., sufficiently close) initial guess for the cost of an additional effort. This concept is discussed in section 5.2 for some realistic examples.

Two additional approaches, leading to logarithmic convergence rates are given in [11, Ch.4.3].

## 5.1 Nonlinearity Condition

To obtain convergence rates for an *iterative* regularization method, in addition to a source condition, a so-called nonlinearity condition on the operator is needed. Such conditions are well known; popular choices are for instance a tangential cone condition for Landweber iteration [9], or a range invariance condition for Gauss-Newton-type iterations [15]. For Tikhonov regularization similar conditions were used, e. g., in [14]. A discussion of several kinds of nonlinearity conditions can be found, e. g., in [13, 18].

In this section, we prove convergence rate estimates for Tikhonov regularization under a source condition on the initial guess and a nonlinearity condition on the operator, but without a smallness condition.

At first, we establish estimates for the deterministic case similar to Theorem 2.4. Instead of condition (iv) in Theorem 2.4, we require the following *tangential cone condition* for all $z \in \mathcal{D}(F)$:

$$\left\|F(z) - F(x^\dagger) - F'(x^\dagger)(z - x^\dagger)\right\| \leq C \left\|F(z) - F(x^\dagger)\right\| . \tag{iv'}$$

Here, $C$ is some constant independent of $x^\dagger$ and $z$. With this condition, we obtain the following theorem.

**Theorem 5.1 (Deterministic Convergence Rate).** *Let $\mathcal{D}(F)$ be convex, $y^\delta \in Y$ such that $\|y^\delta - y\| \leq \delta$, and $x^\dagger$ denote an $x^*$-minimum norm solution of* (3)*. If $F$ is Frechet-differentiable, where* (iv') *holds and a source condition* (5) *is satisfied with $\nu \leq \frac{1}{2}$, we have the estimate*

$$\left\|x_\alpha^\delta - x^\dagger\right\| \leq 2\sqrt{2}\frac{\delta}{\sqrt{\alpha}} + 2(1+C)^{2\nu} \|v\| \alpha^\nu, \tag{7}$$

*where $x_\alpha^\delta$ is a minimizer of* (4)*.*

*Proof.* For the residual and the error we use the abbreviations

$$r_\alpha^\delta := \left\|F(x_\alpha^\delta) - F(x^\dagger)\right\| \quad \text{and} \quad e_\alpha^\delta := \left\|x_\alpha^\delta - x^\dagger\right\| .$$

Starting with the well-known estimate (see [5, Thm.10.4])

$$\left\|F(x_\alpha^\delta) - y^\delta\right\|^2 + \alpha \left\|x_\alpha^\delta - x^\dagger\right\|^2 \leq \delta^2 + 2\alpha \left\langle x_\alpha^\delta - x^\dagger, x^* - x^\dagger\right\rangle ,$$

expansion of the square and use of the source condition lead to

$${r_\alpha^\delta}^2 + \alpha {e_\alpha^\delta}^2 \leq -2\left\langle r_\alpha^\delta, y^\delta - y\right\rangle + 2\delta^2 + 2\alpha \left\langle (F'(x^\dagger)^* F'(x^\dagger))^\nu (x_\alpha^\delta - x^\dagger), v\right\rangle .$$

An interpolation inequality [5, (2.49)] gives (with $\beta := 2\nu$)

$$\left\|F'(x^\dagger)^* F'(x^\dagger))^\nu (x_\alpha^\delta - x^\dagger)\right\| \leq \left\|x_\alpha^\delta - x^\dagger\right\|^{1-\beta} \left\|F'(x^\dagger)(x_\alpha^\delta - x^\dagger)\right\|^\beta .$$

From the nonlinearity condition (iv') we conclude that

$$\left\|F'(x^\dagger)(x_\alpha^\delta - x^\dagger)\right\| \leq (1+C) r_\alpha^\delta .$$

These inequalities together with the Cauchy- and the Young-inequality[6] give

$$r_\alpha^{\delta\,2} + \alpha e_\alpha^{\delta\,2} \le \frac{1}{2} r_\alpha^{\delta\,2} + 4\delta^2 + 2\,\|v\|\,(1+C)^\beta \alpha e_\alpha^{\delta\,1-\beta} r_\alpha^{\delta\,\beta}.$$

For any positive numbers $\epsilon_1$, $\epsilon_2$, we can rewrite the factor in the last term as

$$\alpha e_\alpha^{\delta\,1-\beta} r_\alpha^{\delta\,\beta} = (\epsilon_1^{\frac{2-2\beta}{1+\beta}} \epsilon_2^{\frac{2\beta}{1+\beta}} \alpha)^{\frac{1+\beta}{2}} (\frac{\sqrt{\alpha} e_\alpha^\delta}{\epsilon_1})^{1-\beta} (\frac{r_\alpha^\delta}{\epsilon_2})^\beta,$$

and apply the following Young's inequality:

$$xyz \le \frac{1}{\gamma_1} x^{\gamma_1} + \frac{1}{\gamma_2} x^{\gamma_2} + \frac{1}{\gamma_3} x^{\gamma_3} \quad \text{if } x, y, z \ge 0, \text{ and } \sum_{i=1}^{3} \frac{1}{\gamma_i} = 1$$

(this follows from the convexity of the exponential function). With

$$\gamma_1 = 2, \quad \gamma_2 = \frac{2}{1-\beta}, \quad \gamma_3 = \frac{2}{\beta},$$

this leads to

$$\begin{aligned} r_\alpha^{\delta\,2} + \alpha e_\alpha^{\delta\,2} &\le \frac{1}{2} r_\alpha^{\delta\,2} + 4\delta^2 \\ &\quad + 2(1+C)^\beta \|v\| \left( \frac{1}{2} \left( \epsilon_1^{\frac{2-2\beta}{1+\beta}} \epsilon_2^{\frac{2\beta}{1+\beta}} \alpha \right)^{(1+\beta)} + \frac{1}{\gamma_2 \epsilon_1^2} \alpha e_\alpha^{\delta\,2} + \frac{1}{\gamma_3 \epsilon_2^2} r_\alpha^{\delta\,2} \right). \end{aligned}$$

We choose

$$\epsilon_1^2 = 2(1-\beta)(1+C)^\beta \|v\|, \quad \epsilon_2^2 = 2\beta(1+C)^\beta \|v\|$$

to get

$$\frac{\alpha}{2} e_\alpha^{\delta\,2} \le 4\delta^2 + 2\left((1+C)^\beta \|v\|\right)^2 \alpha^{1+\beta} \beta^\beta (1-\beta)^{1-\beta}$$

Since $\beta^\beta(1-\beta)^{1-\beta} \le 1$ and $\sqrt{x^2+y^2} \le |x| + |y|$, we obtain the assertion. □

With this global convergence rate result, we are able to derive a convergence rate result similar to Theorem 4.1, but without a smallness condition. As a by-product of the nonlinearity condition, we can establish rates under more general stochastic source conditions of Hölder type.

[6] $2|ab| \le \frac{1}{2}a^2 + 2b^2$

**Theorem 5.2 (Stochastic Convergence Rate).** *Let $\mathcal{D}(F)$ be convex, $y^\delta(\omega) \in Y$ such that $\rho_\kappa(y^\delta, y) \leq \delta$ and $x^\dagger(\omega)$ be an $x^*$-minimum norm solution of* (3) *for almost all $\omega$. Furthermore, let for some $0 < \nu \leq 1/2$ the following conditions hold.*

*(i). $F(\cdot,\omega)$ is Frechet-differentiable for almost all $\omega$.*

*(ii). $\forall z \in \mathcal{D}(F)$, $F(\cdot,\omega)$ satisfies*

$$\left\|F(z,\omega) - F(x^\dagger,\omega) - F'(x^\dagger,\omega)(z - x^\dagger)\right\| \leq C(\omega) \left\|F(z,\omega) - F(x^\dagger,\omega)\right\|$$

*(iii). (source condition) $\mu(\Omega_{sc,\nu}) = 1$ where*

$$\Omega_{sc,\nu} := \left\{\omega \mid \exists v(\omega),\, x^\dagger(\omega) - x^*(\omega) = (F'(x^\dagger(\omega),\omega)^* F'(x^\dagger(\omega),\omega))^\nu v(\omega)\right\} .$$

*(iv). (decay condition)*

$$\mu\left(\omega \in \Omega_{sc,\nu} \mid (1 + C(\omega))^{2\nu} \left\|v(\omega)\right\| > \tau\right) < \varphi_{\mathrm{de}}(\tau), \quad \lim_{\tau\to\infty} \varphi_{\mathrm{de}}(\tau) = 0\,.$$

*Then for any $\alpha > 0$ we can estimate*

$$\rho_\kappa(x^\dagger,\, x_\alpha^\delta) \leq \inf_{\tau<\infty} \max\left\{\delta + \varphi_{\mathrm{de}}(\tau),\, 2\sqrt{2}\frac{\delta}{\sqrt{\alpha}} + 2\tau\alpha^\nu\right\} .$$

*Proof.* If the decay condition and the nonlinearity condition holds, we may conclude as in the proof of Theorem 4.1 for $\delta > \rho_\kappa(y^\delta,\, y)$ with (7) that

$$\left\|x^\dagger(\omega) - x_\alpha^\delta(\omega)\right\| \leq 2\sqrt{2}\frac{\delta}{\sqrt{\alpha}} + 2\tau\alpha^\nu$$

with probability larger than $1 - (\delta + \varphi_{\mathrm{de}}(\tau))$. By definition of the Ky-Fan metric the assertion follows with technical arguments as in the proof of Theorem 4.1. $\square$

The previous theorem establishes convergence rates as in Theorem 4.1 under more general source conditions. Note that as in section 4, with an appropriate parameter choice rule of $\alpha$ we get convergence rates for the error in terms of the Ky-Fan metric. Moreover, the estimates in Theorem 5.1 are of similar order in terms of $\alpha$ as for the linear case (compare [5]), and hence can be expected to be order optimal.

Of course, the nonlinearity conditions in this sections are only one possibility, other ones are conceivable as well. The aim of the previous theorem is mainly to exemplify that such conditions give way to stochastic convergence rates without closeness restrictions on the initial guess $x^*(\omega)$ such as (iv) in Theorem 4.1; similar results under alternative conditions are certainly possible and can be derived as above once the deterministic rates are established.

Admittedly, in applications nonlinearity conditions are hard to verify and for many interesting inverse problems this task is still an open problem. However, it should be noted, that there are numerical examples where local convergence can be observed, even though the nonlinearity condition cannot be proven.

In some cases, it might be unrealistic to assume that nonlinearity conditions holds, in such a situation a good initial guess is vital. In the next section, we discuss this case in certain applications.

## 5.2 Increasing the Effort to generate an Initial Guess

In this section, we present a second approach to circumvent the smallness condition (iv) of Theorem 4.1. Instead of imposing additional assumptions on the operator $F$, we replace condition (iv) by a more applicable condition. We do not assume that the norm of the source element $v(\omega)$ is uniformly bounded, but introduce an additional parameter, that represents the effort to obtain a good initial guess (and consequently a small source element $v(\omega)$). The new condition can be formulated as

$$\mu\{\omega \in \Omega \mid \gamma(\omega)\,\|v(\omega)\| > c\} < \varphi_{\mathrm{cl}}(c,\kappa), \qquad \kappa \in K\,. \tag{iv”}$$

Using this assumption, (6) turns, for fixed $c < 1$, into

$$\rho_{\kappa}(x^\dagger,\, x^\delta_{\alpha(\delta)}) \le \inf_{\tau<\infty} \max\left\{\delta + \varphi_{\mathrm{cl}}(c,\kappa) + \varphi_{\mathrm{de}}(\tau), \mathcal{O}\left(\sqrt{\delta}(1+\tau)\right)\right\}. \tag{6”}$$

In the following, we demonstrate that this approach is well applicable to problems, where accurate direct measurements are in principle possible but too expensive in one or another sense. Such a situation occurs in many applied inverse problems.

**Example 5.3.** An important class of inverse problems arises in the determination of atmospheric parameters, which are necessary to predict e. g., weather and climate. Such quantities are for instance the temperature, the humidity or the concentration of gases such as ozone and carbon dioxide in dependence of the altitude.

In principle many of these values could be measured directly using weather balloons giving high resolution with respect to the altitude and accurate measurements.

An alternative approach is to use indirect measuring techniques. Instead of observing the parameters directly, some derived quantity is measured: For instance the satellite TES[7] observes radiation, which is emitted by the earth's surface at various frequencies; afterwards, by solving a nonlinear integral equation, the concentration of ozone in dependence of the altitude can be recovered. Given such a satellite, a vast number of measurements can be taken. The disadvantage of this technique is of course that a nonlinear and ill-posed problem (see e. g., [1]) must be solved.

Let us now turn to the discussion of the source-conditions (iii) and (iv) in Theorem 4.1. It is clear that there is no possibility to influence the smoothness of the solution (e. g., the distribution of ozone as a function of the altitude). Thus, if (iii) is not satisfied with probability 1, we would have to use a setup as in Theorem 5.2, and choose the exponent $\nu$ in condition (iii) sufficiently small to guarantee that this condition is fulfilled for the problem under consideration. (If this is not possible, still an interpolation approach as in [11, Ch. 4.3] may be applicable.)

In contrast, we do have a possibility to influence condition (iv). As initial guess $x^*$ in the nonlinear problem we will of course use the available data, obtained by the traditional direct method involving the weather balloons. The accuracy of this initial guess depends on

- the distance $\Delta x$ between the region the satellite is observing and the point where the measurement with the balloon was made, and
- the time $\Delta t$ that has elapsed since the balloon has made its observation and in particular also on the weather conditions since then.

In particular we cannot guarantee that $x^*$ at a given point is sufficiently good ($\|x^\dagger - x^*\| \leq c$), but we have only a probability of the form

$$\mathbb{P}\left(\left\|x^\dagger - x^*\right\| > c\right) < \varphi_{\mathrm{cl}}(c, \Delta t, \Delta x)\,,$$

where $\Delta t$ and $\Delta x$ denote the distance in time and space respectively. These two parameters play the role of $\kappa$ in (iv"). By performing more balloon starts at more locations the probability that the distance of $x^*$ and $x^\dagger$ is sufficiently small can be increased, but of course at the same time the cost $\kappa$ of the initial guess rises. As described above, in this example there is a tradeoff between quality of the initial guess and effort for obtaining it.

**Example 5.4.** Not only direct measurements may be used as initial guess. It may even happen that another indirect method yields reliable results, but is replaced by a "cheaper" one:

[7] "Tropospheric Emission Spectrometer", homepage at `http://tes.jpl.nasa.gov/`.

Computerized tomography has been applied for decades in medical imaging, but since the patient is exposed to $X$-ray radiation, there are certain restrictions on how often this reliable diagnosis tool can be used. Alternative methods such as e. g. *SPECT* (see e. g. [17]) are less harming to the patient, but lead to nonlinear problems, with lower resolution. Of course available CT-images will serve as valuable initial guess for resulting reconstructions. The parameter $\kappa$ in (iv") again corresponds to this possibility of improving the initial guess by paying some additional price (e. g., by exposing the patient to $X$-rays more frequently). Increasing the cost $\kappa$, we can increase the probability that the initial guess is sufficiently good.

Observe that the second term in the maximum in (6") does not depend on $\kappa$. This essentially means that $\kappa$ has to increase fast enough, to not diminish the convergence rate $\mathcal{O}(\delta^{2\nu/(2\nu+1)})$. In contrast to e. g., the parameters $\xi$ and $\tau$ in Theorem 4.1, there is no danger of letting $\kappa$ grow too fast.

# Acknowledgments

The authors thank Prof. H.W. Engl for useful and stimulating discussions.

This work has been supported by the Austrian National Science Foundation "Fonds zur Förderung der wissenschaftlichen Forschung" (FWF) under project grant SFB F013 "Numerical and Symbolical Scientific Computing" Project F1208.

# References

[1] G. Bal and K. Ren, *Atmospheric concentration profile reconstructions from radiation measurements*, Inverse Problems, 21 (2005), pp. 153–168.

[2] J. Davidson, *Stochastic limit theory*, Oxford University Press, Oxwford, 1994.

[3] R. M. Dudley, *Real Analysis and Probability*, Wadsworth & Brooks/Cole Mathematics Series. Pacific Grove, CA, 1989.

[4] J. Elstrodt, *Maß- und Integrationstheorie. (Measure and integration theory). 2nd ed.*, Springer, Berlin, 1999.

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

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

[7] H. W. Engl, K. Kunisch, and A. Neubauer, *Convergence rates for Tikhonov regularization of nonlinear ill-posed problems.*, Inverse Problems, 5 (1989), pp. 523–540.

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

[9] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math, 72 (1995), pp. 21–37.

[10] A. Hofinger, *Assessing uncertainty in linear inverse problems with the metrics of Ky Fan and Prokhorov*, tech. rep., SFB F013, Linz, 2006.

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

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

[13] B. Hofmann and O. Scherzer, *Factors influencing the ill-posedness of nonlinear problems.*, Inverse Problems, 10 (1994), pp. 1277–1297.

[14] Q. Jin and Z. Hou, *A parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems.*, Chin. Ann. Math., Ser. A 18 (1997), pp. 483–490.

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

[16] P. Kügler, *A derivative free Landweber iteration for parameter identification in certain elliptic PDEs*, Inverse Problems, 19 (2003), pp. 1407–1426.

[17] R. Ramlau and G. Teschke, *Tikhonov replacement functionals for iteratively solving nonlinear operator equations*, Inverse Problems, 21 (2005), pp. 1571–1592.

[18] O. Scherzer, *Convergence criteria of iterative methods based on Landweber iteration for solving nonlinear problems.*, J. Math. Anal. Appl., 194 (1995), pp. 911–933.

[19] T. I. Seidman and C. R. Vogel, *Well posedness and convergence of some regularization methods for nonlinear ill posed problems.*, Inverse Problems, 5 (1989), pp. 227–238.