Convergence rates for the Bayesian approach to linear inverse problems*

Andreas Hofinger† Hanna K. Pikkarainen†‡

Abstract

Recently, the metrics of Ky Fan and Prokhorov were introduced as a tool for studying convergence in stochastic ill-posed problems. In this work, we show that the Bayesian approach to linear inverse problems can be examined in the new framework as well. We consider the finite-dimensional case where the measurements are disturbed by an additive normal noise and the prior distribution is normal. Convergence and convergence rate results are obtained when the covariance matrices are proportional to the identity matrix.

1 Introduction

We are interested in the linear inverse problem

$$y = Ax \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$ is a known matrix, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. In this work, we consider problems where the matrix A is ill-conditioned. Such problems arise, in particular, when A is a discretized version of a compact operator between infinite-dimensional Hilbert spaces.

Given the exact data y, the *least square minimum norm solution* to problem (1) is defined as

$$x^\dagger := \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \left\{ \|x\| : \|Ax - y\| = \min_{z\in\mathbb{R}^n} \|Az - y\| \right\}.$$

For the linear problem (1) the least square minimum norm solution is

$$x^\dagger = A^\dagger y$$

*This work has been supported by the Austrian National Science Foundation FWF through the project SFB F1308.

†Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria.

‡Email: hanna.pikkarainen@ricam.oeaw.ac.at

where $A^\dagger$ is the Moore–Penrose inverse of A.

Since problem (1) is unstable, observed inexact data y^δ cannot be used directly to infer an approximate solution x^δ to (1), but some regularization technique must be applied. Using, e. g., the well-known method of Tikhonov regularization one obtains an approximate solution x_α^δ to (1). The quantity of interest is the distance between the true solution $x^\dagger$ and the approximate solution x_α^δ.

In the deterministic setup, it turns out that this distance can in general be arbitrarily large (at least in the infinite-dimensional case, cf. [5]); an explicit a-priori bound for the error $\|x^\dagger - x_\alpha^\delta\|$ is therefore not possible. Since a regularization method should naturally be such that less noise leads to better approximations, an accepted quality criterion are convergence rate results in terms of the noisy data, i. e., results of the form

$$\|x^\dagger - x_\alpha^\delta\| = \mathcal{O}(f(\|y - y^\delta\|)).$$

The deterministic theory of inverse problems is well-developed (see, e. g., [5]); but a main criticism is that the above convergence rate result depends on a norm bound of the noise which can be seen as a worst-case scenario for the noise.

In practice, less restrictive stochastic error concepts are more suitable. In the frequentist inversion theory it is assumed that all probabilities appearing in the model of an inverse problem are based on frequencies of random events. Usually, only the noise in the data is modelled by a random variable, but, e. g., in the biological applications also the true solution may have a distribution. In addition, the model itself, i. e., the operator A may be disturbed by a stochastic noise.

In the frequentist approach the regularized solution to an inverse problem is obtained by same regularization techniques as in the deterministic theory. However the regularized solution is now a random variable and hence convergence results must be given in a metric appropriate for random variables.

Often, convergence results in the frequentist framework are given in the terms of the mean square error (cf. [1, 26, 20, 10]). In [6, 12, 11, 14], the Ky Fan metric (a quantitative version of the convergence in probability) and the Prokhorov distance (a quantitative version of the convergence in distribution) were used to deduce convergence results for linear and nonlinear inverse problems. As was demonstrated in [13, 12], there are cases where convergence can be observed in these two metrics while at the same time the mean square error remains constant or diverges.

In addition to the frequentist approach, the Bayesian inversion theory is a widely used tool to tackle stochastic inverse problems. The Bayesian framework has been applied to various inverse problems with success (see, e. g., [17] for an overview). Nonetheless, so far the question of convergence in a sense similar to the deterministic and the frequentist theories has not been studied. In this paper, coupling the metric of Ky Fan with the metric of Prokhorov, we obtain convergence and convergence rate results for the posterior distribution.

The proofs of the main results of this paper are based on lifting point-wise obtained deterministic results to the space of random variables equipped with the Ky Fan metric. Thus this paper can be seen as a step towards the building of a bridge between two—seemingly different—approaches to inverse problems, i.e., between the functional analytic and the statistical inversion theories.

2 Bayesian approach to linear inverse problems

In this section, we summarise the main ideas of the Bayesian inversion theory. A comprehensive introduction into the topic can be found in [17]. We also present a question of convergence related to the Bayesian framework.

The basis of the Bayesian approach to inverse problems is different from the deterministic and the frequentist inversion theories since here all quantities included in the model are treated as random variables. In contrast to the frequentist approach, the probabilities appearing in the Bayesian approach need not correspond to frequencies of random events but they are also used to code the confidence or the degree of belief one has into a particular initial guess.

In the Bayesian approach the solution to an inverse problem is obtained via the Bayes formula. The prior information about the quantities of primary interest is presented in the form of the *prior distribution*. The *likelihood function* is given by the model for the indirect measurements. The solution to the inverse problem after performing the measurements is the *posterior distribution* of the random variables of interest. By the Bayes formula the posterior distribution is proportional to the product of the prior distribution and the likelihood function.

Consequently, in the Bayesian inversion theory not just a single regularized solution to (1) is obtained (as it is done in the deterministic and the frequentist settings) but instead a whole distribution is computed.

We examine the common case where all distributions are assumed to be normal. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X and Y be random variables with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. We suppose that the random variable X is unobservable and of our primary interest and Y is directly observable. We call X the *unknown*, Y the *measurement* and its realization y_{data} in the actual measurement process the *data*. We assume that we have a linear model for the measurements with additive noise

$$Y = AX + E \tag{2}$$

where $A \in \mathbb{R}^{m\times n}$ is a known matrix and $E : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^m$ is a random variable. We suppose that X and E are mutually independent normal random variables with probability densities

$$\pi_{\mathrm{pr}}(x) \propto \exp\left(-\frac{1}{2}(x-x_0)^T\Gamma_{\mathrm{pr}}^{-1}(x-x_0)\right) \tag{3}$$

and

$$\pi_{\mathrm{noise}}(e) \propto \exp\left(-\frac{1}{2}(e-e_0)^T\Gamma_{\mathrm{noise}}^{-1}(e-e_0)\right) \tag{4}$$

where $x_0 \in \mathbb{R}^n$, $e_0 \in \mathbb{R}^m$, and $\Gamma_{\mathrm{pr}} \in \mathbb{R}^{n\times n}$ and $\Gamma_{\mathrm{noise}} \in \mathbb{R}^{m\times m}$ are positive definite symmetric matrices.

For the additive noise model (2), the Bayes theorem yields (independently of the particular structure of π_{pr} and π_{noise}) the posterior distribution

$$\pi_{\mathrm{post}}(x) \propto \pi_{\mathrm{pr}}(x)\pi_{\mathrm{noise}}(y-Ax).$$

For the case of normal random variables, this posterior distribution can be computed explicitly:

Theorem 1. ([17, theorem 3.7]) *Let X and E be independent random variables with probability densities* (3) *and* (4)*, respectively. Assume that the measurement Y satisfies the additive noise model* (2)*. Then the posterior distribution μ_{post} of X conditioned on the data y_{data} is normal and has the probability density*

$$\pi_{\mathrm{post}}(x) \propto \exp\left(-\frac{1}{2}(x-x_{\mathrm{post}})^T\Gamma_{\mathrm{post}}^{-1}(x-x_{\mathrm{post}})\right)$$

where the posterior mean is

$$x_{\mathrm{post}} = \left(\Gamma_{\mathrm{pr}}^{-1} + A^T\Gamma_{\mathrm{noise}}^{-1}A\right)^{-1}\left(A^T\Gamma_{\mathrm{noise}}^{-1}(y_{\mathrm{data}}-e_0) + \Gamma_{\mathrm{pr}}^{-1}x_0\right) \tag{5}$$

and the posterior covariance matrix is

$$\Gamma_{\mathrm{post}} = \left(\Gamma_{\mathrm{pr}}^{-1} + A^T\Gamma_{\mathrm{noise}}^{-1}A\right)^{-1}. \tag{6}$$

For the case where the covariance matrices are proportional to the identity matrix, the theorem above can be simplified as follows.

Remark 2. ([17, example 5]) Suppose that $e_0 = 0$, $\Gamma_{\mathrm{pr}} = \gamma^2 I$, and $\Gamma_{\mathrm{noise}} = \sigma^2 I$ with some $\gamma, \sigma > 0$. Then the posterior distribution of X with the data y_{data} is $\mu_{\mathrm{post}} = \mathcal{N}(x_{\mathrm{post}}, \Gamma_{\mathrm{post}})$ where

$$x_{\mathrm{post}} = \left(A^TA + \frac{\sigma^2}{\gamma^2}I\right)^{-1}A^Ty_{\mathrm{data}} + \frac{\sigma^2}{\gamma^2}\left(A^TA + \frac{\sigma^2}{\gamma^2}I\right)^{-1}x_0 \tag{7}$$

and

$$\Gamma_{\mathrm{post}} = \sigma^2\left(A^TA + \frac{\sigma^2}{\gamma^2}I\right)^{-1}. \tag{8}$$

In the Bayesian approach the posterior distribution μ_{post} defined in theorem 1 is the regularized solution to the linear problem (1) given the data y_{data}. Imitating the deterministic inversion theory, a possible question of convergence in the Bayesian framework is "*Does the posterior distribution μ_{post} converges to the point measure $\delta_{x^\dagger}$ at the least square minimum norm solution $x^\dagger$ when the distribution of the noise E tends to the point measure at the origin?*". A suitable measure of the distance between probability distributions is the Prokhorov metric (see definition 4 below). However, also the stochastic nature of the data must be taken into account.

The data y_{data} is a realization of the random variable Y. Thus the posterior mean x_{post} given by (5) is also a realization of a random variable, namely the random variable

$$X_{\text{post}}(\omega) = \left(\Gamma_{\text{pr}}^{-1} + A^T\Gamma_{\text{noise}}^{-1}A\right)^{-1}\left(A^T\Gamma_{\text{noise}}^{-1}(Y(\omega) - e_0) + \Gamma_{\text{pr}}^{-1}x_0\right).$$

The posterior covariance matrix Γ_{post} is according to definition (6) deterministic. Hence the posterior distribution μ_{post} given in theorem 1 is a realization of the random variable

$$\mu_{\text{post}} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{M}(\mathbb{R}^n), \rho_{\text{P}}), \quad \omega \mapsto \mathcal{N}(X_{\text{post}}(\omega), \Gamma_{\text{post}})$$

where $\mathcal{M}(\mathbb{R}^n)$ is the set of all Borel measures in $\mathbb{R}^n$ and ρ_{P} is the Prokhorov metric in $\mathcal{M}(\mathbb{R}^n)$. Therefore instead of the above question, we could ask "*Does the random variable μ_{post} converges to the constant random variable $\delta_{x^\dagger}$ when the noise E tends to the zero random variable?*". An appropriate measure of the distance between random variables is the Ky Fan metric (see definition 3 below). Convergence and convergence rate results corresponding to the situation of remark 2 are presented in section 4.

In the following section, we revise the definition of the Ky Fan and the Prokhorov metrics and present some related results needed in the proofs of the main theorems of the paper.

3 The metrics of Ky Fan and Prokhorov and multidimensional normal distributions

In the setup of this work, we treat the posterior distribution as a probability measure valued random variable. The set $\mathcal{M}(\mathbb{R}^n)$ of Borel measures in $\mathbb{R}^n$ is only a metric space. We want to quantify the convergence in probability for $\mathcal{M}(\mathbb{R}^n)$-valued random variables. Therefore we utilize the *metric of Ky Fan* to measure distances between random variables on a metric space:

Definition 3 (Ky Fan metric). *Let ξ_1 and ξ_2 be random variables in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a metric space (X, d_{X}). The distance between ξ_1 and ξ_2 in the* Ky Fan metric *is defined as*

$$\rho_{\text{K}}(\xi_1, \xi_2) := \inf\left\{\varepsilon > 0 : \mathbb{P}\left(d_{\text{X}}(\xi_1(\omega), \xi_2(\omega)) > \varepsilon\right) < \varepsilon\right\}.$$

The convergence results we obtain are then formulated in terms of this metric. The Ky Fan metric gives a quantitative version of the convergence in probability; for some background on this metric see [8, 13, 4].

To be able to use the Ky Fan metric for the posterior distribution, we need to equip the set $\mathcal{M}(\mathbb{R}^n)$ with a metric. Here we focus on the *Prokhorov metric*, which is defined as follows:

Definition 4 (Prokhorov metric)**.** *Let μ_1 and μ_2 be Borel measures in a metric space (X, d_{X}). The distance between μ_1 and μ_2 in the* Prokhorov metric *is defined as (see, e. g., [2, 4, 16, 23])*

$$\rho_{\mathrm{P}}(\mu_1,\, \mu_2) := \inf\left\{\varepsilon > 0 : \mu_1(B) \le \mu_2\left(B^{\varepsilon}\right) + \varepsilon \;\forall\, B \in \mathcal{B}(X)\right\}$$

where $\mathcal{B}(X)$ is the Borel σ-algebra in X. The set B^{ε} is the ε-neighbourhood of B, i.e.,

$$B^{\varepsilon} := \left\{x \in X : \inf_{z\in B} d_{\mathrm{X}}(x, z) < \varepsilon\right\}.$$

For some additional background and a comparison with the Ky Fan metric see, e. g., [13].

An important theorem by Strassen [24] (and also Dudley [3]) connects the Prokhorov and the Ky Fan metrics. In this work, we need only the following corollary of Strassen's theorem.

Proposition 5. ([13, proposition 1.5]) *Let (X, d_{X}) be a metric space, $x \in X$, and ξ be a random variable on X with the distribution μ. Then*

$$\rho_{\mathrm{K}}(\xi,\, x) = \rho_{\mathrm{P}}(\mu,\, \delta_x)$$

where δ_x denotes the point measure at x.

The following theorem shows that convergence rates are essentially preserved when they are lifted from a metric space to a space of random variables equipped with the Ky Fan metric.

Theorem 6. *Let ξ_1, ξ_2 and ζ_1, ζ_2 be random variables on metric spaces (X, d_{X}) and (Y, d_{Y}), respectively. Let*

$$d_{\mathrm{X}}(\xi_1(\omega), \xi_2(\omega)) \le \Phi\left(d_{\mathrm{Y}}(\zeta_1(\omega), \zeta_2(\omega))\right) \tag{9}$$

for almost all $\omega \in \Omega$ with some strictly monotonically increasing right-continuous function Φ. Then

$$\rho_{\mathrm{K}}(\xi_1,\, \xi_2) \le \max\left\{\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2),\, \Phi\left(\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2)\right)\right\}. \tag{10}$$

Proof. For an arbitrary $\eta > 0$, due to (9), the monotonicity of Φ, and the definition of the Ky Fan metric,

$$\begin{aligned}
&\mathbb{P}\left(d_{\mathrm{X}}(\xi_1(\omega), \xi_2(\omega)) \leq \Phi\left(\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta\right)\right)\\
&\quad \geq \mathbb{P}\left(\Phi\left(d_{\mathrm{Y}}(\zeta_1(\omega), \zeta_2(\omega))\right) \leq \Phi\left(\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta\right)\right)\\
&\quad = \mathbb{P}\left(d_{\mathrm{Y}}(\zeta_1(\omega), \zeta_2(\omega)) \leq \rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta\right)\\
&\quad > 1 - \left(\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta\right).
\end{aligned}$$

Hence by the definition of the Ky Fan distance,

$$\rho_{\mathrm{K}}(\xi_1,\, \xi_2) \leq \max\{\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta,\ \Phi\left(\rho_{\mathrm{K}}(\zeta_1,\, \zeta_2) + \eta\right)\}.$$

Since Φ is right-continuous, inequality (10) is attained by letting $\eta \to 0$. $\square$

Similar lifting results have been obtained in the Prokhorov and the Ky Fan metrics when $\xi_1 = F(\zeta_1)$ and $\xi_2 = F(\zeta_2)$ with a Lipschitz, a locally Lipschitz and a Hölder continuous function F in [27], [7] and [6, 12, 11], respectively.

In this paper, we examine the case in which all distributions are normal. Let $y_0 \in \mathbb{R}^m$ and $\Sigma \in \mathbb{R}^{m\times m}$ be a positive definite symmetric matrix. A normal m-dimensional random variable with mean y_0 and covariance matrix Σ is a random variable in $\mathbb{R}^m$ whose distribution is absolutely continuous with respect to the m-dimensional Lebesgue measure and has the probability density

$$\pi(y) = \left(\frac{1}{(2\pi)^m|\Sigma|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y-y_0)^T\Sigma^{-1}(y-y_0)\right)$$

where $|\cdot|$ is the determinant of matrices. The corresponding distribution is denoted by $\mathcal{N}(y_0, \Sigma)$.

In the following lemma, we give an upper bound for the Ky Fan distance between a normal random variable and its mean when its covariance matrix is sufficiently small. The proof is given in appendix A.

Lemma 7. *Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(y_0, \Sigma)$. Let us define $\kappa(m) := \max\{1,\, m-2\}$ and $C(m)$ to be*

$$C(m) := \begin{cases} \frac{2\pi}{(m+1)^2}, & \text{if } m \text{ is odd,}\\ \frac{2^m}{m^2}, & \text{if } m \text{ is even.}\end{cases}$$

Then there exists a positive constant $\theta(m)$ such that

$$\rho_{\mathrm{K}}(\xi,\, y_0) \leq \sqrt{-\|\Sigma\| \log\left(C(m)\|\Sigma\|^{\kappa(m)}\right)}$$

for all Σ with $\|\Sigma\| < \theta(m)$.

According to proposition 5, an upper bound for the Prokhorov distance between a normal distribution and the point measure at its mean is also given by lemma 7 when the covariance matrix of the distribution is small enough.

In the following corollary, we present some important special cases of lemma 7, which are needed in this paper.

Corollary 8. *Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(y_0, \Sigma)$. Let $C(m)$ and $\kappa(m)$ be as in lemma 7 and $I \in \mathbb{R}^{m\times m}$ be the identity matrix.*

(i) If $\Sigma = \sigma^2 I$ for some $\sigma > 0$, there exists a positive constant $\sigma(m)$ such that

$$\rho_{\mathrm{K}}(\xi,\, y_0) \le \sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}$$

for all $\sigma < \sigma(m)$.

(ii) If $\Sigma = \sigma^2(B + \frac{\sigma^2}{\gamma^2}I)^{-1}$ where $B \in \mathbb{R}^{m\times m}$ is a symmetric positive semidefinite matrix and $\gamma, \sigma > 0$, there exist positive constants $\gamma(m)$ and $\sigma(m)$ such that

$$\rho_{\mathrm{K}}(\xi,\, y_0) \le \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda_{\mathrm{MIN}} + \sigma^2}}\sqrt{-\log\left(C(m)\frac{\gamma^{2\kappa(m)}\sigma^{2\kappa(m)}}{(\gamma^2\lambda_{\mathrm{MIN}} + \sigma^2)^{\kappa(m)}}\right)}$$

for all $\gamma < \gamma(m)$ and $\sigma < \sigma(m)$ where λ_{MIN} is the minimal eigenvalue of B.

Proof. The first claim is obvious. In case (ii)

$$\|\Sigma\| = \frac{\gamma^2\sigma^2}{\gamma^2\lambda_{\mathrm{MIN}} + \sigma^2}.$$

If $\lambda_{\mathrm{MIN}} = 0$, the norm of Σ is equal to γ^2. Hence the claim follows from lemma 7 when $\gamma(m)$ is chosen to be $\sqrt{\theta(m)}$ and $\sigma(m)$ is arbitrary.

If $\lambda_{\mathrm{MIN}} \neq 0$, the norm of Σ is less than $\theta(m)$ only if γ and σ satisfy the inequality

$$\lambda_{\mathrm{MIN}}\theta(m)\gamma^2 + \theta(m)\sigma^2 - \gamma^2\sigma^2 > 0.$$

Therefore an upper bound for either γ or σ can be chosen freely, but then the upper bound for the second parameter is given by the equation $\lambda_{\mathrm{MIN}}\theta(m)\gamma^2 + \theta(m)\sigma^2 - \gamma^2\sigma^2 = 0$. If either γ or σ is selected to be small enough, i.e., $\gamma < \sqrt{\theta(m)}$ or $\sigma < \sqrt{\lambda_{\mathrm{MIN}}\theta(m)}$, the second parameter can be arbitrary. By an appropriate choice of upper bounds for γ and/or σ the claim is a consequence of lemma 7. □

4 Convergence rates for the Bayesian approach

In this section, we investigate convergence and convergence rate results for the Bayesian approach to linear inverse problems. In particular, we answer the question of convergence posed in section 2 for the situation of remark 2.

We assume that the prior distribution of the unknown X is $\mathcal{N}(x_0, \gamma^2 I)$ and the noise E has the distribution $\mathcal{N}(0, \sigma^2 I)$ with some $\gamma, \sigma > 0$ and $x_0 \in \mathbb{R}^n$ (like in remark 2). As we have seen in theorem 1, the posterior distribution with the data y_{data} is given as

$$\mu_{\text{post}} = \mathcal{N}(x_{\text{post}},\, \Gamma_{\text{post}})$$

with x_{post} and Γ_{post} defined in (7) and (8), respectively.

As noticed before, the mean of the posterior distribution is actually a realization of the random variable X_{post} while the covariance matrix is constant. To deduce a full convergence result for the Bayesian approach, we first of all bound the error $\rho_{\text{P}}(\mathcal{N}(X_{\text{post}}(\omega),\, \Gamma_{\text{post}}),\, \delta_{x^\dagger})$ given a concrete realization $y_{\text{data}} = Y(\omega)$ of the measurement.

In the following proposition, we give an upper bound for the Prokhorov distance between the posterior distribution and the point measure at the least square minimum norm solution for a particular data.

Proposition 9. *Let the assumptions of theorem 1 and remark 2 be satisfied. Then the distance between the posterior distribution and the point measure at the solution $x^\dagger$ is bounded in the Prokhorov metric by*

$$\begin{aligned}
&\rho_{\text{P}}(\mathcal{N}(X_{\text{post}}(\omega), \Gamma_{\text{post}}),\, \delta_{x^\dagger}) \\
&\quad \le \frac{\sigma^2 \left\| x^\dagger - (I-P)x_0 \right\|}{\gamma^2 \lambda_p^2 + \sigma^2} + \rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) + \frac{\gamma}{2\sigma} \|Y(\omega) - y\| + \|P x_0\|
\end{aligned} \tag{11}$$

for all $\omega \in \Omega$ where λ_p is the minimal positive singular value of A and P is the orthogonal projection onto $\mathcal{N}(A)$.

Let λ_{MIN} denote the minimal eigenvalue of $A^T A$. Then there exist positive constants $\gamma(n)$ and $\sigma(n)$ such that

$$\begin{aligned}
&\rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) \\
&\quad \le \frac{\gamma\sigma}{\sqrt{\gamma^2 \lambda_{\text{MIN}} + \sigma^2}} \sqrt{-\log\left(C(n) \frac{\gamma^{2\kappa(n)} \sigma^{2\kappa(n)}}{(\gamma^2 \lambda_{\text{MIN}} + \sigma^2)^{\kappa(n)}} \right)}
\end{aligned} \tag{12}$$

for all $\gamma < \gamma(n)$ and $\sigma < \sigma(n)$ where the constants $C(n)$ and $\kappa(n)$ are given in lemma 7.

Proof. By the triangle inequality of the Prokhorov metric (e. g., [16]), and proposition 5,

$$\rho_{\text{P}}(\mathcal{N}(X_{\text{post}}(\omega), \Gamma_{\text{post}}),\, \delta_{x^\dagger}) \le \rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) + \|X_{\text{post}}(\omega) - x^\dagger\|$$

for all $\omega \in \Omega$. We may rewrite

$$X_{\text{post}}(\omega) = \left(A^T A + \frac{\sigma^2}{\gamma^2} I\right)^{-1} A^T(Y(\omega) - Ax_0) + x_0.$$

Thus $X_{\text{post}}(\omega) - x_0$ is the Tikhonov regularized solution to the linear inverse problem

$$Az = y - Ax_0 \tag{13}$$

with the regularization parameter σ^2/γ^2 and the noisy data $Y(\omega) - Ax_0$. The least square minimum norm solution to (13) is

$$z^\dagger = A^\dagger y - A^\dagger A x_0 = x^\dagger - (I - P)x_0.$$

By using the singular value decomposition of the matrix A we can estimate

$$\|X_{\text{post}}(\omega) - x^\dagger\| \leq \frac{\sigma^2}{\gamma^2 \lambda_p^2 + \sigma^2} \|x^\dagger - (I - P)x_0\| + \frac{\gamma}{2\sigma}\|Y(\omega) - y\| + \|Px_0\|$$

for all $\omega \in \Omega$. Bound (12) is a consequence of proposition 5 and item (ii) in corollary 8. □

Remark 10. Since the representation for μ_{post} given in theorem 1 is only valid when the noise E is normal, proposition 9 gives no immediate bound on the error $\rho_{\text{P}}(\mathcal{N}(X_{\text{post}}(\omega), \Gamma_{\text{post}}), \delta_{x^\dagger})$ for all $\omega \in \Omega$. In principle, $\|Y(\omega) - y\|$ can be arbitrarily large and hence the right hand side of (11) is in general unbounded.

For any positive definite symmetric matrix $\Sigma \in \mathbb{R}^{n \times n}$ the function $x \mapsto \mathcal{N}(x, \Sigma)$ is continuous from $\mathbb{R}^n$ to $\mathcal{M}(\mathbb{R}^n)$ equipped with the Prokhorov metric. Therefore, to turn (11) into an actual bound on the error, we may consider the posterior distribution as a random variable,

$$\mu_{\text{post}}(\omega) := \mathcal{N}(X_{\text{post}}(\omega), \Gamma_{\text{post}}),$$

i. e., as a measurable function from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{M}(\mathbb{R}^n), \rho_{\text{P}})$. Since μ_{post} is a random variable on a metric space, we can compute the error between the posterior distribution and the constant random variable $\delta_{x^\dagger}$ in the Ky Fan metric.

Theorem 11. *Let the assumptions of theorem 1 and remark 2 be satisfied. Then the distance between the posterior distribution and the constant random variable $\delta_{x^\dagger}$ is bounded in the Ky Fan metric by*

$$\begin{aligned}\rho_{\text{K}}(\mu_{\text{post}}, \delta_{x^\dagger}) \leq \max \Bigg\{ & \rho_{\text{K}}(Y, y), \\ & \frac{\sigma^2 \|x^\dagger - (I - P)x_0\|}{\gamma^2 \lambda_p^2 + \sigma^2} + \rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}), \delta_0) + \frac{\gamma}{2\sigma}\rho_{\text{K}}(Y, y) + \|Px_0\| \Bigg\}\end{aligned}$$

where λ_p is the minimal positive singular value of A and P is the orthogonal projection onto $\mathcal{N}(A)$.

Proof. The proof follows by combining proposition 9 and theorem 6. The right hand side of (11) plays the role of the function Φ in theorem 6 and the metric d_{x} is given by the Prokhorov metric. $\square$

We are interested in the case when the noise E tends to zero in the Ky Fan metric, i. e., the distribution of the noise tends to the point distribution at the origin in the Prokhorov metric.

Lemma 12. *Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(0, \sigma^2 I)$ for some $\sigma > 0$. Then*

$$\rho_{\mathrm{K}}(\xi, 0) \to 0 \quad \Longleftrightarrow \quad \sigma \to 0.$$

Proof. According to the proof of lemma 7 the Ky Fan distance $\rho_{\mathrm{K}}(\xi, 0)$ is equal to the unique positive solution to the equation

$$z = \frac{2}{\Gamma(m/2)} I_{m-1}\left(\frac{z}{\sqrt{2}\sigma}\right) \tag{14}$$

where Γ is the gamma function and the function I_{m-1} is

$$I_{m-1}(x) = \int_x^\infty t^m e^{-t^2} \mathrm{d}t$$

for all $x \geq 0$. The explicit formula of the function I_{m-1} is given in appendix A. The function I_{m-1} is strictly monotonically decreasing and continuous. Furthermore, $I_{m-1}(x) \to 0$ as $x \to \infty$. Hence the claim follows from equation (14). $\square$

Therefore a proper question of convergence is if the error $\rho_{\mathrm{K}}(\mu_{\mathrm{post}}, \delta_{x^\dagger})$ tends to zero as $\sigma \to 0$.

We can use the result of theorem 11 and deduce parameter choice rules for the Bayesian approach to obtain convergence and convergence rates for the posterior distribution. The following theorems are the main results of this work and give the positive answer to the question of section 2.

Theorem 13. *Let the assumptions of theorem 1 and remark 2 be valid and $x_0 \in \mathcal{N}(A)^\perp$. Let $\gamma(\sigma)$ satisfy*

$$\frac{\sigma}{\gamma(\sigma)} \longrightarrow 0 \quad \text{and} \quad \gamma(\sigma)\sqrt{-\log(C(m)\sigma^{2\kappa(m)})} \longrightarrow 0 \tag{15}$$

as $\sigma \to 0$ where the constants $C(m)$ and $\kappa(m)$ are given in lemma 7. Then

$$\rho_{\mathrm{K}}\left(\mu_{\mathrm{post}}, \delta_{x^\dagger}\right) \longrightarrow 0$$

as $\sigma \to 0$.

Proof. Since the distribution of Y is $\mathcal{N}(y, \sigma^2 I)$, according to corollary 8 there exists a positive constant $\sigma(m)$ such that

$$\rho_{\mathrm{K}}(Y,\, y) \leq \sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}$$

for all $\sigma < \sigma(m)$. By combining the results of theorem 11 and proposition 9 there exist positive constants $\gamma(n)$ and $\sigma(m,n)$ such that

$$\begin{aligned}\rho_{\mathrm{K}}\left(\mu_{\mathrm{post}},\, \delta_{x^\dagger}\right) \leq \frac{\sigma^2\left\|x^\dagger - x_0\right\|}{\gamma^2\lambda_p^2 + \sigma^2} + \max\left\{\frac{\gamma}{2}, \sigma\right\}\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)} \\ + \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda_{\mathrm{MIN}} + \sigma^2}}\sqrt{-\log\left(C(n)\frac{\gamma^{2\kappa(n)}\sigma^{2\kappa(n)}}{(\gamma^2\lambda_{\mathrm{MIN}} + \sigma^2)^{\kappa(n)}}\right)}\end{aligned} \tag{16}$$

for all $\gamma < \gamma(n)$ and $\sigma < \sigma(m,n)$ where λ_p is the minimal positive singular value of A and λ_{MIN} is the minimal eigenvalue of $A^T A$.

The second term on the right hand side of (16) tends to zero when $\gamma\sqrt{-\log(C(m)\sigma^{2\kappa(m)})} \to 0$ as $\sigma \to 0$. In the first term it is required that $\gamma/\sigma \to \infty$ as $\sigma \to 0$. If $\lambda_{\mathrm{MIN}} = 0$, the third term tends to zero as $\sigma \to 0$ under the above assumptions. When $\lambda_{\mathrm{MIN}} \neq 0$, it is enough if $\gamma\sqrt{-\log(C(n)\sigma^{2\kappa(n)})} \to 0$ and $\gamma/\sigma \geq 1$ as $\sigma \to 0$. If $\lambda_{\mathrm{MIN}} \neq 0$, then $n \leq m$. Thus the parameter choice (15) guarantees the convergence. □

For example, $\gamma(\sigma) \sim \sigma^\eta$ with some $0 < \eta < 1$ fulfills the requirements of theorem 13.

Remark 14. The assumption $x_0 \in \mathcal{N}(A)^\perp$ in theorem 13 is necessary. The least square minimum norm solution belongs to the orthocomplement of the null space of the matrix A and $PX_{\mathrm{post}}(\omega) = Px_0$ for all $\omega \in \Omega$ where P is the orthogonal projection onto $\mathcal{N}(A)$. Hence the convergence of the posterior distribution to the point measure at the least square minimum norm solution $x^\dagger$ is possible only when $Px_0 = 0$. If $Px_0 \neq 0$ and $\gamma(\sigma)$ satisfies the parameter choice (15), the posterior distribution converges to the point measure at the least square solution $x^\dagger + Px_0$.

Besides this convergence result, also convergence *rate* results can be obtained.

Theorem 15. *Let the assumptions of theorem 1 and remark 2 be satisfied and $x_0 \in \mathcal{N}(A)^\perp$. Let γ be chosen as*

$$\gamma \sim \left(\sigma^2/\sqrt{-\log(C(m,n)\sigma^{2\kappa(m,n)}}\right)^{\frac{1}{3}} \tag{17}$$

where the constants $C(m,n) := C(\max(m,n))$ and $\kappa(m,n) := \kappa(\max(m,n))$ are defined in lemma 7. Then

$$\rho_{\mathrm{K}}\left(\mu_{\mathrm{post}},\, \delta_{x^\dagger}\right) \leq \mathcal{O}\left(\left(\sigma\sqrt{-\log\left(C(m,n)\sigma^{2\kappa(m,n)}\right)}\right)^{\frac{2}{3}}\right). \tag{18}$$

Proof. As in the proof of theorem 13 we have the upper bound (16) for the distance between the posterior distribution and the point measure at the true solution for small enough γ and σ. To obtain a convergence rate for the error the right hand side of (16) should be minimized for a fixed σ. Since the minimizing $\gamma(\sigma)$ is not easy to derive, we estimate the right hand side of (16) from above. When γ and σ are small enough and $\gamma/\sigma \geq 1$,

$$\rho_{\mathrm{K}}\left(\mu_{\mathrm{post}},\, \delta_{x^\dagger}\right) \leq \frac{\sigma^2}{\gamma^2\lambda_p^2}\left\|x^\dagger - x_0\right\| + \frac{5\gamma}{2}\sqrt{-\log\left(C(m,n)\sigma^{2\kappa(m,n)}\right)}$$

where λ_p is the minimal positive singular value of the matrix A. By choice (17) the two terms on the right-hand side are balanced and hence rate (18) is obtained. $\square$

In theorem 13 as well as 15 we show that the parameter γ must tend to 0 in order to obtain the convergence of μ_{post} to $\delta_{x^\dagger}$. This condition on γ is counter-intuitive compared to the common notion of the Bayesian approach, where $\gamma = 0$ essentially implies that the mean of the prior distribution should be taken as a true solution. To explain this discrepancy, it should be noted that, compared with the variance of the noise, the variance of the prior distribution does tend to the infinity ($\gamma/\sigma \to \infty$), i. e., the prior distribution becomes non-informative. This is also visible from the fact that the mean of the prior distribution need not to converge to the desired solution but it may just remain constant.

The exponent κ depends on the dimension of the measurement. Hence the convergence rate also depends on the dimension unlike in the deterministic regularization theory. Nonetheless, the dimension-dependence appears only in the logarithmic factor, i. e., it diminishes the rate when σ is large, but the influence becomes smaller as $\sigma \to 0$.

While the rate of convergence in theorem 15 is independent of the matrix A, the constant in the convergence rate (18) depends on A. If A is ill-conditioned and hence the minimal positive singular value λ_p of A is tiny, the constant in (18) is huge. Therefore convergence rate results independent of the matrix A are of interest.

In the deterministic inversion theory, explicit results on convergence rates for infinite-dimensional linear inverse problems are only possible when additional assumptions on the solution are imposed (cf. [5, proposition 3.11]). These assumptions can, for instance, be formulated in terms of abstract smoothness conditions, so called *source conditions*.

Definition 16. *The least square minimum norm solution $x^\dagger$ satisfies a* deterministic source condition *with source function f if there exist $v \in \mathbb{R}^n$ and $\tau > 0$ such that*

$$x^\dagger = f(A^T A)v \quad \textit{and} \quad \|v\| \leq \tau. \tag{19}$$

Typical choices of the source function are $f(\lambda) = \lambda^\nu$, $\nu \le 1$ and $f(\lambda) = (-\log \lambda)^{-\nu}$ (see [5, 15]). Source conditions are needed to bound the approximation error:

Definition 17. *The source function f allows the* deterministic convergence rate h *if there exists an increasing function h such that $h(0) = 0$ and*

$$x^\dagger \in \{z \in \mathbb{R}^n : z = f(A^T A)\, v,\ \|v\| \le \tau\} \implies \|x^\dagger - x_\alpha\| \le \tau h(\alpha)$$

for any $A \in \mathbb{R}^{m\times n}$ and $\tau > 0$ where $x_\alpha := (A^T A + \alpha I)^{-1} A^T A x^\dagger$.

For the Hölder and the logarithmic source functions f above, it has been shown that $f = h$ (see [5] and [15], respectively). Nevertheless, this is not the case in general, e. g., when saturation occurs (cf., e. g., [5]). For some general results on connections between f and h, using weak assumptions only (e. g., monotonicity or concavity of f) we refer to [22, 25].

Theorem 18. *Let the assumptions of theorem 1 and remark 2 be valid and $x_0 \in \mathcal{N}(A)^\perp$. Let $x^\dagger - x_0$ satisfy the deterministic source condition* (19) *with the source function $f(\lambda) = \lambda^\nu$ for some $0 \le \nu \le 1$. Furthermore, let γ be chosen as*

$$\gamma \sim \left(\sigma^{2\nu} / \sqrt{-\log(C(m,n)\sigma^{2\kappa(m,n)}}\right)^{\frac{1}{2\nu+1}} \tag{20}$$

where the constants $C(m,n) := C(\max(m,n))$ and $\kappa(m,n) := \kappa(\max(m,n))$ are defined in lemma 7. Then

$$\rho_{\kappa}\left(\mu_{\text{post}},\, \delta_{x^\dagger}\right) \le \mathcal{O}\left(\left(\sigma\sqrt{-\log\left(C(m,n)\sigma^{2\kappa(m,n)}\right)}\right)^{\frac{2\nu}{2\nu+1}}\right). \tag{21}$$

Proof. The source function $f(\lambda) = \lambda^\nu$ allows the deterministic convergence rate $h(\lambda) = \lambda^\nu$ [5, (5.18)]. The source condition is used to bound the approximation error in the proof of proposition 9 instead of the singular value decomposition of the matrix A. Hence by lifting a result similar to theorem 11 can be obtained. By combining that result with proposition 9 and corollary 8 there exist positive constants $\gamma(n)$ and $\sigma(m,n)$ such that

$$\begin{aligned}\rho_{\kappa}\left(\mu_{\text{post}},\, \delta_{x^\dagger}\right) \le{}& \tau\left(\frac{\sigma}{\gamma}\right)^{2\nu} + \max\left\{\frac{\gamma}{2}, \sigma\right\}\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)} \\ &+ \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda_{\text{MIN}} + \sigma^2}}\sqrt{-\log\left(C(n)\frac{\gamma^{2\kappa(n)}\sigma^{2\kappa(n)}}{(\gamma^2\lambda_{\text{MIN}} + \sigma^2)^{\kappa(n)}}\right)}\end{aligned}$$

for all $\gamma < \gamma(n)$ and $\sigma < \sigma(m,n)$ where λ_p is the minimal positive singular value of A and λ_{MIN} is the minimal eigenvalue of $A^T A$. Similarly as in the proof of theorem 15 when γ and σ are small enough and $\gamma/\sigma \ge 1$,

$$\rho_{\kappa}\left(\mu_{\text{post}},\, \delta_{x^\dagger}\right) \le \tau\left(\frac{\sigma}{\gamma}\right)^{2\nu} + \frac{5\gamma}{2}\sqrt{-\log\left(C(m,n)\sigma^{2\kappa(m,n)}\right)}.$$

By choice (20) the two terms on the right-hand side are balanced and hence rate (21) is obtained. □

5 Conclusions

In this paper, we have examined convergence and convergence rate results for the Bayesian inversion theory. As a tool for quantifying convergence of the posterior distribution, a coupling of the Ky Fan and the Prokhorov metrics appeared to be the natural choice. As far as we know, the paper contains the first convergence result for the Bayesian approach to inverse problems.

In this work, we considered only normal distributions. While this is an accepted choice for the noise, for the prior information in the Bayesian framework also different options are in use (see [17, chapter 3] for an overview). Alternative prior distributions may lead to better reconstructions, but it is often not possible to deduce explicit solutions, either posterior distributions or point estimates. Instead, Markov chain Monte Carlo (MCMC) methods are used to determine approximate estimates. Convergence results similar to the ones presented in this paper are not straight-forward to achieve for arbitrary prior distributions.

Furthermore, this work is based on the assumption that the model of the inverse problem is finite-dimensional. The received results are dimension-dependent in a way that prevents generalization to the infinite-dimensional case. The dimension-dependence can be attributed to the fact that in an infinite-dimensional Hilbert space the noise E in the additive noise model (2) (and consequently Y as well) cannot as a Gaussian random variable belong to the underlying Hilbert space since the covariance of E is not a trace-class operator (cf. [18, theorem 2.3]).

The Bayesian inversion theory in infinite-dimensional spaces is not completely developed. The existence of the regular version of conditional distributions is well-known in Polish spaces, i. e., in complete separable metric spaces [9, theorem I.3.3], in the dual space of nuclear countable Hilbert spaces [19, theorem 3.2], and in the space of distributions in an open domain of $\mathbb{R}^n$ [19, theorem 5.2]. For Gaussian linear inverse problems in an infinite-dimensional Hilbert space, in the dual space of a nuclear countable Hilbert space, and in the space of distributions the form of the posterior distribution is presented in [21] and [19], respectively. Convergence results for the infinite-dimensional Bayesian inversion theory require more sophisticated stochastic analysis than used in this paper.

So far, the functional analytic and the statistical approach to inverse problems have mainly been studied by separate communities. In this paper, we have deduced convergence results for the Bayesian framework by using well-known results of the deterministic theory and by lifting them to the space of random variables. We hope that this paper will provide a step towards the building of a bridge between the deterministic and the statistical inversion theories.

Acknowledgements

The authors are grateful to Professor Heinz W. Engl, Johannes Kepler Universität Linz and Johann Radon Institute for Computational and Applied Mathematics, for the suggestion to study the subject and for valuable discussions.

This work has been supported by the Austrian National Science Foundation FWF through the project SFB F1308.

A Proof of lemma 7

Proof of lemma 7. Let ζ be an $\mathbb{R}^m$-valued random variable with distribution $\mathcal{N}(0,\Sigma)$. We need to estimate the Ky Fan distance

$$\rho_{\mathrm{K}}(\xi,\, y_0) = \rho_{\mathrm{K}}(\zeta,\, 0) = \inf\{\varepsilon > 0 : \mathbb{P}(\|\zeta(\omega)\| > \varepsilon) < \varepsilon\}.$$

Let $z > 0$. For the normal random variable ζ, the probability of realizations with large norms is

$$\mathbb{P}(\|\zeta(\omega)\| > z) = \left(\frac{1}{(2\pi)^m|\Sigma|}\right)^{\frac{1}{2}} \int_{\|x\|>z} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right) \mathrm{d}x$$

where $|\cdot|$ denotes the determinant of matrices. Transformation to spherical coordinates leads to

$$\mathbb{P}(\|\zeta(\omega)\| > z) \leq \frac{2}{\Gamma(m/2)} \int_{\frac{z}{\sqrt{2\|\Sigma\|}}}^{\infty} t^{m-1}e^{-t^2}\mathrm{d}t$$

where Γ is the gamma function. If $\Sigma = \sigma^2 I$ with some $\sigma > 0$, the equality holds. We denote

$$I_m(x) = \int_x^{\infty} t^m e^{-t^2}\mathrm{d}t$$

for all $m \in \mathbb{N}_0$ and $x \geq 0$. With this notation,

$$\rho_{\mathrm{K}}(\xi,\, y_0) \leq \inf\left\{ z > 0 : \frac{2}{\Gamma(m/2)}\, I_{m-1}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right) < z \right\}.$$

There exists a unique point $z(m, \|\Sigma\|)$ such that

$$\frac{2}{\Gamma(m/2)}\, I_{m-1}\left(\frac{z(m,\|\Sigma\|)}{\sqrt{2\|\Sigma\|}}\right) = z(m,\|\Sigma\|) \tag{22}$$

and

$$\frac{2}{\Gamma(m/2)}\, I_{m-1}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right) < z \quad \text{for all } z > z(m, \|\Sigma\|),$$
$$\frac{2}{\Gamma(m/2)}\, I_{m-1}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right) > z \quad \text{for all } 0 < z < z(m, \|\Sigma\|).$$

Thus $\rho_{\kappa}(\xi, y_0) \leq z(m, \|\Sigma\|)$. To derive bounds on $\rho_{\kappa}(\xi, y_0)$ it therefore suffices to deduce estimates on $z(m, \|\Sigma\|)$.

By induction, we obtain the following recursive representation for the integrals I_m

$$\begin{cases} I_0(x) = \frac{\sqrt{\pi}}{2}\operatorname{erfc}(x), \\ I_1(x) = \frac{1}{2}e^{-x^2}, \\ I_m(x) = \frac{1}{2}x^{m-1}e^{-x^2} + \frac{m-1}{2}I_{m-2}(x), \quad m \geq 2 \end{cases}$$

where the *complementary error function* $\operatorname{erfc}(\cdot)$ is defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\mathrm{d}t$$

for $x \in \mathbb{R}$. The explicit formulae for the functions I_m, $m \geq 2$, are

$$I_m(x) = \begin{cases} \frac{1}{2}e^{-x^2}\sum_{i=0}^{\frac{m-1}{2}} \frac{\left(\frac{m-1}{2}\right)!}{\left(\frac{m-1}{2}-i\right)!}x^{m-2i-1} \\ \frac{1}{2}e^{-x^2}\sum_{i=0}^{\frac{m}{2}-1} \frac{1}{2^i}\frac{(m-1)!!}{(m-2i-1)!!}x^{m-2i-1} + \frac{\sqrt{\pi}}{2^{m/2+1}}(m-1)!!\operatorname{erfc}(x) \end{cases}$$

for an odd and even m, respectively, and for all $x \geq 0$ where

$$(2l+1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2l+1)$$

for all $l \in \mathbb{N}_0$.

When exploring equation (22) we distinguish the cases (i) $m = 1$, (ii) $m = 2$, (iii) $m > 2$ is even and (iv) $m > 2$ is odd. Notice that the solution $z(m, \|\Sigma\|)$ to equation (22) satisfies $z(m, \|\Sigma\|) \to 0$ and $z(m, \|\Sigma\|)/\sqrt{\|\Sigma\|} \to \infty$ as $\|\Sigma\| \to 0$ in all four cases.

(i) When $m = 1$, the matrix Σ is a positive number σ^2. Hence, we are interested in the equation

$$z = \operatorname{erfc}\left(\frac{z}{\sqrt{2}\sigma}\right).$$

Since $\operatorname{erfc}(x) \leq \exp(-x^2)/x\sqrt{\pi}$ for all $x > 0$, the solution $\tilde{z}(1, \sigma)$ to the equation

$$z = \frac{\sigma}{z}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{z^2}{2\sigma^2}\right) \tag{23}$$

satisfies $z(1,\sigma^2) \leq \tilde{z}(1,\sigma)$ for all $\sigma \geq 0$. We make the ansatz $\tilde{z}(1,\sigma) = \sigma A(1,\sigma)$ and take the logarithm of both sides of (23). Then

$$\log\left(\sqrt{\frac{\pi}{2}}\sigma\right) + 2\log A(1,\sigma) = -\frac{1}{2}A(1,\sigma)^2.$$

For the ansatz we need to have $A(1,\sigma) \to \infty$ as $\sigma \to 0$. Therefore there exists a positive constant σ_1 such that $A(1,\sigma) \geq 1$ for all $\sigma < \sigma_1$. Hence

$$A(1,\sigma) \leq \sqrt{-\log\left(\frac{\pi}{2}\sigma^2\right)}$$

for all $\sigma < \sigma_1$ and thus

$$\rho_{\mathrm{K}}(\xi,\, y_0) \leq \sigma\sqrt{-\log\left(\frac{\pi}{2}\sigma^2\right)}$$

for all $\sigma < \min\left\{\sigma_1, \sqrt{2/\pi}\right\}$. This result resembles theorem 6.9 in [12], but while there the result was only stated in an asymptotic form, here we gave a more explicit relation for the region of validity.

(ii) When $m = 2$, taking the logarithm from both sides of equation (22) yields

$$\log z = -\frac{z^2}{2\|\Sigma\|}.$$

Since $z(m, \|\Sigma\|)/\sqrt{\|\Sigma\|} \to \infty$ as $\|\Sigma\| \to 0$, there exists a positive constant σ_2 such that $z(m, \|\Sigma\|) \geq \sqrt{\|\Sigma\|}$ when $\|\Sigma\| < \sigma_2$. Therefore

$$\log\left(\sqrt{\|\Sigma\|}\right) \leq -\frac{z^2}{2\|\Sigma\|}$$

when $\|\Sigma\| < \sigma_2$. Hence

$$\rho_{\mathrm{K}}(\xi,\, y_0) \leq \sqrt{-\|\Sigma\|\log\left(\|\Sigma\|\right)}$$

for all Σ such that $\|\Sigma\| < \min\{\sigma_2, 1\}$.

(iii) When $m > 2$ is even, i.e., $m = 2l$ for some $l \geq 2$, equation (22) is of the form

$$z = \exp\left(-\frac{z^2}{2\|\Sigma\|}\right)\sum_{i=0}^{l-1}\frac{1}{(l-i-1)!}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right)^{2l-2i-2}.$$

We notice that

$$e^{-x^2}\sum_{i=0}^{l-1}\frac{x^{2l-2i-2}}{(l-i-1)!} \leq l x^{2l-2}e^{-x^2}$$

when $x \geq 1$ and $x^{2l-2} \leq x$ when $x \leq 1$. Since $z(2l, \|\Sigma\|)/\sqrt{\|\Sigma\|} \to \infty$ and $z(2l, \|\Sigma\|) \to 0$ as $\|\Sigma\| \to 0$, there exists a positive constant σ_{2l} such that

$\sqrt{2\|\Sigma\|} \le z(2l, \|\Sigma\|) \le 1$ when $\|\Sigma\| < \sigma_{2l}$. Therefore the solution $\tilde{z}(2l, \|\Sigma\|)$ to the equation

$$\frac{2^{l-1}}{l}\|\Sigma\|^{l-1} = \exp\left(-\frac{z^2}{2\|\Sigma\|}\right)$$

is greater than $z(2l, \|\Sigma\|)$ when $\|\Sigma\| < \sigma_{2l}$. Hence

$$\rho_\kappa(\xi,\, y_0) \le \sqrt{-\|\Sigma\| \log\left(\frac{2^{2l-2}}{l^2}\|\Sigma\|^{2l-2}\right)}$$

for all Σ such that $\|\Sigma\| < \min\{\sigma_{2l}, l^{1/(l-1)}/2\}$.

(iv) When $m > 2$ is odd, i.e., $m = 2l+1$ for some $l \ge 1$, equation (22) is equal to

$$z = \operatorname{erfc}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right) + \exp\left(-\frac{z^2}{2\|\Sigma\|}\right)\sum_{i=0}^{l-1}\frac{2^{l-i}}{\sqrt{\pi}\,(2(l-i)-1)!!}\left(\frac{z}{\sqrt{2\|\Sigma\|}}\right)^{2(l-i)-1}.$$

We notice that

$$e^{-x^2}\sum_{i=0}^{l-1}\frac{2^{l-i}\,x^{2(l-i)-1}}{\sqrt{\pi}\,(2(l-i)-1)!!} + \operatorname{erfc}(x) \le \frac{2^l}{\sqrt{\pi}}\,(l+1)x^{2l-1}\,e^{-x^2}$$

when $x \ge 1$ and $x^{2l-1} \le x$ when $x \le 1$. Since $z(2l+1, \|\Sigma\|) \to 0$ and $z(2l+1, \|\Sigma\|)/\sqrt{\|\Sigma\|} \to \infty$ as $\|\Sigma\| \to 0$, there exists a positive constant σ_{2l+1} such that $\sqrt{2\|\Sigma\|} \le z(2l+1, \|\Sigma\|) \le 1$ when $\|\Sigma\| < \sigma_{2l+1}$. Therefore the solution $\tilde{z}(2l+1, \|\Sigma\|)$ to the equation

$$\frac{1}{(l+1)}\sqrt{\frac{\pi}{2}}\,\|\Sigma\|^{l-\frac{1}{2}} = \exp\left(-\frac{z^2}{2\|\Sigma\|}\right)$$

is greater than $z(2l+1, \|\Sigma\|)$ when $\|\Sigma\| < \sigma_{2l+1}$. Hence

$$\rho_\kappa(\xi,\, y_0) \le \sqrt{-\|\Sigma\| \log\left(\frac{\pi}{2(l+1)^2}\|\Sigma\|^{2l-1}\right)}$$

for all Σ such that $\|\Sigma\| < \min\{\sigma_{2l+1}, (2(l+1)^2/\pi)^{1/(2l-1)}\}$. □

References

[1] Mario Bertero and G. A. Viano. On probabilistic methods for the solution of improperly posed problems. *Boll. Un. Mat. Ital. B (5)*, 15: 453–508, 1978.

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

[3] R. M. Dudley. Distances of probability measures and random variables. *Ann. Math. Statist*, 39:1563–1572, 1968.

[4] R. M. Dudley. *Real Analysis and Probability*. Wadsworth, 2003.

[5] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, 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*, 21:399–412, 2005.

[7] Heinz W. Engl and Anton Wakolbinger. Continuity properties of the extension of a locally Lipschitz continuous map to the space of probability measures. *Monatsh. Math.*, 100:85–103, 1985.

[8] Ky Fan. Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit. *Mat. Z.*, 49:681–683, 1944.

[9] I I Gihman and A. V. Skorohod. *The Theory of Stochastic Processes I*. Springer-Verlag, 1974.

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

[11] Andreas Hofinger. Assessing uncertainty in linear inverse problems with the metrics of Ky Fan and Prokhorov. Technical report, SFB F013, 2006.

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

[13] Andreas Hofinger. 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.

[14] Andreas Hofinger and Stefan Kindermann. Assessing uncertainty in nonlinear inverse problems with the metric of Ky Fan. Technical report, SFB F013, 2006.

[15] Thorsten Hohage. Regularization of exponentially ill-posed problems. *Numer. Funct. Anal. Optim.*, 21:439–464, 2000.

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

[17] Jari P. Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Springer-Verlag, 2004.

[18] Hui-Hsiung Kuo. *Gaussian Measures in Banach Spaces*. Springer-Verlag, 1975.

[19] Markku S. Lehtinen, Lassi Päivärinta, and Erkki Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5:599–612, 1989.

[20] Mark A. Lukas. Convergence Rates for Regularized Solutions. *Math. Comp.*, 51:107–131, 1988.

[21] Avi Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrsch. Verw. Gebiete*, 65:385–397, 1984.

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

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

[24] V Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist*, 36:423–439, 1965.

[25] U. Tautenhahn. Optimality for ill-posed problems under general source conditions. *Numer. Funct. Anal. Optim.*, 19:377–398, 1998.

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

[27] Ward Whitt. Preservation of rates of convergence under mappings. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 29:39–44, 1974.