Convergence rates for linear inverse problems in the presence of an additive normal noise*

Andreas Hofinger† Hanna K. Pikkarainen†‡

Abstract

In this work, we examine a finite-dimensional linear inverse problem where the measurements are disturbed by an additive normal noise. The problem is solved both in the frequentist and in the Bayesian frameworks. Convergence of the used methods when the noise tends to zero is studied in the Ky Fan metric. The obtained convergence rate results and parameter choice rules are of a similar structure for both approaches.

1 Introduction

We are interested in the linear 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, a *least squares solution* to the inverse problem (1) is a vector $z \in \mathbb{R}^n$ such that

$$\|Az - y\| = \min_{w\in\mathbb{R}^n} \|Aw - y\|.$$

If the null space $\mathcal{N}(A)$ of the matrix A is nontrivial, there exist several least squares solutions. An additional requirement is needed for the uniqueness.

The *least squares minimum norm solution* $x^\dagger$ to (1) is the least squares solution with the minimal norm, i. e.,

$$x^\dagger := \arg\min_{z\in\mathbb{R}^n} \left\{ \|z\| : \|Az - y\| = \min_{w\in\mathbb{R}^n} \|Aw - y\| \right\}.$$

*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

For the linear problem (1),

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

where $A^\dagger$ is the Moore–Penrose inverse of A. Furthermore, the set of all least squares solutions is $x^\dagger + \mathcal{N}(A)$.

We assume that the measurements are disturbed by an additive noise. Since problem (1) is unstable, observed inexact data y^ε cannot be used directly to infer an approximate solution $A^\dagger y^\varepsilon$ to (1) but some regularization technique must be applied to obtain an approximate solution x^ε.

A regularization method should naturally be such that less noise leads to a better approximation. An accepted quality criterion are convergence rate results in terms of the noisy data, i. e., results of the form

$$\rho_x(x^\varepsilon, x^\dagger) = \mathcal{O}(f(\rho_y(y^\varepsilon, y))) \tag{2}$$

where ρ_x and ρ_y are suitable metrics.

In the deterministic regularization theory, the noise in the data is assumed to be bounded, i. e., $\|y^\varepsilon - y\| \leq \delta$ for some $\delta > 0$. We are interested in the case where the noise can be modelled by a normal random variable. Then, in principle, $\|y^\varepsilon - y\|$ can be arbitrarily large. Hence the results of the deterministic theory cannot be utilized and less restrictive stochastic concepts must be used.

When the noise is considered as a random variable, the regularized solution is also a random variable. Hence the distance between $x^\dagger$ and the regularized solution needs to be measured in an appropriate metric in the space of random variables. Often, convergence results in the presence of a stochastic noise are given in the terms of the mean squares error (cf. [1, 25, 20, 8]). In [10, 9], the Ky Fan metric (a quantitative version of the convergence in probability) was used to deduce convergence results for linear inverse problems.

In addition to the frequentist approach (see section 2), the Bayesian inversion theory (see section 4) is a widely used tool to tackle stochastic inverse problems. A main argument in favor of the Bayesian approach is that not only a single regularized solution is computed (as it is done in the frequentist setting) but that a whole distribution can be obtained. The first convergence results for the posterior distribution were presented in [13].

In this work, we use the metric of Ky Fan to derive convergence rate results in the framework of an additive normal noise for Tikhonov type of regularized solutions (i. e., the frequentist setting). Moreover, we show that with an analogous approach, similar results can be obtained for the MAP estimate in the Bayesian framework. By coupling the metric of Ky Fan with the metric of Prokhorov, we can even obtain corresponding results for the posterior distribution.

This paper can be seen as a step towards the building of a bridge between these two—seemingly different—statistical approaches to inverse problems, i.e., between the frequentist and the Bayesian inversion theories.

2 Frequentist approach to linear inverse problems

The frequentist inversion theory is the stochastic counterpart to the deterministic inversion theory. In the frequentist approach stochastic inverse problems are solved by using some regularization technique. Possible prior information about the true solution to the problem is taken into account in regularization. Since the measured data y^ε usually can be modelled by a random variable, the regularized solution is also a random variable. Convergence results in the frequentist framework are given in the terms of an appropriate metric in the space of random variables.

We suppose that model (1) for the measurement situation is exact and the true solution x is deterministic. We assume that the measurements are perturbed by an additive noise, i. e.,

$$y^\varepsilon(\omega) = Ax + \varepsilon(\omega) \tag{3}$$

where ε is a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}^m$. The distribution of the noise ε is known and possibly allows also large values with small probability.

In the frequentist framework, all probabilities correspond to the frequences of random events. Hence, for a known $x \in \mathbb{R}^n$ the frequences of events $\{y^\varepsilon \in B\}$ related to repeated measurements approximate the probabilities of the random events $\{\omega \in \Omega : Ax + \varepsilon(\omega) \in B\}$ where B is a Borel set in $\mathbb{R}^m$.

In some application, another least squares solution to the linear problem (1) than the least squares minimum norm solution $x^\dagger$ may be of interest, e. g., the least squares solution that minimized another norm than the Euclidean norm. Let $G \in \mathbb{R}^{n\times n}$ be a positive definite symmetric matrix and $x_0 \in \mathbb{R}^n$. The *least squares* (G, x_0)*-minimum norm solution* $x^\dagger_{G,x_0}$ to (1) is

$$x^\dagger_{G,x_0} := \arg\min_{z\in\mathbb{R}^n} \left\{ \|z - x_0\|_G : \|Az - y\| = \min_{w\in\mathbb{R}^n} \|Aw - y\| \right\}$$

where $\|z\|_G := (z, Gz)^{1/2}$ is the norm defined by the matrix G. For the linear problem (1),

$$x^\dagger_{G,x_0} = x^\dagger + \left(G^{1/2} P_{\mathcal{N}(A)}\right)^\dagger G^{1/2} \left(x_0 - x^\dagger\right)$$

where $P_{\mathcal{N}(A)}$ is the orthogonal projection onto $\mathcal{N}(A)$ [18, theorem (20.9)]. Obviously, $x^\dagger = x^\dagger_{I,0}$.

Let $G \in \mathbb{R}^{n\times n}$ and $S \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices and $x_0 \in \mathbb{R}^n$. Given a measurement $y^\varepsilon(\omega)$, we construct the regularized solution $x^\varepsilon_{\alpha,S,G,x_0}$ to the linear problem (1) as a minimizer of the functional

$$\|Ax - y^\varepsilon(\omega)\|^2_S + \alpha\|x - x_0\|^2_G$$

where $\alpha > 0$. Pointwise regularization is considered here, i. e., the functional depends on the realization $y^\varepsilon(\omega)$, not on the random variable y^ε. When $G = I$, $S = I$, and $x_0 = 0$, the minimizer is called the Tikhonov regularized solution.

For the linear problem (1) the regularized solution can be given explicitly; with the regularization parameter α and the noisy data $y^\varepsilon(\omega)$ it is

$$x^\varepsilon_{\alpha,S,G,x_0}(\omega) = \left(A^T S A + \alpha G\right)^{-1} \left(A^T S y^\varepsilon(\omega) + \alpha G x_0\right). \tag{4}$$

As a linear transformation of y^ε, the regularized solution $x^\varepsilon_{\alpha,S,G,x_0}$ is also a random variable.

In deterministic convergence results it is assumed that $\|y - y^\varepsilon(\omega)\| \leq \delta$ for some $\delta > 0$, i.e., the noise is bounded. In contrast, in this work we want to allow unbounded noise. Hence the results of the deterministic theory cannot be utilized and less restrictive stochastic concepts must be used.

Distances between random variables can be measured with several different metrics. Here, we utilize the *metric of Ky Fan*:

Definition 1 (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*

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

The Ky Fan metric gives a quantitative version of the convergence in probability; for some background on this metric see [7, 11, 3]. The Ky Fan distance has been used to study convergence in stochastic inverse problems in [10, 9, 12, 13].

The following theorem combines the Ky Fan distance between the regularized solution (4) and the least squares (G, x_0)-minimum norm solution with the Ky Fan distance between the noisy and the noiseless measurements.

Theorem 2. *Let $G \in \mathbb{R}^{n\times n}$ and $S \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices and $x_0 \in \mathbb{R}^n$. Furthermore, let $x^\varepsilon_{\alpha,S,G,x_0}$ be the regularized solution* (4) *to the linear problem* (1) *with the regularization parameter α and the data y^ε where y^ε satisfies the additive noise model* (3)*. Then*

$$\begin{aligned}&\rho_{\mathrm{K}}\left(x^\varepsilon_{\alpha,S,G,x_0},\, x^\dagger_{G,x_0}\right) \\ &\leq \max\left\{\frac{\alpha}{\lambda_p^2 + \alpha}\sqrt{\frac{\lambda^G_{\mathrm{MAX}}}{\lambda^G_{\mathrm{MIN}}}}\left\|x^\dagger_{G,x_0} - x_0\right\| + \frac{1}{2\sqrt{\alpha}}\sqrt{\frac{\lambda^S_{\mathrm{MAX}}}{\lambda^G_{\mathrm{MIN}}}}\,\rho_{\mathrm{K}}(y^\varepsilon,\, y),\, \rho_{\mathrm{K}}(y^\varepsilon,\, y)\right\}\end{aligned} \tag{5}$$

where λ^G_{MIN} and λ^G_{MAX} are the minimal and the maximal eigenvalues of G, respectively, λ^S_{MAX} is the maximal eigenvalue of S, and λ_p is the minimal positive singular value of $S^{1/2}AG^{-1/2}$.

Proof. We may rewrite

$$\begin{aligned} &x^{\varepsilon}_{\alpha,S,G,x_0}(\omega) \\ &= G^{-1/2}\left(G^{-1/2}A^TSAG^{-1/2}+\alpha I\right)^{-1}G^{-1/2}A^TS\left(y^{\varepsilon}(\omega)-Ax_0\right)+x_0. \end{aligned}$$

Thus $G^{1/2}(x^{\varepsilon}_{\alpha,S,G,x_0}(\omega)-x_0)$ is the Tikhonov regularized solution to the linear inverse problem

$$S^{1/2}AG^{-1/2}z = S^{1/2}(y-Ax_0) \tag{6}$$

with the regularization parameter α and the noisy data $S^{1/2}(y^{\varepsilon}(\omega)-Ax_0)$. The least squares minimum norm solution to (6) is

$$z^{\dagger} = (AG^{-1/2})^{\dagger}(y-Ax_0).$$

By using the singular value decomposition of the matrix $S^{1/2}AG^{-1/2}$ we can estimate

$$\begin{aligned} &\left\|x^{\varepsilon}_{\alpha,S,G,x_0}(\omega)-(x_0+G^{-1/2}z^{\dagger})\right\| \\ &\qquad\leq \frac{\alpha}{\lambda_p^2+\alpha}\sqrt{\|G^{-1}\|}\left\|z^{\dagger}\right\|+\frac{1}{2\sqrt{\alpha}}\sqrt{\|G^{-1}\|\,\|S\|}\,\|y^{\varepsilon}(\omega)-y\|. \end{aligned}$$

According to the definition of the Moore-Penrose inverse $x_0+G^{-1/2}z^{\dagger} = x^{\dagger}_{G,x_0}$. Therefore

$$\begin{aligned} &\left\|x^{\varepsilon}_{\alpha,S,G,x_0}(\omega)-x^{\dagger}_{G,x_0}\right\| \\ &\qquad\leq \frac{\alpha}{\lambda_p^2+\alpha}\sqrt{\frac{\lambda^G_{\text{MAX}}}{\lambda^G_{\text{MIN}}}}\left\|x^{\dagger}_{G,x_0}-x_0\right\|+\frac{1}{2\sqrt{\alpha}}\sqrt{\frac{\lambda^S_{\text{MAX}}}{\lambda^G_{\text{MIN}}}}\,\|y^{\varepsilon}(\omega)-y\|. \end{aligned} \tag{7}$$

The point-wise obtained norm bound (7) can be lifted to be bound (5) in the Ky Fan metric according to [13, theorem 6]. □

Remark 3. The Ky Fan metrics in theorem 2 are calculated by assuming that the underlying metrics in $\mathbb{R}^n$ and $\mathbb{R}^m$ are the Euclidean metrics, not the metrics introduced by the G-norm and the S-norm, respectively. The appearance of the eigenvalues of G and S in (5) is the consequence of that choice.

The above estimate gives a bound on the error in the Ky Fan metric and leads to parameter choice rules for α that ensure convergence rates for the error as in (2) as $\rho_{\text{K}}(y^{\varepsilon}, y) \to 0$. Such results have been obtained in, e. g., [10, chapter 3] when $G = I$, $S = I$, and $x_0 = 0$.

Since the upper bound (5) depends on the matrix A through the minimal positive singular value of $S^{1/2}AG^{-1/2}$, convergence rate results will not

be independent of A. Especially, when A is ill-conditioned, constants in convergence rate results can be significantly big.

In the deterministic inversion theory, convergence rate results for linear inverse problems independent of the matrix A are only possible when additional assumptions on the solution are imposed (see, e. g., [4]). These assumptions can, for instance, be formulated in terms of abstract smoothness conditions, so called *source conditions*.

Definition 4. *The least squares minimum norm solution $x^\dagger$ to the linear problem* (1) *satisfies a* deterministic source condition *with source function f if there exist $v \in \mathbb{R}^n$ and $\tau > 0$ such that*

$$x^\dagger = f(A^T A)v \quad \text{and} \quad \|v\| \leq \tau. \tag{8}$$

Typical choices of the source function are $f(\lambda) = \lambda^\nu$, $\nu \leq 1$ and $f(\lambda) = (-\log \lambda)^{-\nu}$ (see [4, 14]).

Definition 5. *The source function f allows the* deterministic convergence rate *h if there exists an increasing function h such that $h(0) = 0$ and*

$$x \in \left\{ z \in \mathcal{N}(A)^\perp : z = f\left(A^T A\right)\, v,\ \|v\| \leq \tau \right\} \implies \|x - x_\alpha\| \leq \tau h(\alpha)$$

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

For the Hölder and the logarithmic source functions f above, it has been shown that $f = h$ (see [4] and [14], respectively). Nevertheless, this is not the case in general, e. g., when saturation occurs (cf., e. g., [4]). 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, 24].

Theorem 6. *Let $G \in \mathbb{R}^{n\times n}$ and $S \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices and $x_0 \in \mathbb{R}^n$. Furthermore, let $x^\varepsilon_{\alpha,S,G,x_0}$ be the regularized solution* (4) *to the linear problem* (1) *with the regularization parameter α and the data y^ε where y^ε satisfies the additive noise model* (3)*. Assume that the source function f allows the deterministic convergence rate h and that there exist $v \in \mathbb{R}^n$ and $\tau > 0$ such that*

$$G^{1/2}\left(x^\dagger_{G,x_0} - x_0\right) = f\left(G^{-1/2} A^T S A G^{-1/2}\right) v \quad \text{and} \quad \|v\| \leq \tau. \tag{9}$$

Then

$$\rho_{\text{K}}\left(x^\varepsilon_{\alpha,S,G,x_0},\, x^\dagger_{G,x_0}\right) \leq \max\left\{ \frac{\tau h(\alpha)}{\sqrt{\lambda^G_{\text{MIN}}}} + \frac{1}{2\sqrt{\alpha}} \sqrt{\frac{\lambda^S_{\text{MAX}}}{\lambda^G_{\text{MIN}}}}\, \rho_{\text{K}}(y^\varepsilon,\, y),\, \rho_{\text{K}}(y^\varepsilon,\, y) \right\} \tag{10}$$

where λ^G_{MIN} is the minimal eigenvalue of G and λ^S_{MAX} is the maximal eigenvalue of S.

Proof. According to assumption (9) the least squares minimum norm solution $z^\dagger$ to problem (6) satisfy the deterministic source condition (8) with the source function f and the constant $\tau > 0$. Hence by the stability bounds for (deterministic) Tikhonov regularization (see [4, chapter 5]),

$$\left\| x^{\varepsilon}_{\alpha,S,G,x_0}(\omega) - x^{\dagger}_{G,x_0} \right\| \leq \frac{\tau h(\alpha)}{\sqrt{\lambda^G_{\text{MIN}}}} + \frac{1}{2\sqrt{\alpha}} \sqrt{\frac{\lambda^S_{\text{MAX}}}{\lambda^G_{\text{MIN}}}} \, \|y^{\varepsilon}(\omega) - y\|. \tag{11}$$

The point-wise obtained norm bound (11) can be lifted to be bound (10) in the Ky Fan metric according to [13, theorem 6]. □

If y^ε has a known distribution, the bounds in (5) and (10) can be translated into a more concrete shape. In particular, if the distribution of y^ε is normal, a convergence result in terms of the covariance matrix is of interest.

3 Convergence rates for the frequentist approach

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 measurable function from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}^m$ whose probability 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 this paper, we examine the situation in which there is an additive normal noise in the measurements. Let $\Sigma \in \mathbb{R}^{m\times m}$ be a positive definite symmetric matrix. We suppose that the distribution of the noise ε is normal with mean 0 and covariance matrix $\sigma^2\Sigma$ with some $\sigma > 0$. The matrix Σ describes the cross-correlation between the coordinates of the noise and σ is a tuning parameter. Without loss of generality we may assume that $\|\Sigma\| = 1$.

According to the model (3) the distribution of y^ε is $\mathcal{N}(y, \sigma^2\Sigma)$. An upper bound for the Ky Fan distance $\rho_{\text{K}}(y^\varepsilon, y)$ between the noisy and the exact measurements is given in the following lemma for small enough σ.

Lemma 7. ([13, lemma 7]) *Let $y_0 \in \mathbb{R}^m$ and $\Sigma \in \mathbb{R}^{m\times m}$ be a positive definite symmetric matrix such that $\|\Sigma\| = 1$. Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(y_0, \sigma^2\Sigma)$ for some $\sigma > 0$. 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 constants $\sigma(m)$ such that

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

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

In deterministic regularization theory, convergence rates are calculated when the noise level tends to zero. If the noise is assumed to be normal, an interesting situation is the case when the noise tends to the zero random variable. When σ is small, the distribution of the noise is more concentrated around the origin and can be seen as a smooth approximation of the point measure at the origin. Therefore, intuitively the noise should approach to zero as $\sigma \to 0$. According to the above lemma this intuitive property is valid in the Ky Fan metric, i. e., $\rho_{\text{K}}(\varepsilon,\, 0) \to 0$ as $\sigma \to 0$. Therefore the situation of interest in this paper is $\sigma \to 0$.

In deterministic regularization theory, the regularization parameter is chosen in dependence of the noise level (and maybe also the noisy data) to obtain convergence (cf. [4]). If the noise is assumed to be normal and the distance between random variables is measured in the Ky Fan metric, the norm of the covariance matrix of the noise and the regularization parameter are related.

Theorem 8. *Let $G \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{m\times m}$, and $\Sigma \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices, $\|\Sigma\| = 1$, and $x_0 \in \mathbb{R}^n$. Let ε be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ε is $\mathcal{N}(0, \sigma^2\Sigma)$ for some $\sigma > 0$. Furthermore, let $x^{\varepsilon}_{\alpha,S,G,x_0}$ be the regularized solution (4) to the linear problem (1) with the regulatization parameter α and the data y^{ε} where y^{ε} satisfies the additive noise model (3). If $\alpha(\sigma)$ fulfils*

$$\alpha(\sigma) \longrightarrow 0 \quad \textit{and} \quad \sigma\sqrt{-\frac{\log\left(C(m)\sigma^{2\kappa(m)}\right)}{\alpha(\sigma)}} \longrightarrow 0 \tag{12}$$

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

$$\rho_{\text{K}}\left(x^{\varepsilon}_{\alpha(\sigma),S,G,x_0},\, x^{\dagger}_{G,x_0}\right) \longrightarrow 0$$

as $\sigma \to 0$.

Proof. By combining the results of theorem 2 and lemma 7 there exists a positive constant $\sigma(m)$ such that

$$\begin{aligned}\rho_{\text{K}}\left(x^{\varepsilon}_{\alpha,S,G,x_0},\, x^{\dagger}_{G,x_0}\right) \leq{}& \frac{\alpha}{\lambda_p^2+\alpha}\sqrt{\frac{\lambda^G_{\text{MAX}}}{\lambda^G_{\text{MIN}}}}\left\|x^{\dagger}_{G,x_0} - x_0\right\| \\ &+ \max\left\{\frac{1}{2\sqrt{\alpha}}\sqrt{\frac{\lambda^S_{\text{MAX}}}{\lambda^G_{\text{MIN}}}},\, 1\right\}\sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}\end{aligned}$$

for all $\sigma < \sigma(m)$ where λ^G_{MIN} and λ^G_{MAX} are the minimal and the maximal eigenvalues of G, respectively, λ^S_{MAX} is the maximal eigenvalue of S, and λ_p is the minimal positive singular value of $S^{1/2}AG^{-1/2}$. By choices (12) the Ky Fan distance between the regularized solution (4) and the least squares (G, x_0)-minimum norm solution converges to zero. □

Theorem 2 and lemma 7 lead also to convergence *rate* results. But when $x^\dagger_{G,x_0}$ satisfies a specific smoothness condition, we may use theorem 6 and obtain convergence rate results independent of the matrix A. In the following theorem, the result for the Hölder type of smoothness condition is presented.

Theorem 9. *Let $G \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{m\times m}$, and $\Sigma \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices, $\|\Sigma\| = 1$, and $x_0 \in \mathbb{R}^n$. Let ε be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ε is $\mathcal{N}(0, \sigma^2\Sigma)$ for some $\sigma > 0$. Let $x^\varepsilon_{\alpha,S,G,x_0}$ be the regularized solution (4) to the linear problem (1) with the regulatization parameter α and the data y^ε where y^ε satisfies the additive noise model (3). Suppose that there exist $v \in \mathbb{R}^n$ and $\tau > 0$ such that*

$$G^{1/2}\left(x^\dagger_{G,x_0} - x_0\right) = \left(G^{-1/2}A^T SAG^{-1/2}\right)^\nu v \quad \text{and} \quad \|v\| \le \tau$$

for some $0 < \nu \le 1$. Furthermore, let α be chosen as

$$\alpha \sim \left(\sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}\right)^{\frac{2}{2\nu+1}} \tag{13}$$

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

$$\rho_{\text{K}}\left(x^\varepsilon_{\alpha,S,G,x_0},\, x^\dagger_{G,x_0}\right) \le \mathcal{O}\left(\left(\sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}\right)^{\frac{2\nu}{2\nu+1}}\right). \tag{14}$$

Proof. The source function $f(\lambda) = \lambda^\nu$ allows the deterministic convergence rate $h(\lambda) = \lambda^\nu$ [4, (5.18)]. Therefore by combining the results of theorem 6 and lemma 7 there exists a positive constant $\sigma(m)$ such that

$$\begin{aligned}\rho_{\text{K}}&\left(x^\varepsilon_{\alpha,S,G,x_0},\, x^\dagger_{G,x_0}\right)\\ &\le \frac{\tau\alpha^\nu}{\sqrt{\lambda^G_{\text{MIN}}}} + \max\left\{\frac{1}{2\sqrt{\alpha}}\sqrt{\frac{\lambda^S_{\text{MAX}}}{\lambda^G_{\text{MIN}}}},\, 1\right\}\sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}\end{aligned}$$

for all $\sigma < \sigma(m)$ where λ^G_{MIN} is the minimal eigenvalue of G and λ^S_{MAX} is the maximal eigenvalue of S. Due to the first term we need $\alpha \to 0$. Consequently, the maximum in the second term is equal to $\sqrt{\lambda^S_{\text{MAX}}}/2\sqrt{\alpha\lambda^G_{\text{MIN}}}$ as $\alpha \to 0$. By choice (13) the two terms are balanced and rate (14) is received. □

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 $\sigma \to 0$.

The dimension-dependence of (14) seems to resemble a general complication when working with Gaussian random variables in infinite-dimensional spaces. The dimension-dependence can be attributed to the fact that in such a setup the perturbation $\varepsilon(\omega)$ in (3) (and consequently $y^\varepsilon(\omega)$ as well) cannot as a Gaussian random variable belong to the underlying Hilbert space if the covariance of $\varepsilon(\omega)$ is not a trace-class operator (cf. [17, theorem 2.3]).

In the following, we show that the conceptually different Bayesian approach leads to similar convergence rate results and parameter choice rules.

4 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 [16].

The basis of the Bayesian approach to inverse problems is different from the deterministic and the frequentist inversion theories. In the Bayesian framework 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 describe the confidence or the degree of belief that one has into a particular initial guess.

All information available before performing the measurements about the quantity of primary interest is coded in a probability distribution, the so-called *prior distribution*. Even though the quantity of primary interest is assumed to be deterministic, it is modelled by a random variable whose distribution is the prior distribution.

The Bayesian inversion theory is based on the Bayes formula. The solution to the inverse problem after performing the measurements is the *posterior distribution* of the random variables of interest. The Bayes formula describes how the prior information and measurements have to be combined to give the posterior distribution; by this formula the posterior distribution is proportional to the product of the prior distribution and the *likelihood function* which is given by the model for the indirect measurements.

Consequently, in the Bayesian approach 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. Since now the prior information is coded via random variables, we need a notation that differs slightly from the previous sections. We

denote random variables by capital letters and their realizations by lower case letters.

Let X and Y be random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}^n$ 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

$$Y = AX + E \tag{15}$$

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

$$\pi_{\text{pr}}(x) \propto \exp\left(-\frac{1}{2\gamma^2}\,(x - x_0)^T\Gamma^{-1}(x - x_0)\right) \tag{16}$$

and

$$\pi_{\text{noise}}(e) \propto \exp\left(-\frac{1}{2\sigma^2}\,e^T\Sigma^{-1}e\right) \tag{17}$$

where $\gamma, \sigma > 0$, $x_0 \in \mathbb{R}^n$, and $\Gamma \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{m\times m}$ are positive definite symmetric matrices. Hence the covariance matrices of the prior distribution and the noise are $\gamma^2\Gamma$ and $\sigma^2\Sigma$, respectively. The matrices Γ and Σ describe the cross-correlation of the coordinates of the prior distribution and the noise, respectively, and γ and σ are tuning parameters.

For the additive noise model (15), the Bayes theorem yields the posterior probability density

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

(independently of the particular structure of π_{pr} and π_{noise}). For the case of normal random variables, this posterior distribution can be computed explicitly:

Theorem 10. ([16, theorem 3.7]) *Let $x_0 \in \mathbb{R}^n$, $\Gamma \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{m\times m}$ be positive definite symmetric matrices, and $\gamma, \sigma > 0$. Let X and E be independent random variables with probability densities* (16) *and* (17)*, respectively. Assume that the measurement Y satisfies the additive noise model* (15)*. Then the posterior distribution μ_{post} of X conditioned on the data y_{data} is normal and has the probability density*

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

where the posterior mean is

$$\bar{x} = \left(A^T\Sigma^{-1}A + \frac{\sigma^2}{\gamma^2}\,\Gamma^{-1}\right)^{-1}\left(A^T\Sigma^{-1}y_{\text{data}} + \frac{\sigma^2}{\gamma^2}\,\Gamma^{-1}\,x_0\right) \tag{18}$$

and the posterior covariance matrix is

$$\Gamma_{\text{post}} = \sigma^2 \left(A^T \Sigma^{-1} A + \frac{\sigma^2}{\gamma^2} \Gamma^{-1} \right)^{-1}. \tag{19}$$

In the following section, we show that the Bayesian approach leads to similar convergence rate results as the frequentist framework when convergence is measured in the Ky Fan metric.

5 Convergence rates for the Bayesian approach

In this section we investigate convergence rate results for the Bayesian approach in the Ky Fan metric. Let $x_0 \in \mathbb{R}^n$, $\Gamma \in \mathbb{R}^{n \times n}$ and $\Sigma \in \mathbb{R}^{m \times m}$ be positive definite symmetric matrices, and $\gamma, \sigma > 0$. We assume that the prior distribution of the unknown is $\mathcal{N}(x_0, \gamma^2 \Gamma)$ and the noise has the distribution $\mathcal{N}(0, \sigma^2 \Sigma)$ (as in theorem 10). The data y_{data} is a realization of the random variable Y. Since the true solution is supposed to be deterministic and hence the noiseless measurement is equal to y, the distribution of Y is $\mathcal{N}(y, \sigma^2 I)$. Without loss of generality we may assume that $\|\Gamma\| = \|\Sigma\| = 1$.

Instead of the posterior distribution μ_{post} the maximum a posteriori (MAP) estimate

$$x_{\text{MAP}} = \arg \max_{x \in \mathbb{R}^n} \pi_{\text{post}}(x)$$

is often given as a solution to an inverse problem in the Bayesian approach.[1] At first we determine convergence results for this particular value. They essentially resemble the results for the frequentist inversion theory, considered in the first part of this paper.

In paper [13] convergence results for the whole posterior distribution in the Bayesian approach were presented when $\Gamma = I$ and $\Sigma = I$. We show that similar results can be obtained for arbitrary positive definite symmetric Γ and Σ with $\|\Gamma\| = \|\Sigma\| = 1$ and emphasise the similarities with the results related to the frequentist approach and the MAP estimate.

5.1 Convergence rates for the MAP estimate

Under the above assumptions the MAP estimate x_{MAP} is actually a realization of the random variable

$$X_{\text{MAP}}(\omega) = \left(A^T \Sigma^{-1} A + \frac{\sigma^2}{\gamma^2} \Gamma^{-1} \right)^{-1} \left(A^T \Sigma^{-1} Y(\omega) + \frac{\sigma^2}{\gamma^2} \Gamma^{-1} x_0 \right)$$

(cf. (18)). The MAP estimate resembles the regularized solution (4) to the linear problem (1) with the regularization parameter $\alpha := \sigma^2/\gamma^2$ and the noisy data $Y(\omega)$ when $G = \Gamma^{-1}$ and $S = \Sigma^{-1}$.

[1] Also the conditional mean (CM) estimate could be used (see [16, chapter 3]). In our setup, the MAP and CM estimates coincide.

For the Tikhonov type of regularization methods the choice of the regularization parameter plays a crucial role. If we have some external information about the quantity of primary interest, e.g., structural information or previous measurements, the Bayesian approach allows us to use all available a-priori information to choose an appropriate regularization parameter since the MAP estimate is a regularized solution.

Because of the specific form of the MAP estimate we may use the convergence results of section 3. The MAP estimate converges to the least squares (Γ^{-1}, x_0)-minimum norm solution as $\sigma \to 0$ if the norm of the covariance matrix of the prior distribution depends on σ and satisfies certain conditions:

Theorem 11. *Let the assumption of theorem 10 be valid and* $\|\Gamma\| = \|\Sigma\| = 1$. *Let us denote the least squares* (Γ^{-1}, x_0)*-minimum norm solution by* $x^\ddagger$. *If* $\gamma(\sigma)$ *satisfies*

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

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

$$\rho_{\kappa}\left(X_{\mathrm{MAP}},\, x^\ddagger\right) \longrightarrow 0$$

as $\sigma \to 0$.

The proof is analogous to the proof of theorem 8. For example, $\gamma(\sigma) \sim \sigma^\eta$ with some $0 < \eta < 1$ fulfills the requirements of theorem 11.

Besides this convergence result, also convergence *rate* results can be obtained. When the least squares (Γ^{-1}, x_0)-minimum norm solution satisfies a specific smoothness condition, the convergence rate results do not depend on the matrix A. The proof of the following theorem is similar to the proof of theorem 9.

Theorem 12. *Let the assumption of theorem 10 be satisfied and* $\|\Gamma\| = \|\Sigma\| = 1$. *Let us denote the least squares* (Γ^{-1}, x_0)*-minimum norm solution by* $x^\ddagger$. *Suppose that there exist* $v \in \mathbb{R}^n$ *and* $\tau > 0$ *such that*

$$\Gamma^{-1/2}\left(x^\ddagger - x_0\right) = \left(\Gamma^{1/2}A^T\Sigma^{-1}A\Gamma^{1/2}\right)^\nu v \quad \textit{and} \quad \|v\| \le \tau$$

for some $0 < \nu \le 1$. *Furthermore, let* γ *be chosen as*

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

where the constants $C(m)$ *and* $\kappa(m)$ *are given in lemma 7. Then*

$$\rho_{\kappa}\left(X_{\mathrm{MAP}},\, x^\ddagger\right) \le \mathcal{O}\left(\left(\sigma\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)}\right)^{\frac{2\nu}{2\nu+1}}\right).$$

In theorem 11 as well as 12 we require that the parameter γ must tend to zero in order to obtain the convergence of X_{MAP} to the least squares (Γ^{-1}, x_0)-minimum norm solution. This condition on γ seems obvious when theorem 8 is combined with the fact that σ^2/γ^2 plays the role of the regularization parameter α in the Bayesian approach. Nonetheless, it is counter-intuitive 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 norm of the covariance matrix of the noise, the norm of the covariance matrix of the prior distribution does tend to infinity ($\gamma/\sigma \to \infty$), i. e., the prior distribution becomes non-informative.

5.2 Convergence rates for the posterior distribution

A main argument in favor of the Bayesian approach is that not only a single solution (such as x_{MAP} above) is computed but that a whole distribution can be obtained. The first convergence results for the posterior distribution were presented in [13] when $\Gamma = I$ and $\Sigma = I$. In this section those results are generalized for arbitrary positive definite symmetric Γ and Σ with $\|\Gamma\| = \|\Sigma\| = 1$. The technique for proving the convergence results in the general case is the same as in [13]. Furthermore, we emphasise the similarities with the results for the frequentist approach and the MAP estimate.

As we have seen in theorem 10, the posterior distribution with the data y_{data} is given as

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

with x_{MAP} and Γ_{post} defined in (18) and (19), respectively. As noticed before, the mean of this distribution is a realization of the random variable X_{MAP} while the covariance matrix is constant. Therefore, we may consider the posterior distribution as a random variable,

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

i. e., a measurable function from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{M}(\mathbb{R}^n), \rho_{\text{P}})$, the space of Borel measures on $\mathbb{R}^n$ equipped with the Prokhorov metric ρ_{P}:

Definition 13 (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, 3, 15, 23])*

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

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

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

For some additional background of the Prokhorov distance and a comparison with the Ky Fan metric see, e. g., [11]. The Prokhorov metric has been used to deduce convergence results for stochastic inverse problems in [6, 5, 10, 9, 12, 13].

Since μ_{post} is a random variable on a metric space, we can compute the distance between μ_{post} and the constant random variable $\delta_{x^\ddagger}$ in the Ky Fan metric where $\delta_{x^\ddagger}$ is the point measure at the least squares (Γ^{-1}, x_0)-minimum norm solution $x^\ddagger$.

Theorem 14. *Let the assumption of theorem 10 be satisfied and* $\|\Gamma\| = \|\Sigma\| = 1$. *Let us denote the least squares* (Γ^{-1}, x_0)*-minimum norm solution by* $x^\ddagger$. *Then*

$$\rho_{\text{K}}\left(\mu_{\text{post}},\, \delta_{x^\ddagger}\right) \leq \max\left\{ \rho_{\text{K}}(Y,\, y), \right. \tag{20}$$
$$\left. \frac{\sigma^2}{\gamma^2\lambda_p^2 + \sigma^2}\, \frac{1}{\sqrt{\lambda^{\Gamma}_{\text{MIN}}}}\, \left\| x^\ddagger - x_0 \right\| + \rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) + \frac{\gamma}{2\sigma\sqrt{\lambda^{\Sigma}_{\text{MIN}}}}\, \rho_{\text{K}}(Y,\, y) \right\}$$

where $\lambda^{\Gamma}_{\text{MIN}}$ *and* $\lambda^{\Sigma}_{\text{MIN}}$ *are the minimal eigenvalues of* Γ *and* Σ, *respectively, and* λ_p *is the minimal positive singular value of* $\Sigma^{-1/2} A \Gamma^{1/2}$.

In addition, assume that the source function f *allows the deterministic convergence rate* h *and that there exist* $v \in \mathbb{R}^n$ *and* $\tau > 0$ *such that*

$$\Gamma^{-1/2}(x^\ddagger - x_0) = f\left(\Gamma^{1/2} A^T \Sigma^{-1} A \Gamma^{1/2}\right) v \quad \textit{and} \quad \|v\| \leq \tau.$$

Then

$$\rho_{\text{K}}\left(\mu_{\text{post}},\, \delta_{x^\ddagger}\right) \tag{21}$$
$$\leq \max\left\{ \tau h\left(\frac{\sigma^2}{\gamma^2}\right) + \rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) + \frac{\gamma}{2\sigma\sqrt{\lambda^{\Sigma}_{\text{MIN}}}}\, \rho_{\text{K}}(Y,\, y),\, \rho_{\text{K}}(Y,\, y) \right\}.$$

Let $\lambda^{A,\Sigma}_{\text{MIN}}$ *and* $\lambda^{A,\Sigma}_{\text{MAX}}$ *denote the minimal and the maximal eigenvalues of the matrix* $A^T\Sigma^{-1}A$, *respectively. Then there exist positive constants* $\gamma(n)$ *and* $\sigma(n)$ *such that*

$$\rho_{\text{P}}(\mathcal{N}(0, \Gamma_{\text{post}}),\, \delta_0) \tag{22}$$
$$\leq \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda^{A,\Sigma}_{\text{MIN}} + \sigma^2}} \sqrt{ - \log\left(C(n) \frac{\gamma^{2\kappa(n)}\sigma^{2\kappa(n)}}{\left(\gamma^2\lambda^{A,\Sigma}_{\text{MAX}} + \sigma^2(\lambda^{\Gamma}_{\text{MIN}})^{-1}\right)^{\kappa(n)}} \right)}$$

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

Proof. The proof is similar to the proofs of [13, proposition 11 and theorem 13]. Estimates (20) and (21) follow from the triangle inequality of the Prokhorov metric (e. g., [15]), the translation invariance of the Euclidean norm, the stability bound (7) and (11) where $x^{\varepsilon}_{\alpha,S,G,x_0}(\omega)$, S, G, α, $y^{\varepsilon}(\omega)$, and $x^{\dagger}_{G,x_0}$ are replaced by $X_{\mathrm{MAP}}(\omega)$, Σ^{-1}, Γ^{-1}, σ^2/γ^2, $Y(\omega)$, and $x^{\ddagger}$, respectively, and the lifting argument [13, theorem 6]. In addition, $\lambda^{\Gamma}_{\mathrm{MAX}} = \|\Gamma\| = 1$.

Bound (22) is a consequence of [13, proposition 5 and lemma 7]. By definition (19),

$$\frac{\gamma^2\sigma^2}{\gamma^2\lambda^{A,\Sigma}_{\mathrm{MAX}} + \sigma^2(\lambda^{\Gamma}_{\mathrm{MIN}})^{-1}} \leq \|\Gamma_{\mathrm{post}}\| \leq \frac{\gamma^2\sigma^2}{\gamma^2\lambda^{A,\Sigma}_{\mathrm{MIN}} + \sigma^2}.$$

Therefore the upper bounds $\gamma(n)$ and $\sigma(n)$ need to be chosen such that they satisfies the inequality

$$\lambda^{A,\Sigma}_{\mathrm{MIN}}\theta(n)\gamma^2 + \theta(n)\sigma^2 - \gamma^2\sigma^2 > 0$$

where $\theta(n)$ is the constant of [13, lemma 7]. □

As for the frequentist approach, we can now use this result and deduce parameter choice rules for the Bayesian approach to obtain convergence and convergence rates for the posterior distribution.

Theorem 15. *Let the assumptions of theorem 10 be valid and* $\|\Gamma\| = \|\Sigma\| = 1$. *Let us denote the least squares* (Γ^{-1}, x_0)*-minimum norm solution by* $x^{\ddagger}$. *Let* $\gamma(\sigma)$ *satisfy*

$$\frac{\sigma}{\gamma(\sigma)} \longrightarrow 0 \quad \textit{and} \quad \gamma(\sigma)\sqrt{-\log(C(m,n)\sigma^{2\kappa(m,n)})} \longrightarrow 0 \tag{23}$$

as $\sigma \to 0$ *where the constants* $C(m,n) := C(\max(m,n))$ *and* $\kappa(m,n) := \kappa(\max(m,n))$ *are defined in lemma 7. Then*

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

as $\sigma \to 0$.

Proof. By combining the results of theorem 14 and lemma 7 there exist positive constants $\gamma(n)$ and $\sigma(m,n)$ such that

$$\begin{aligned}
&\rho_{\mathrm{K}}\left(\mu_{\mathrm{post}},\, \delta_{x^{\ddagger}}\right) \\
&\leq \frac{\sigma^2}{\gamma^2\lambda_p^2 + \sigma^2}\,\frac{\|x^{\ddagger} - x_0\|}{\sqrt{\lambda^{\Gamma}_{\mathrm{MIN}}}} + \max\left\{\frac{\gamma}{2\sqrt{\lambda^{\Sigma}_{\mathrm{MIN}}}}, \sigma\right\}\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)} \\
&\quad + \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda^{A,\Sigma}_{\mathrm{MIN}} + \sigma^2}}\sqrt{-\log\left(C(n)\frac{\gamma^{2\kappa(n)}\sigma^{2\kappa(n)}}{\left(\gamma^2\lambda^{A,\Sigma}_{\mathrm{MAX}} + \sigma^2(\lambda^{\Gamma}_{\mathrm{MIN}})^{-1}\right)^{\kappa(n)}}\right)}
\end{aligned} \tag{24}$$

for all $\gamma < \gamma(n)$ and $\sigma < \sigma(m,n)$ where $\lambda_{\text{MIN}}^{A,\Sigma}$ and $\lambda_{\text{MAX}}^{A,\Sigma}$ are the minimal and the maximal eigenvalues of $A^T\Sigma^{-1}A$, respectively, $\lambda_{\text{MIN}}^{\Gamma}$ and $\lambda_{\text{MIN}}^{\Sigma}$ are the minimal eigenvalues of Γ and Σ, respectively, and λ_p is the minimal positive singular value of $\Sigma^{-1/2}A\Gamma^{1/2}$

The second term on the right hand side of (24) 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 $\sigma \to 0$. For the third term it is enough if $\gamma/\sigma \geq 1$ and $\gamma\sqrt{-\log(C(n)\sigma^{2\kappa(n)})} \to 0$ as $\sigma \to 0$. Thus the parameter choice (23) guarantees the convergence. □

Finally, we turn once more to the case of the Hölder type of smoothness condition and obtain that the posterior distribution attains (measured in a combination of the Ky Fan and the Prokhorov metrics) the similar convergence rate as the regularized solution in the frequentist approach (theorem 9) and the MAP estimate (theorem 12).

Theorem 16. *Let the assumption of theorem 10 be valid and* $\|\Gamma\| = \|\Sigma\| = 1$. *Let us denote the least squares* (Γ^{-1}, x_0)*-minimum norm solution by* $x^\ddagger$. *Suppose that there exist* $v \in \mathbb{R}^n$ *and* $\tau > 0$ *such that*

$$\Gamma^{-1/2}(x^\ddagger - x_0) = \left(\Gamma^{1/2}A^T\Sigma^{-1}A\Gamma^{1/2}\right)^\nu v \quad \textit{and} \quad \|v\| \leq \tau$$

for some $0 < \nu \leq 1$. *Furthermore, let* γ *be chosen as*

$$\gamma \sim \left(\sigma^{2\nu}/\sqrt{-\log(C(m,n)\sigma^{2\kappa(m,n)}}\right)^{\frac{1}{2\nu+1}} \tag{25}$$

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

$$\rho_{\text{K}}\left(\mu_{\text{post}},\, \delta_{x^\ddagger}\right) \leq \mathcal{O}\left(\left(\sigma\sqrt{-\log\left(C(m,n)\sigma^{2\kappa(m,n)}\right)}\right)^{\frac{2\nu}{2\nu+1}}\right). \tag{26}$$

Proof. The source function $f(\lambda) = \lambda^\nu$ allows the deterministic convergence rate $h(\lambda) = \lambda^\nu$ [4, (5.18)]. By combining the results of theorem 14 and lemma 7 there exist positive constants $\gamma(n)$ and $\sigma(m,n)$ such that

$$\begin{aligned}
&\rho_{\text{K}}\left(\mu_{\text{post}},\, \delta_{x^\ddagger}\right) \\
&\leq \tau\left(\frac{\sigma}{\gamma}\right)^{2\nu} + \max\left\{\frac{\gamma}{2\sqrt{\lambda_{\text{MIN}}^{\Sigma}}}, \sigma\right\}\sqrt{-\log\left(C(m)\sigma^{2\kappa(m)}\right)} \\
&\quad + \frac{\gamma\sigma}{\sqrt{\gamma^2\lambda_{\text{MIN}}^{A,\Sigma} + \sigma^2}}\sqrt{-\log\left(C(n)\frac{\gamma^{2\kappa(n)}\sigma^{2\kappa(n)}}{\left(\gamma^2\lambda_{\text{MAX}}^{A,\Sigma} + \sigma^2(\lambda_{\text{MIN}}^{\Gamma})^{-1}\right)^{\kappa(n)}}\right)}
\end{aligned} \tag{27}$$

for all $\gamma < \gamma(n)$ and $\sigma < \sigma(m,n)$ where $\lambda_{\text{MIN}}^{A,\Sigma}$ and $\lambda_{\text{MAX}}^{A,\Sigma}$ are the minimal and the maximal eigenvalues of $A^T\Sigma^{-1}A$, respectively, and $\lambda_{\text{MIN}}^{\Gamma}$ and $\lambda_{\text{MIN}}^{\Sigma}$ are the minimal eigenvalues of Γ and Σ, respectively.

To obtain a convergence rate for the error the right hand side of (27) should be minimized for a fixed σ. Since the minimizing $\gamma(\sigma)$ is not easy to derive, we estimate the right hand side of (27) from above. When γ and σ are small enough and $\gamma/\sigma \geq \max\{1, 2\sqrt{\lambda_{\text{MIN}}^{\Sigma}}\}$,

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

By choice (25) the two terms on the right-hand side are balanced and hence rate (26) is obtained. □

6 Conclusions

In this paper, we have examined convergence results in the Ky Fan metric for different statistical inversion theories, namely the frequentist and the Bayesian approaches to linear inverse problems. It turned out that convergence rate results and parameter choice rules for both approaches are similar. This remains also true when in the Bayesian framework the convergence of the whole posterior distribution, not just the MAP estimate, is considered.

Because the MAP estimate under the assumptions of theorem 10 is of the form of the regularized solution (4) in the frequentist setting, the convergence results for the MAP estimate (theorems 11 and 12) are special cases of the results for the regularized solution (theorems 8 and 9).

If the linear problem (1) is over- or exactly determined, i. e., $m \geq n$, the parameter choice rules and the convergence rate obtained in this paper for the posterior distribution (theorems 15 and 16) are the same as for the MAP estimate. Typically, the true solution x is a discretized version of an infinite-dimensional object. To assure computational accuracy x lies in a high dimensional space. On the other hand, only a limited number of measurements can be performed and hence the data belongs to a low dimensional space. Therefore, the linear inverse problem (1) is often underdetermined, i. e., $m < n$. Then the obtained convergence rate for the posterior distribution is slower than for the MAP estimate since the dimension of the unknown, not the dimension of the measurements, determines the speed of convergence (cf. theorem 16). Also the parameter choice rules are effected by the dimension of the unknown (cf. theorems 15 and 16).

In the frequentist approach, the penalty term in the regularization, i. e., the matrix G and the vector x_0 define which least squares solution the regularized solution (4) converges to. The choice of the used metric in the

least squares term , i. e., the matrix S appears only in the constant of the convergence rate result.

For the MAP estimate the matrices G and S are replaced by the inverses of the normalized covariance matrices of the prior and the noise distributions, i. e., the matrices Γ^{-1} and Σ^{-1}, respectively. In addition, the regularization parameter α is defined as the quotient of the norms of the covariance matrices of the noise and the prior distributions. Hence the prior distribution in the Bayesian approach should be chosen with a careful consideration.

Throughout this paper, we considered normal distributions. The normality is an accepted property for the noise, but for the prior information in the Bayesian inversion theory also different choices are in use (see [16, chapter 3] for an overview). In the normal case, the MAP estimate and the posterior distribution can be easily deduced. Alternative prior distributions may lead to better reconstructions, but it is often not possible to calculate explicit solutions, either posterior distributions or point estimates. Hence, 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 obtained convergence results are dimension-dependent in a way that prevents generalization to the infinite-dimensional case. In the frequentist approach to infinite-dimensional linear inverse problems convergence results in the Ky Fan metric have been published (see, e. g., [10, 9]). The Bayesian inversion theory in infinite-dimensional spaces is not completely developed. For Gaussian linear inverse problems the forms of the posterior distribution in some special cases are presented in [21, 19]. Convergence results in the Bayesian approach to infinite-dimensional inverse problems require more sophisticated stochastic analysis than used in this paper.

So far, different statistical inversion theories, i. e., the frequentist and the Bayesian approaches to inverse problems, have mainly been studied by separate communities. In this paper, we have examined both the frequentist and the Bayesian frameworks with the same method to better understand properties and differences of both approaches. We hope this paper will provide a step towards the building of a bridge between the frequentist and the Bayesian inversion theories and allow connections with the deterministic approach to inverse problems.

Acknowledgements

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

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

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. *Real Analysis and Probability*. Wadsworth, 2003.

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

[5] 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.

[6] 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.

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

[8] 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.

[9] Andreas Hofinger. Assessing uncertainty in linear inverse problems with the metrics of Ky Fan and Prokhorov. Technical report 2006–23, SFB F013, 2006.

[10] 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).

[11] 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.

[12] Andreas Hofinger and Stefan Kindermann. Assessing uncertainty in nonlinear inverse problems with the metric of Ky Fan. Technical report 2006–31, SFB F013, 2006.

[13] Andreas Hofinger and Hanna K. Pikkarainen. Convergence rates for the Bayesian approach to linear inverse problems. Technical report 2006–32, SFB F013, 2006.

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

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

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

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

[18] C. Lawson and R. Hanson. *Solving Least Squares Problems*. Prentice-Hall, 1974.

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

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