EARLY STOPPING FOR ENSEMBLE KALMAN–BUCY INVERSION

MAIA TIENSTRA

Abstract. Bayesian linear inverse problems aim to recover an unknown signal from noisy observations, incorporating prior knowledge. This paper analyses a data dependent method to choose the scale parameter of a Gaussian prior. The method we study arises from early stopping methods, which have been successfully applied to a range of problems for statistical inverse problems in the frequentist setting. These results are extended to the Bayesian setting. We study the use of a discrepancy based stopping rule in the setting of random noise. Our proposed stopping rule results in optimal rates under certain conditions on the prior covariance operator. We furthermore derive for which class of signals this method is adaptive. It is also shown that the associated posterior contracts at the optimal rate and provides a conservative measure of uncertainty. We implement the proposed stopping rule using the continuous-time ensemble Kalman–Bucy filter (EnKBF). The fictitious time parameter replaces the scale parameter, and the ensemble size is appropriately adjusted in order to not lose statistical optimality of the computed estimator.

1. Introduction

Bayesian inference methods are widely used in statistical inverse problems. A major challenge is the selection and computational implementation of suitable prior distributions. This problem can be addressed by using hierarchical Bayesian methods and Bayesian model selection. More recently, however, a frequentist analysis of Bayesian methods has gained popularity [24], where choosing the prior is synonymous with choosing the amount of regularisation. Furthermore, a frequentist analysis of posterior credible intervals has also become an active area of research [17, 16, 11]. In this paper, we follow both lines of research and analyse an adaptive choice of the prior. We build on recent results on adaptive choice of the regularisation parameter for statistical inverse problems, which have been studied extensively in [4, 5, 23]. In these papers, the regularisation parameter is chosen using statistical early stopping. We extend these methods as an empirical Bayesian method for selecting the scale parameter of the prior covariance. In addition, the ensemble Kalman Filter (EnKF) and its continuous-time ensemble Kalman-Bucy Filter (EnKBF) have become popular methods for performing Bayesian inference on high-dimensional inverse problems. See [6] for an overview of EnKF and the closely related ensemble Kalman inversion (EKI) [14]. The convergence rates of adaptive EKI have previously been studied for deterministic linear inverse problems in [18]. Here we combine this work with the Bayesian frequentist perspective [17, 16, 11], statistical early stopping [4, 5, 23], and continuous-time EnKBF implementations [19].

Date: March 28, 2024.

1.1. **Problem Formulation.** We will now recall the Bayesian inverse problem setting of [17, 25]. We are interested in recovering the ground truth signal $\theta^\dagger$ from the following observations Y, which we believe to be generated by the following model

$$Y = G\theta^\dagger + \delta\Xi, \tag{1}$$

where $\delta = \frac{\nu}{\sqrt{n}} > 0$, and ν is the noise level assumed to be unknown. Here $G : H_1 \to H_2$ denotes a known linear, compact, continuous operator between two infinite dimensional Hilbert spaces H_1 and H_2 with inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$, respectively. The norms of H_1 and H_2 are denoted by $||\cdot||_1$ and $||\cdot||_2$ respectively. We will denote the adjoint of an operator A between two Hilbert spaces by A^*.

The measurement error Ξ is assumed to be Gaussian white noise and δ denotes the noise level which we will study in the limit $n \to \infty$. The noise Ξ is not an element of H_2, but we can define it as a Gaussian process $(\Xi_h : h \in H_2)$ with mean 0, and covariance $\mathrm{cov}(\Xi_\mathrm{h}, \Xi_{\mathrm{h}'}) = \langle \mathrm{h}, \mathrm{h}'\rangle_2$. The observations are then driven by this process. Thus we observe a Gaussian process $Y = (Y_h : h \in H_2)$ with mean and covariance given by

$$\mathbb{E}Y_h = \langle G\theta^\dagger, h\rangle_2, \quad \mathrm{cov}(\mathrm{Y_h}, \mathrm{Y_{h'}}) = \frac{1}{\mathrm{n}}\langle \mathrm{h}, \mathrm{h}'\rangle_2. \tag{2}$$

In this paper, we follow a Bayesian perspective and place a Gaussian prior over the unknown parameter θ, which is conjugate to (1), implying that the posterior will be Gaussian and analytic up to some normalizing constant. We also further assume that the true $\nu = 1$, to simplify the notation. Particularly, we consider a family of Gaussian priors $\mathcal{N}(0, \tau_n^2 C_0)$ with covariance operator $C_0 : H_1 \to H_1$ and where $\tau_n > 0$ is the scaling parameter of interest.

Proposition 1.1. *(Prop. 3.1 in [16]) For given $\tau_n > 0$, the prior distribution for θ is $\mathcal{N}(0, \tau_n^2 C_0)$ and Y given θ is $\mathcal{N}(G\theta, n^{-1}I)$ distributed. Then the conditional distribution of θ given Y, the posterior, is Gaussian $\mathcal{N}(\widehat{\theta}_{\tau_n}, C_{\tau_n})$ on H_1 with mean*

$$\widehat{\theta}_{\tau_n} := K_{\tau_n} Y \tag{3}$$

and covariance operator

$$C_{\tau_n} := \tau_n^2 C_0 - \tau_n^2 K_{\tau_n}\left(GC_0G^* + \frac{1}{n\tau_n^2}I\right)K_{\tau_n}^*, \tag{4}$$

where the Kalman gain $K_{\tau_n} : H_2 \to H_1$ is the linear continuous operator given by

$$K_{\tau_n} := C_0G^*\left(GC_0G^* + \frac{1}{n\tau_n^2}I\right)^{-1}. \tag{5}$$

We remark that a rigorous construction of the Bayesian set up for infinite dimensional Hilbert space can be found in [10, 11]. We recall that the mean $\widehat{\theta}_{\tau_n}$ also arises formally as the minimizer of the Tikhonov functional

$$\mathcal{L}(\theta) = \frac{1}{2}\|G\theta - Y\|_2^2 + \frac{1}{2n\tau_n^2}||C_0^{-1/2}\theta||_1^2. \tag{6}$$

Moreover, we see that the scale parameter τ_n of the prior becomes the regularisation parameter and that the estimator crucially depends on the choice of τ_n. This connection between the Bayesian inverse problem and Tikhonov regularisation has been extensively studied in [24]. The question we wish to answer is; can we choose τ_n depending on

Y, such that $\widehat{\theta}_{\tau_n}$ provides an adaptively optimal frequentist estimator for $\theta^\dagger$ and C_{τ_n} covers the frequentist uncertainty in the asymptotic limit $n \to \infty$. This paper positively answers this question. To choose the τ_n, we will use early stopping which is defined as follows: Suppose that for some given iterative method and for each $\tau_n \in \mathbb{R}_+ \cup \{0\}$ we have a sequence of estimators

$$(\widehat{\theta}_{\tau_n})_{\tau_n}$$

such that they minimize,

$$\widehat{\theta}_{\tau_n} = \text{argmin } \mathcal{L}_{(\tau_n)}(\theta)$$

and are ordered in decreasing bias and increasing variance. The goal of early stopping is to stop this iterative method exactly when the bias and variance of the estimator are balanced. An estimator for the asymptotic bias is given by the residuals, something that we can always measure. The residuals at time τ_n are given as

$$R_{\tau_n} := ||P_n(Y - G\widehat{\theta}_{\tau_n})||_2^2, \tag{7}$$

where P_n is an appropriate projection operator onto a finite dimensional subspace of H_2. The projection is necessary since the H_2-norm of Y is unbounded. To stop the iterative process we must define a stopping rule

$$\tau_{\text{dp}}(n) := \inf\{\tau_n > 0 : R_{\tau_n} \leqslant \kappa\} \tag{8}$$

for some threshold $\kappa > 0$ that is also a function of the data and is thus depending on n.

Remark 1.1. *We will further on drop the dependency on n to simplify the notation, but we remark that τ_{dp} is an estimator that depends on the data Y and is a function of n.*

Formulation (8) is often referred to as discrepancy principle [9, 3] in the case that the noise level is known and κ is chosen proportional to this noise level. Other choice for κ have been studied, see Subsection 1.2.

Definition 1.1. *(See Chapter 6 in [15]) A scale parameter τ_n and its associated estimator $\widehat{\theta}_{\tau_n}$ are called optimal if it achieves the minimax rate as $n \to \infty$ for a given Sobolev regularity of the unknown $\theta^\dagger$. We call the estimator adaptive if it does not require knowledge of the Sobolev regularity. A method is then adaptively optimal if it produces an estimator $\widehat{\theta}_{\tau_n}$ that is both adaptive and optimal.*

The challenge then is how to choose P_n and κ such that stopping according to (8) is adaptively optimal. In the following section, we summarise some of the previous contributions to solving this problem from a statistical perspective.

1.2. **Previous work on early stopping and ensemble Kalman inversion.** The works of Blanchard, Hoffman, and Reiß ([5]) and that of Blanchard and Mathé ([3]) study the discrepancy principle (8) in a frequentist setting. More precisely, in [3], they consider a variety of regularisation methods, such as Tikhonov, and consider the weighted residuals $||(G^*G)^{1/2}(Y - G\widehat{\theta}_{\tau_n})||_2^2$ instead of the projection operator P_n in (7). In the case that the true Sobolev regularity is known, this method can achieve optimal rates. However, no adaptive generalisation is currently available. In [5], they consider the discretized version of (1), with identity covariance matrix. The discretization is dependent on the estimation of the effective dimension, and is an integral part of their analogue of (8). In this regime, the method is adaptive for a small range of signals. In [25], the Bayesian perspective is studied, and they use an empirical Bayesian method which maximizes the log-likelihood for τ_n. In this setting they are able to achieve optimal rates as long as the prior is smooth enough. In this work, we instead extend the discrepancy principle (8) to Bayesian estimators of the form (3).

The ensemble Kalman filter (EnKF) has become a popular derivative-free method for approximating posterior distributions. Time-continuous formulations, the so-called ensemble Kalman-Bucy filter (EnKBF), have been first proposed in [1, 19, 2]. These formulations become again exact for linear Gaussian problems. A frequentist perspective on the EnKBF has already been explored in [22, 20]. Moreover in the early work [19], the issue of stopping time base on the discrepancy principle was posed as an open problem

The EnKF has also been utilised as a derivative-free optimisation method. These variants of the EnKF are often collected under the notion of ensemble Kalman inversion (EKI). EKI can be used for finding the minimizer of the cost functional (6) see [14, 7] and the review paper [6]. Algorithms for selecting the Tikhonov regularisation parameter within EKI have been discussed, for example, in [26]. Similarly, a discrepancy principle based stopping rule has been implemented for EKI in [13, 12]. Finally, [18], the paper closest to this work, examines choosing τ_n under the known noise regime, but does not provide a frequentist nor Bayesian analysis.

1.3. **Main contributions and paper outline.** In this paper, we study the stopping rule considered in [4] as an empirical method for choosing the scale parameter of the prior covariance. We provide a Bayesian analysis of this empirical method by extending the setting of [4] to the Bayesian setting of [16], and then provide an analysis of the asymptotic behaviour of the posterior stopped at time τ_{dp}. Furthermore, we derive for which ℓ^2-Sobelov regularities this method is optimal, and provide an adaptation interval for this method. We then reformulate the regularisation parameter as a time parameter, where we sequentially approach the final posterior by studying the time-continuous ensemble Kalman-Bucy filter. In the linear setting, the associated McKean-Vlasov evolution equations for mean and covariance are exact and provide an iterative process to transform the prior distribution.

This paper is structured as follows: In Section 2 we introduce the mathematical assumptions of our model. We provide the details of the projection matrix P_n and the stopping rule (8). We then formulate the filter function associated with the regularisation (6), and show that the results of [4] can be applied. We then show that the stopped posterior contracts optimally and provides a conservative measure of uncertainty. In Section 3 we formulate the Bayesian inverse problem of (1) in terms of the time-continuous ensemble Kalman-Bucy filter and introduce the associated McKean-Vlasov evolution equations. In this section, we formulate the regularisation parameter as a time parameter. We then show that taking a finite particle finite dimensional approximation of the posteriors leads to an error that is smaller than the statistical minimax error. In Section 4, we provide the discrete-time approximation of the process of the continuous time EnKBF and give the associated algorithm. The numerical results can then be found in Section 5. Finally, the conclusions can be found in Section 6.

1.4. **Further notation.** We define the following additional notation. For two numbers a and b we denote the minimum of a and b by $a \wedge b$. For two sequences $(a_n)_n$ and $(b_n)_n$ in $\mathbb{R}_+$, $a_n \lesssim b_n$, respectively $a_n \gtrsim b_n$ denote inequalities up to a multiplicative constant. $a_n \asymp b_n$ denotes that $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ holds. ℓ^2 denotes space of sequences in that are square summable, and ℓ^1 denotes the space of summable sequences both with indices running through $\mathbb{N}$. For a random variable Y denote the distribution of Y by $\mathbb{P}_Y$. The space of bounded linear operators mapping from H_1 to H_2 is denoted by $\mathcal{L}(H_1, H_2)$, with respective norm denoted by $\|\cdot\|_{\mathcal{L}(H_1,H_2)}$. For T a trace class operator with singular values $(a_i)_{i\in\mathbb{N}}$ the trace norm is $\|T\|_{\mathbb{T}} = \mathrm{Tr}(TT^*)^{1/2} = \sum_{i=1}^{\infty} a_i$. We can then view the class of trace class operators as sequences in ℓ^1 via their associated sequences of singular values.

2. Theoretical results on adaptive estimation

The results of this section use the fact that G is a linear compact operator. Then, by the Spectral Theorem, the eigenfunctions $(v_i)_{i\in\mathbb{N}}$, of G^*G form an orthonormal basis of H_1. Denote the bounded eigenvalues of $(G^*G)^{1/2}$ by,

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots > 0. \tag{9}$$

The following sequence space model is equivalent to observing (1), see [15], and is written as

$$Y_i = \sigma_i \theta_i^\dagger + \frac{1}{\sqrt{n}}\epsilon_i, \tag{10}$$

for $i \geqslant 1$, where $\theta_i^\dagger = \langle \theta^\dagger, v_i \rangle_1$ for $i \in \mathbb{N}$. Furthermore, all ϵ_i are i.i.d. $\mathcal{N}(0,1)$ with respect to the conjugate basis $(u_i)_{i\in\mathbb{N}}$ of the range of G in H_2 defined by

$$Gv_i = \sigma_i u_i \tag{11}$$

and $Y_i = \langle Y, u_i \rangle_2$.

In applications, it is necessary to truncate the infinite dimensional inverse problem to a finite dimensional one. Such a truncation plays a crucial role in previous studies on discrepancy based inference. See [4, 5, 23]. The truncation is then performed in sequence space (10) by setting all coefficients, i, larger than an appropriate dimension d_n to zero. That is, the required projection operator $P_n : H_2 \to H_2$ is defined in sequence space by

$$\langle P_n Y, u_i \rangle_2 = \begin{cases} Y_i & \text{if } i \leqslant d_n \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

The truncation is chosen to depend on the noise level, so when we consider the asymptotic limit $n \to \infty$ we have that $d_n \to \infty$. In this way, we are in the non-parametric setting, and in the no noise limit return to the infinite dimensional setting.

2.1. **Structural Assumptions.** We assume that the inverse problem is polynomially ill-posed where the degree of ill-posedness is given by some parameter $p > 0$. That is, the eigenvalues decay as

$$\sigma_i \asymp i^{-p}, \quad i \in \mathbb{N}. \tag{13}$$

We choose an entry-wise Gaussian prior over the coefficients of the variable of interest θ of the form, $\theta_i \overset{iid}{\sim} \mathcal{N}(0, \tau_n \lambda_i)$. For $\tau_n > 0$, we assume that the $\lambda_i's$ decay as

$$\lambda_i \asymp i^{-2\alpha-1}, \quad i \in \mathbb{N} \tag{14}$$

for $\alpha > 0$. We furthermore assume that the observations are generated given some true signal that lies in the Sobolev space S^β where β denotes the regularity of the signal. More specifically that

$$\theta^\dagger \in S^\beta := \{\theta \in H_1 : \|\theta\|_\beta^2 < \infty\} \tag{15}$$

where the norm is defined as

$$\theta = (\theta_i)_{i\in\mathbb{N}} \mapsto \|\theta\|_\beta^2 := \sum_i i^{2\beta}(\theta_i)^2. \tag{16}$$

Intuitively, these spaces consist of sequences of coefficients, $(\theta_i) \in \ell^2$ that decay faster to zero than the sequence $(i^{2\beta})$ for $i \in \mathbb{N}$.

With these assumptions, the posterior is denoted by Π_{n,τ_n} to indicate the dependence on both the n and the scale parameter τ_n. The entry-wise posterior is given as

$$\theta_i \mid Y_i \sim \mathcal{N}\left(\frac{n\tau_n^2\lambda_i\sigma_i Y_i}{1 + n\tau_n^2\lambda_i\sigma_i^2}, \frac{\tau_n^2\lambda_i}{1 + n\tau_n^2\lambda_i\sigma_i}\right). \tag{17}$$

The estimator (3) is then given in sequence space as

$$\widehat{\theta}_{i,\tau_n} = \frac{n\tau_n^2\lambda_i\sigma_i Y_i}{1 + n\tau_n^2\lambda_i\sigma_i^2}. \tag{18}$$

We consider signals in the Sobolev ellipsoid $S^\beta(r, d_n)$, which is defined as

$$\theta^\dagger \in S^\beta(r, d_n) := \{\theta \in \mathbb{R}^{d_n} : \sum_{i=1}^{d_n} i^{2\beta}\theta_i^2 < r\}. \tag{19}$$

We recall that the minimax rate of estimation over the unit ball in $S^\beta(1, \infty)$ is $n^{-\beta/(2\beta+2p+1)}$. The effective dimension of our observed signal is given by

$$d_{\text{eff}} \asymp n^{1/(2\beta+2p+1)}. \tag{20}$$

If we truncate the signal Y to P_nY with $d_n = d_{\text{eff}}$, then the approximation error we make is less than the minimax error. As we do not know β, the true smoothness of our signal, we choose $d_n \leqslant n$ to be at least as fine as

$$d_n \asymp n^{1/(2p+1)}. \tag{21}$$

This is an upper bound for d_{eff}, as $\beta \geqslant 0$. Our goal is now to recover the first d_n coefficients of the signal $\theta^\dagger$.

In deterministic inverse problems it is assumed that the noise level is known. Then it is possible to implement the discrepancy principle such that we stop at the first iteration when $R_{\tau_n} \lesssim d_n/n$.

Remark 2.1. *In [4], they consider the setting when $C_0 = I$, the dimension of the approximation space is d_n, with unknown noise δ in (1). They show that for*

$$\kappa \asymp d_n\delta^2 \tag{22}$$

stopping according to (8) is adaptively optimal. The theory holds for slight deviations of this choice of κ, mainly that κ can be chosen such that $|\kappa - d_n\delta^2| \leqslant c_n\sqrt{d_n}\delta$ [4] due to estimation of δ. We emphasize that the stopping rule depends on the truncation dimension d_n as also discussed in this paper. However, the choice $C_0 = I$ prevents an infinite-dimensional Bayesian interpretation of the estimators.

2.2. **Theoretical Results.** We first state the definition and assumptions of the class of spectral filters considered in [4].

Assumption 2.1. *Denote the regularisation function by $g(t, \sigma_i)$ where $g(t, \sigma_i) : \mathbb{R}_+ \times \mathbb{R}_+ \to [0, 1]$. Then the following must hold in order for $g(t, \sigma_i)$ to be in the class of regularising functions considered.*

(1) The function $g(t, \sigma)$ is non-decreasing in t and σ, continuous in t with $g(0, \sigma) = 0$ and $lim_{t\to\infty} g(t, \sigma) = 1$ for any fixed $\sigma > 0$.
(2) For all $t \geqslant t' \geqslant t_0$. the function $\lambda \mapsto \frac{1-g(t',\sigma)}{1-g(t,\sigma)}$ is non-decreasing.
(3) There exist positive constants α, β_-, β_+ such that for all $t \geqslant t_0$, and $\sigma \in [0, 1]$ we have that

$$\beta_- \min((t\sigma)^\rho, 1) \leqslant g(t, \sigma) \leqslant \min(\beta_+(t\sigma)^\rho, 1).$$

Lemma 2.1. *Let $(\sigma_i; v_i, u_i)$ be a singular system for G. We also assume that C_0 has the same eigenfunctions as G and further recall that the eigenvalues of C_0 are written as λ_i for $i = 1, ..., d_n$. Then the spectral filter for this generalised Tikhonov regularisation*

$$\widehat{\theta}_{\tau_n} \in \mathrm{argmin}||\mathrm{G}\theta_{\tau_\mathrm{n}} - \mathrm{Y}||^2 + \tau_\mathrm{n}^{-2}||\mathrm{C}_0^{-1/2}\theta_{\tau_\mathrm{n}}||^2 \tag{23}$$

is given by

$$g(\tau_n, \tilde{\sigma}) = \frac{1}{(1 + \tau_n{}^{-2}\lambda^{-1}\sigma^{-2})} \tag{24}$$

where $\tilde{\sigma} = (\lambda^{-1/2})\sigma$ and $n \geqslant 0$ $n \in \mathbb{R}$ and satisfies Assumption 2.1.

Proof. Denote the eigenvalues of G^TG by σ_i^2 and assume $C_0^{-1/2}$ has the same eigenfunctions of G^TG. Denote also the eigenvalues of $C_0^{-1/2}$ by $(\lambda^{-1/2})_i$. Then define

$$g_\alpha(\tilde{\sigma}) = \frac{1}{\sigma + \alpha(\lambda^{-1/2})^2} \tag{25}$$

Then

$$\begin{aligned}\widehat{\theta}_\alpha &= \sum_{i=1}^{\infty} \frac{1}{\sigma_i^2 + \alpha\left(\lambda_i^{-1/2}\right)^2} \langle G^T y, v_n \rangle v_n \\ &= \sum_{i=1}^{\infty} \frac{1}{\sigma_i^2 + \alpha\left(\lambda_i^{-1/2}\right)^2} \langle y, G v_n \rangle v_n \\ &= \sum_{i=1}^{\infty} \frac{1}{\sigma_i^2 + \alpha\left(\lambda_i^{-1/2}\right)^2} \langle y, \sigma_i u_n \rangle v_n \\ &= \sum_{i=1}^{\infty} \frac{\sigma_i}{\sigma_i^2 + \alpha\left(\lambda_i^{-1/2}\right)^2} \langle y, u_n \rangle v_n.\end{aligned}$$

With $g_\alpha(\tilde{\sigma}) = \tilde{\sigma}^2 g_{\tau_n{}^{-2}}(\tilde{\sigma}^2) = g(\tau_n, \tilde{\sigma})$ it follows that

$$g(\tau_n, \tilde{\sigma}) = \frac{1}{(1 + \tau_n{}^{-2}(\lambda^{-1/2})^2 \sigma^{-2})}. \tag{26}$$

Then, $\widehat{\theta}_{\tau_n}$ can be written as

$$\widehat{\theta}_{\tau_n} = (G^T G + \tau_n{}^{-2}(C_0^{-1/2})^* C_0^{-1/2})^{-1} G^T y \tag{27}$$

which is the unique minimiser of

$$||G\theta - y||^2 + \tau_n{}^{-2}||C_0^{-1/2}\theta||^2. \tag{28}$$

This can be seen by writing the normal equation for θ. Note that $\lambda_i^{-1/2} \geqslant 1$, for $i \in \mathbb{N}$. For, $\sigma \in (0, 1]$ Assumption (2.1) is satisfied with $\rho = 2$, $\beta_- = 0$. For the Tikhonov regularisation with $C_0^{-1/2} = I$, the spectral filter is $g(\tau_n, \sigma) = \frac{1}{(1+\tau_n^{-2}\sigma^{-2})}$, and β_+ was computed to be 1. We also have that $\frac{1}{(1+\tau_n^{-2}\lambda^{-1}\sigma^{-2})} \leqslant \frac{1}{(1+\tau_n^{-2}\sigma^{-2})}$ so $\beta_+ = 1$. □

We then have that the Bayesian estimator as given by the mean of the stopped posterior achieves the frequentist minimax rate. We first state the following definition:

Definition 2.1. *A sequence $(\epsilon_n)_n$ of positive numbers is a posterior contraction rate at the parameter $\theta^\dagger$ wrt to the semi-metric d if for every sequence $(M_n)_n \to \infty$, it holds that,*

$$\Pi_n(\theta : d(\theta, \theta^\dagger) \geqslant M_n \epsilon_n \mid Y) \xrightarrow{P_{\theta^\dagger}} 0 \tag{29}$$

as $n \to \infty$. Where $\Pi_n(\cdot \mid Y)$ is the posterior given observations Y and given prior Π_n.

Intuitively, $(\epsilon_n)_n$ is the rate at which the d_n ball of radius $M_n \epsilon_n$ decreases such that "most" of the posterior mass is inside this ball.

Theorem 2.1. *Provided that $\kappa_n \asymp d_n/n$ in (8), the truncation dimension d_n in (7) satisfies (21), and*

$$\beta < 2\alpha + 2p + 1, \tag{30}$$

it follows that

$$\sup_{\theta^\dagger \in S^\beta(1)} \mathbb{E}\|\widehat{\theta}_{\tau_{\mathrm{dp}}} - \theta^\dagger\|^2 \lesssim n^{-\beta/(2\beta+2p+1)} \tag{31}$$

in the limit $1/n \to 0$*, where* $\widehat{\theta}_{\tau_{\mathrm{dp}}}$ *is the estimator (3).*

Proof. We have that $\mathbb{E}_{\theta^\dagger}\Pi_n(\theta : d(\theta,\theta^\dagger) \geqslant M_n\epsilon_n \mid Y) \xrightarrow{P_{\theta^\dagger}} 0$, for every $M_n \to \infty$ where ϵ is computed as

$$\epsilon_n = (n\tau_n^2)^{-\beta/(1+2\alpha+2p)\wedge 1} + (n\tau_n^2)^{-\alpha/(1+2\alpha+2p)}$$

Then from Theorem 4.1 and proof of in [17] the rate can be split into the following cases:

- If $\tau_n \equiv 1$ the rate is $\epsilon_n = n^{(\alpha\wedge\beta)/(1+2\alpha+2p)}$
- if $\beta \leqslant 1+2\alpha+2p$, and $\tau_n \asymp n^{(\alpha-\beta)/(1+2\beta+2p)}$, then $\epsilon_n = n^{(-\beta)/(1+2\beta+2p)}$
- if $\beta > 1+2\alpha+2p$, no matter the scaling of τ_n we do not get an optimal rate as $\epsilon_n \gg n^{(-\beta)/(1+2\beta+2p)}$

Note when $\beta < 1+2p+2\alpha$, there exists an oracle choice for τ_n which is

$$\tau_n^* \asymp n^{(\alpha-\beta)/(2\beta+2p+1)} \tag{32}$$

such that the minimax rate is achieved. In the case where $\beta \geqslant 1+2p+2\alpha$, the minimal τ_n, such that the true bias equals the variance, returns a rate larger than $n^{-\beta/(2\beta+2p+1)}$. Moreover ϵ_n consists of three parts the order of the frequentist basis part, the order of the frequentist variance part, and the order of the posterior spread. The order of the posterior speed dominates that of the frequentist variance, so only two terms are necessary to consider. From this decomposition of ϵ_n, we can see that the smoothness condition on the Bayesian setting is necessary also in the frequentist setting. Then as the filter function (24) satisfies Assumptions 2.1, the rate estimate (31) follows by applying Corollary 3.7 in [4]. □

Corollary 2.1. *If* $\kappa \asymp d_n/n$*, then* $\mathbb{P}_Y(\tau_{\mathrm{lo}} \leqslant \tau_{\mathrm{dp}} \leqslant \tau_{\mathrm{up}}) \to 1$*, where* $\tau_{\mathrm{lo}}, \tau_{\mathrm{up}} \asymp n^{(\alpha-\beta)/1+2\beta+2p}$*, and the adaptation interval is* $[0,\beta_+]$*, such that* $\beta \leqslant \beta_+ \leqslant 1+2p+2\alpha$*.*

Proof. We can further conclude by Theorem 2.1 that $\mathbb{P}_Y(\tau_{\mathrm{dp}} \leqslant \tau_{\mathrm{up}}) \to 1$, as the expected loss of $\widehat{\theta}_{\tau_{dp}}$ is upper bounded by the expected loss of $\widehat{\theta}_{\tau_n^*}$ and the expected ℓ^2 loss is strictly convex. If $\kappa \gtrsim d_n/n$, then $\mathbb{P}_Y(\tau_{\mathrm{dp}} \geqslant \tau_{\mathrm{lo}}) \to 1$ as asympototically $R_{\tau_{\mathrm{dp}}} \gtrsim \kappa$ and by the monotonicity of the bias. We further conclude that the adaptation interval, is $[\beta_-,\beta_+]$ where $\beta_- = 0$, as the estimator for the effective dimension was chosen as $\rho^{-1/(2p+1)}$ impying $\beta = 0$. We then have, that $\beta_+ = 1+2p+2\alpha$, as only in this case can the prior for fixed α be rescaled and achieve optimal rates see proof of Theorem 4.1 in [17]. □

We will further show that as the truncation dimension $d_n \to \infty$, the posterior dependent on τ_{dp} also contracts at the optimal rate. We use the results from Theorem 2.1, and only need to consider the behaviour of the tails knowing that the mean of the posterior is moving towards the truth at the optimal rate.

Theorem 2.2. *Let* G^*G *and* C_0 *have the same eigenfunctions. Denote the eigenvalues of* G^*G *and* C_0 *by* σ_i^2 *and,* λ_i *respectively. Recall the structural assumptions,*

$$\sigma_i \asymp i^{-p} \tag{33}$$

and entry-wise prior

$$\theta_i \sim \mathcal{N}(0, \tau_n^2 \lambda_i) \quad \lambda_i \asymp i^{-1-2\alpha} \tag{34}$$

Let $\theta^\dagger \in S^\beta$, and denote the posterior associated to the estimated stopping time τ_{dp} by $\Pi_{n,\tau_{\mathrm{dp}}}(\cdot \mid Y)$. Then

$$\mathbb{E}_{\theta^\dagger} \Pi_{n,\tau_{\mathrm{dp}}} \left(\widehat{\theta}_{\tau_{\mathrm{dp}}} : ||\widehat{\theta}_{\tau_{\mathrm{dp}}} - \theta^\dagger|| \geqslant M_n \epsilon_n \right) \to 0 \tag{35}$$

for every $M_n \to \infty$, and with $\epsilon_n = n^{-\beta/(2\beta+2p+1)}$.

Proof. The proof follows accordingly:

- Find and interval I such that $\tau_{\mathrm{dp}} \in I$ a.s. where τ_{dp} is estimated from the data (8).
- Then find ϵ_n such that

$$\sup_{\tau_n \in I} ||\widehat{\theta}_{\tau_n} - \theta^\dagger|| \leqslant \epsilon_n \tag{36}$$

  as $d_n \to \infty$.
- Show that $\epsilon_n^2 \inf_{\tau_n \in I} ||C_{\tau_n}||_{op}^{-1} \to \infty$.
- Then $\Pi_{n,\tau_{\mathrm{dp}}}(\cdot \mid Y)$ contracts at rate ϵ_n

We have that

$$\sup_{\tau_n \in I} ||\widehat{\theta}_{\tau_n} - \theta^\dagger|| \leqslant \epsilon_n \tag{37}$$

for $\epsilon_n = n^{-\beta/(2\beta+2p+1)}$ and $I = [\tau_{\mathrm{lo}}, \tau_{\mathrm{up}}]$, where $\tau_{\mathrm{lo}}, \tau_{\mathrm{up}}$ are of order $n^{(\alpha-\beta)/(2\beta+2p+1)}$. From Theorem (2.1) and Corollary 2.1 when $d_n \to \infty$ as $1/n \to 0$ we have that $\mathbb{P}_Y(\tau_{\mathrm{dp}} \in I) \to 1$. Furthermore, we know that the posterior spread for each time τ_n is given by

$$C_{\tau_n} := \tau_n^2 C_0 - \tau_n^2 K_{\tau_n} \left(GC_0G^* + \frac{1}{n\tau_n^2} I \right) K_{\tau_n}^* \tag{38}$$

So then

$$||C_{\tau_n}||_{op}^{-1} \lesssim \frac{1 + n\tau_n^2}{\tau_n^2} \tag{39}$$

as the maximum eigenvalues of G and C_0 are 1. Then

$$\epsilon_n^2 \inf_{\tau_n \in I} ||C_{\tau_n}||_{op}^{-1} \to \infty \tag{40}$$

So

$$\lim_{n\to\infty} \Pi_{n,\tau_n}(||\widehat{\theta}_{\tau_n} - \theta^\dagger||_{H_1}^2 \geqslant \epsilon_n) = 0 \tag{41}$$

So then

$$\sup_{\tau_n \in I} \Pi_{n,\tau_n}(||\widehat{\theta}_{\tau_n} - \theta^\dagger||_{H_1}^2 \geqslant \epsilon_n) = o_{\mathbb{P}_\theta^\dagger}(1) \tag{42}$$

where $\mathbb{P}_\theta^\dagger$ is the true distribution of $\theta^\dagger$ □

The next question we would like to address is if the spread of the stopped posterior $\Pi_{n,\tau_{\mathrm{dp}}}$ has good frequentist coverage. First, we define a credible ball.

Definition 2.2. *Denote the mean of the posterior by KY. Then the credible ball centred at KY is defined as*

$$KY + B(r_{n,c}) := \{\theta \in H_1 : ||\theta - KY||_{H_1} < r_{n,c}\} \tag{43}$$

where $B(r_{n,c})$ is the ball centred at 0 with radius $r_{n,c}$, $c \in (0,1)$ denotes the desired credible level of $1-c$. The $r_{n,c}$ is chosen such that

$$\Pi_{n,\tau_n}(KY + B(r_{n,c}) \mid Y) = 1 - c. \tag{44}$$

We can define the frequentist coverage of the set $KY + B(r_{n,c})$ as

$$\Pi_{\theta^\dagger}(\theta^\dagger \in KY + B(r_{n,c})) \tag{45}$$

Corollary 2.2. *For fixed $\alpha > 0$, if $\theta_i^\dagger = C_i i^{-1-2\beta}$ for all $i = 1, ..., d_n$, then as $n \to \infty$, $\Pi_{n,\tau_{\mathrm{dp}}}$ asymptotically has frequentist coverage 1.*

Proof. It was assumed that $\beta < 1 + 2p + 2\alpha$, and if $\kappa \asymp d_n/n$, then $\tau_{\mathrm{dp}} \gtrsim n^{(\alpha-\beta)/1+2p+2\alpha}$ see Corollary 2.1, so the result follows from by Theorem 4.2 in [17]. □

Moreover, it can be seen from conditions $(1, 2)$ of Theorem 4.2, where we view τ_n as the time parameter, that it is important to not stop prematurely to have the desired frequentist coverage, and the lower bound of τ_{dp} is also of order $n^{(\alpha-\beta)/(1+2p+2\beta)}$.

3. Ensemble Kalman–Bucy inversion

In this section, we reformulate the Bayesian inverse problem defined by (1) and the Gaussian process prior $\mathcal{N}(0, \tau_n C_0)$ in terms of the time-continuous ensemble Kalman–Bucy filter [19, 8, 6]. In order to do so, we introduce a new time-like variable $t \geqslant 0$ and a H_1-valued and time-dependent Gaussian process denoted by Θ_t. This process satisfies the McKean–Vlasov evolution equation

$$\frac{\mathrm{d}\Theta_t}{\mathrm{d}t} = n\Sigma_t G^* \left\{ Y - \frac{1}{2} G\left(\Theta_t + \widetilde{\theta}_t\right)\right\}, \qquad \Theta_0 \sim \mathcal{N}(0, C_0), \tag{46}$$

where the mean and covariance operator of Θ_t are denoted by $\widetilde{\theta}_t$ and Σ_t, respectively.

Despite the fact, that the evolution equation (46) are nonlinear, closed form solutions in terms of the mean and the covariance operator are available and given by

$$\widetilde{\theta}_t = C_0 G^* \left(GC_0G^* + \frac{1}{n\,t} I\right)^{-1} Y, \tag{47a}$$

$$\Sigma_t = C_0 - C_0 G^* \left(GC_0G^* + \frac{1}{n\,t} I\right)^{-1} GC_0. \tag{47b}$$

Comparison to the estimator (3) reveals that

$$\widetilde{\theta}_t = \widehat{\theta}_{\tau_n} \tag{48}$$

for $t = \tau_n^2$. Furthermore, under the same relation between t and τ_n, it also holds that

$$C_{\tau_n} = t\Sigma_t. \tag{49}$$

Hence, upon defining d_n, P_n, and κ as before, the evolution equations (46) are integrated in time until

$$t_{\mathrm{dp}} := \inf \left\{ t > 0 : \|P_n(Y - G\widetilde{\theta}_t)\|_2^2 \leqslant \kappa \right\}. \tag{50}$$

We next propose an adapted ensemble implementation of the mean-field EnKBF formulation (46). Such an idea has been proposed in [18], where the ensemble size is grown with each iteration. We will keep our ensemble sized fixed throughout the iterations. First note that by the Schmidt–Eckhardt–Young–Mirsky theorem for deterministic low rank approximation of self adjoint trace class operators truncating the SVD of C_0 at J gives us the approximation error of

$$\|V_{J-1}V_{J-1}^* - C_0\|_{\mathcal{L}(H_1,H_1)} = J^{-2\alpha-1}. \tag{51}$$

where $V_{J-1}V_{J-1}^*$ is the truncated SVD of C_0. Let us generate J independent H_1-valued random ensemble members $\Theta_0^{(i)}$, $i = 1, \ldots, J$ such that their mean is zero and the covariance of each member is C_0. Denote their empirical mean by

$$\widetilde{\theta}_0^J = \frac{1}{J}\sum_{i=1}^{J} \Theta_0^{(i)} \tag{52}$$

and their empirical covariance operator by

$$\Sigma_0^J = \frac{1}{J}\sum_{i=1}^{J} (\Theta_0^{(i)} - \widetilde{\theta}_0^J) \otimes (\Theta_0^{(i)} - \widetilde{\theta}_0^J). \tag{53}$$

The optimal approximation error of the empirical covariance is given by

$$\|\Sigma_0^J - C_0\|_{\mathcal{L}(H_1,H_1)} \asymp J^{-2\alpha-1}. \tag{54}$$

We can achieve this error by using a truncated singular value decomposition of C_0 and using this in place of the empirical approximation of the true covariance. Alternatively, the ensemble can be constructed using the Nyström method, in which case (54) holds in expectation only. See [18] for proof that this method has approximation error $J^{-2\alpha-1}$. If we set

$$J = d_n + 1 \tag{55}$$

the approximation error (54) becomes smaller than the minimax error provided (30) is replaced by the smoothing condition

$$\beta \leqslant 2\alpha + 1 \tag{56}$$

on the prior Gaussian distribution. Note that (56) implies (30) since $p > 0$.

Corollary 3.1. *If $\tilde{\theta}_0^J = \widehat{\theta}_0$, and for $J = d_n + 1$, then $\|\tilde{\theta}_t^J - \widehat{\theta}_t\|_1 \lesssim J^{-2\alpha-1}$ for all $t > 0$.*

Proof. For t fixed, we have that

$$\|\tilde{\theta}_t^J - \widehat{\theta}_t\|_1 = \left\| \left[\Sigma_0^J G^* \left(G\Sigma_0^J G^* + \frac{1}{n\,t} I \right)^{-1} Y \right] - \left[C_0 G^* \left(G C_0 G^* + \frac{1}{n\,t} I \right)^{-1} Y \right] \right\|_1 \tag{57}$$

Then, let

$$K^J = \Sigma_0^J G^* \left(G\Sigma_0^J G^* + \frac{1}{n\,t} I \right)^{-1}, \tag{58}$$

$$K = C_0 G^* \left(G C_0 G^* + \frac{1}{n\,t} I \right)^{-1}. \tag{59}$$

We then have that

$$\|K^J - K\|_{\mathcal{L}(H_1,H_1)} \asymp J^{-2\alpha-1} \tag{60}$$

by linearity of the norm and the assumption on the approximation error (54). It follows that

$$\|\tilde{\theta}_t^J - \widehat{\theta}_t\|_1 = \|Y(K^J - K)\|_1 \tag{61}$$

$$\lesssim \|Y\|_2 \|(K^J - K)\|_{\mathcal{L}(H_1,H_1)} \tag{62}$$

$$\lesssim J^{-2\alpha-1}. \tag{63}$$

Plugging in $J = d_n + 1$ we get that $J^{-2\alpha-1} = n^{-2\alpha-1/2p+1} + 1 \leqslant n^{\beta/(2\alpha+2p+1)}$ when (56) is satisfied. □

The mean-field EnKBF formulation (46) is now replaced by interacting finite particle EnKBF formulation

$$\frac{\mathrm{d}\Theta_t^{(i)}}{\mathrm{d}t} = n\Sigma_t^J G^* \left\{ Y - \frac{1}{2} G \left(\Theta_t^{(i)} + \widetilde{\theta}_t^J \right) \right\} \tag{64}$$

for $i = 1, \ldots, J$.

Corollary 3.2. *Denote singular value decomposition of C_0 truncated at the J^{th} singular value by $T^J(T^J)^* = C^0$, where $T^J : \mathbb{R}^J \to H_1$ is a linear operator. By truncated, we mean that $T^J(T^J)^* = P_J(T^\infty(T^\infty)^*)$ where P_J is defined in* (12)*. Let Q_J be the projector of H_1 onto the $\mathbb{R}^J$ subspace of H_1, that maps the first J coordinates to $\mathbb{R}^J$. Then, for all $t > 0$ we have that*

$$\|Q_J \tilde{\theta^J}_t - \widehat{\theta}_t\| \lesssim J^{-2\alpha-1}. \tag{65}$$

Proof. Fix t. Then we have that

$$\|Q_J \tilde{\theta}_t^J - \tilde{\theta}_t^J\| = 0$$

□

Again plugging in $J = d_n + 1$ we get that $J^{-2\alpha-1} = n^{-2\alpha-1/2p+1} + 1 \leqslant n^{\beta/(2\alpha+2p+1)}$ when (56) is satisfied. So using a finite dimension finite ensemble size does not add error of order higher than the minimax rate. We can then consider the projected version of (64) which is given as

$$\frac{\mathrm{d}\Theta_t^{(i)}}{\mathrm{d}t} = nQ_J(\Sigma_t^J G^*) \left\{ Q_J(Y - G\widetilde{\theta}_t^J) - \frac{1}{2} Q_J(G\left(\Theta_t^{(i)} - \widetilde{\theta}_t^J \right)) \right\}. \tag{66}$$

The stopping criterion (50) can be replaced by using the projected residuals

$$t_n := \inf \left\{ t > 0 : \|Q_J(Y - G\widetilde{\theta}_t^J)\|_2^2 \leqslant \kappa \right\}. \tag{67}$$

It should be noted that (66) effectively constitutes a finite-dimensional system of ordinary differential equations since

$$\Theta_t^{(j)} = \sum_{i=1}^{J} \Theta_0^{(i)} m_t^{(i,j)} \tag{68}$$

for appropriate time-dependent coefficients $m_t^{(i,j)}$, $i, j = 1, \ldots, J$ [21].

4. Algorithmic details

We now give the discrete time formulation for implementing (66) which we rewrite as

$$\frac{\mathrm{d}}{\mathrm{d}t}\Theta_t = -\frac{1}{2} n \Sigma_t^J G^{\mathrm{T}} \left(G\Theta_t + Gm_t - 2Y \right). \tag{69a}$$

with the understanding that here everything has been projected onto $\mathbb{R}^J$. Denote the ensemble of J parameter values at time $t_k \geqslant 0$ by $\theta_k^{(i)}$. The initial conditions are given by

$$\theta_0^{(i)} \sim \mathrm{N}(0, C_0) \tag{70}$$

for $i = 1, \ldots, J$. Let us denote the empirical mean of the ensemble by m_k^J and the empirical covariance matrices by C_k^J. We introduce the empirical covariance matrix between Θ and $G\Theta$, denoted by $\mathcal{C}_n^J \in \mathbb{R}^{d_n \times d_n}$, as well as the empirical covariance matrix of $G\Theta$, denoted by $S_k^J \in \mathbb{R}^{d_n \times d_n}$, which is given by

$$S_k^J = \frac{1}{J-1} \sum_{i=1}^{J} (G\theta_k^{(i)} - G\tilde{\theta}_k^J)(G\tilde{\theta}_k^J - G\theta_k^{(i)})^{\mathrm{T}}, \tag{71}$$

where $G\tilde{\theta}_k^J$ denotes the empirical mean of $G\theta$. Similarly,

$$\mathcal{C}_k^J = \frac{1}{J-1} \sum_{i=1}^{J} (\theta_k^{(i)} - \tilde{\theta}_k^J)(G\theta_k^{(i)}) - G\tilde{\theta}_k^J)^{\mathrm{T}}. \tag{72}$$

The resulting deterministic discrete time update formulas, which follow from (69), are given as

$$\theta_{k+1}^{(i)} = \theta_k^{(i)} - \frac{1}{2} K_k \left(G\theta_k^{(i)} + G\tilde{\theta}_k^J - 2Y \right) \tag{73a}$$

with Kalman gain matrix

$$K_k = \Delta t \, \mathcal{C}_k^J \left(\Delta t S_k^J + I \right)^{-1}. \tag{74}$$

The standard discrepancy principle, stops the iteration of the EnKBF whenever

$$k_{\mathrm{dp}} = \inf \left\{ k \geqslant k_0 : \| G(\tilde{\theta}_k^J) - Y \|^2 \leqslant \kappa \right\}. \tag{75}$$

where $\kappa = Cd_n/n$ and $0 < C \leqslant 1$ is such that $|\kappa - d_n\rho| \leqslant C\sqrt{d_n}/n$ and k_0 is the initial time. See Section 2. Combining the deterministic updates, the recalibration of the initial ensemble, and the discrepancy based stopping criterion, we have the resulting algorithm.

Algorithm 1 Deterministic EnKF

Require: $J > 0,\ m_0,\ \tilde{\Delta}_0,\ Y,\ G,\ \rho$

$\Theta_0 \leftarrow \mathtt{initialize}(J, m_0, \tilde{\Delta}_0)$ ▷ $\Theta \in \mathbb{R}^{d_n \times J}$

$\Theta_0 \leftarrow \mathtt{calibrate}(J, m_0, \Theta_0)$ ▷ *see* (54)

$R_0 \leftarrow ||G(\tilde{\theta}_0^J) - Y||^2$

$\kappa_{\mathrm{dp}} = d_n \delta^2$ ▷ *see* (22)

while $R_k < \kappa_{\mathrm{dp}}$ ***do***

$K_k \leftarrow \Delta t\, \mathcal{C}_k^J \left(\Delta t S_k^J + I\right)^{-1}$ ▷ $\mathcal{C}_k^J$ *see* (72), Σ_k^J *see* (71)

for $i \in \{1, .., J\}$ ***do***

$$\theta_{k+1}^{(i)} \leftarrow \theta_k^{(i)} - \frac{1}{2} K_k \left(G\theta_k^{(i)} + Gm_k^J - 2Y\right)$$

$R_{k+1} \leftarrow ||G\left(\tilde{\theta}_{k+1}^J\right) - Y||^2$

Return Θ_k

5. Numerical Examples

In this section, we demonstrate the performance of the discrepancy principle based stopping rule. We choose the forward operator to have $p = 1/2$. We choose $\theta^\dagger \in S^\beta := \{\theta \in H_1 : \sum_{i=1}^\infty \theta_i^2 i^{2\beta} < \infty\}, \quad \beta > 0$. We initialize the ensemble with entry wise prior $\theta_i \overset{ind}{\sim} \mathcal{N}(0, i^{-1-2\alpha})$, where $\alpha = 1/2$. We generate noisy observation by setting the dimension of $\theta^\dagger$ to 100 and draw independent standard Gaussian noise with noise level $1/n = 0.001$. We set $\kappa_{\mathrm{dp}} = Cd_n/n$, and run the ensemble Kalman filter for 1000 iterations with $\Delta t_n = 0.04$. For each iteration, we compute the residual and record the first iteration such that $R_n^2 \leqslant \kappa_{\mathrm{dp}}$. We consider three different, $\theta^\dagger$ which are shown in Figure 1 where $\theta_{1,i} = 5\sin(0.5i)i^{-1}, \theta_{2,i} = 5\sin(0.5i)i^{-3/2}, \theta_{3,i} = 5\exp(-i)$ for $i = 1, ..., 100$. We

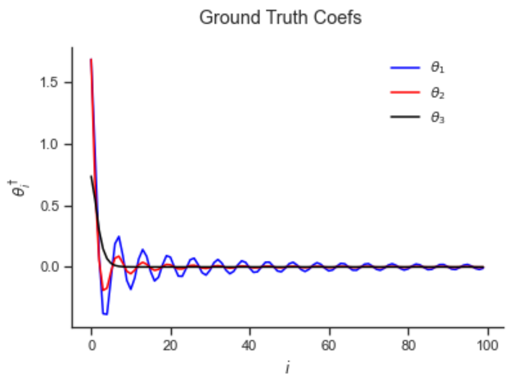

Fig. 1. Plot of ground truth coefficients with differing decay.

show our results using the deterministic algorithm. We chose the ensemble size to be 90, which was required to estimate the posterior spread correctly. We note that a smaller ensemble size can work in practice if one is interested in the mean alone. The results are based on 100 simulations. In Figure 2, we see that the posterior, given our choice of α, p, κ, associated with the stopped ensemble has the correct coverage.

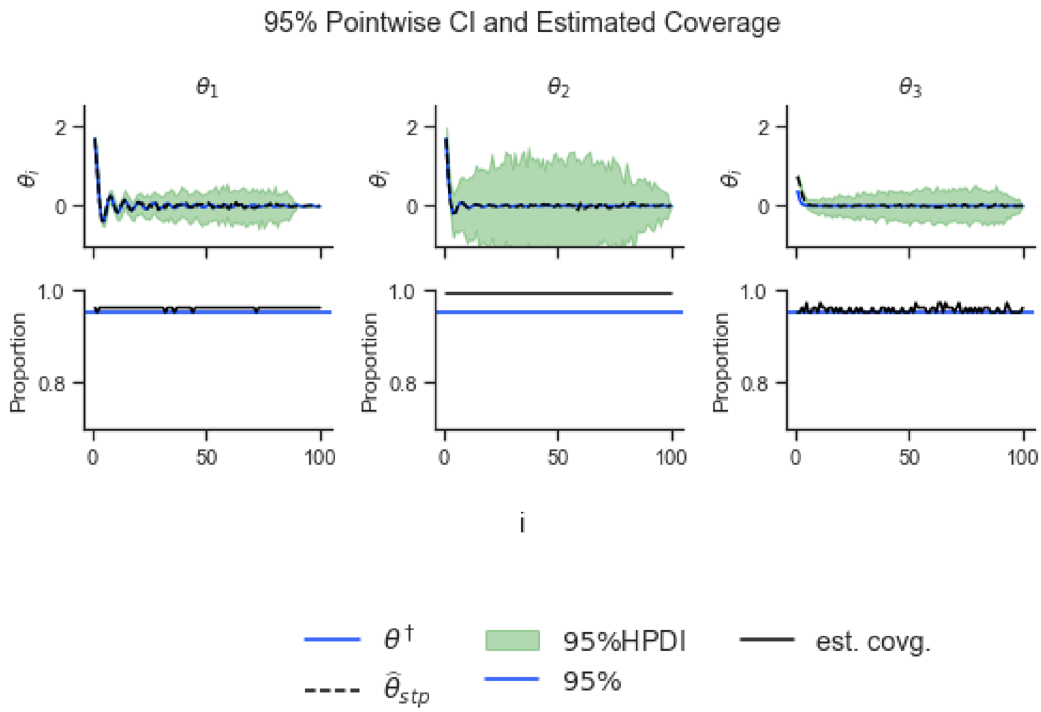

Fig. 2. Plot of MAP estimator versus ground truth and the respective 95% credible intervals and frequentist coverage resulting from the stopped ensemble using the discrepancy principle. The estimated τ_{dp} was, $32.48, 35.52, 40$ respectively.

To run the algorithm, one must choose α, and additionally, choose κ, which depends on estimating the noise level. From Section 2 one should choose a relatively smooth prior so that it can adapt to the true smoothness. Additionally, one should choose a small κ, to mitigate against stopping too early, as the lower bound on τ_{dp} should be satisfied in order to have the appropriate posterior variance.

6. Conclusion

Prior selection is an important decision to make in Bayesian methods. We have seen that the scale parameter determines the amount of regularisation the prior introduces into the problem. We have also seen in the analysis in Section 2 that the scale parameter strongly influences the posterior variance. Therefore, in order to derive a posterior that has the desired properties, the correct choice of scale parameter is critical. We have considered an adaptive empirical method for choosing the scale parameter of the covariance of a Gaussian prior using early stopping. It was shown in the linear case that the link between the choice of how much regularisation is needed and the choice of the prior is, in effect, the choice of how much to scale the covariance operator. The stopping rule was dependent on the dimension of the discretisation, which is an upper bound on the effective dimension of the observations. In addition, the asymptotic behaviour of the stopped Bayesian estimator and the associated stopped posterior were analysed. It was explicitly shown that the stopped estimator is minimax optimal and that the posterior contracts at the optimal rate. We also showed that the stopped posterior's frequentist coverage tends to 1. We showed that this method is adaptive if the prior smoothness α is chosen to be at such that $\beta < 1 + 2p + 2\alpha$ holds. Finally, we tested our method numerically using the EnKF. This algorithm allowed us to reformulate our problem in terms of an iterative process of the evolving Gaussian distributions.

The results of this paper depend on the structural assumptions made in Section 2. In particular, we point out that the proofs depend on the linearity of the inverse problem. The ensemble Kalman filter introduced in Section 3 is an exact method only in the linear setting, but ensemble Kalman methods have achieved success in cases where the forward operator is non-linear and a Gaussian approximation of the posterior is appropriate. Early stopping, as a separate field, has considered nonlinear inverse problems in the deterministic setting, where hence no adaptation can occur. For these reasons, an extension of these methods to the non-linear setting is of interest. Another open question is whether there are stopping rules that do not depend on the effective dimension and are adaptively optimal in the Bayesian setting.

Acknowledgements. This work has been funded by Deutsche Forschungsgemeinschaft (DFG) - Project-ID 318763901 - SFB1294. MT thanks Prof. Sebastian Reich for his many inputs and feedback. MT also thanks Jakob Walter for his many revisions and feedback.

References

[1] K. Bergemann and S. Reich. A localization technique for ensemble Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, 136:701–707, 2010. doi: 10.1002/qj.591.

[2] K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation. *Meteorol. Z.*, 21:213–219, 2012. doi: 10.1127/0941-2948/2012/0307.

[3] G. Blanchard and P. Mathé. Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems*, 28, 2012. doi: 10.1088/0266-5611/28/11/115011.

[4] G. Blanchard, M. Hoffmann, and M. Reiß. Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 6(3):1043–1075, 2018. doi: 10.1137/17M1154096.

[5] G. Blanchard, M. Hoffmann, and M. Reiß. Early stopping for statistical inverse problems via truncated SVD estimation. *Electr. J. Stat.*, 12(2):3204–3231, 2018. doi: 10.1214/18-EJS1482.

[6] E. Calvello, S. Reich, and A. M. Stuart. Ensemble Kalman methods: A mean-field perspective. *arXiv:2209.11371*, 2022.

[7] N. K. Chada, A. M. Stuart, and X. T. Tong. Tikhonov regularization within ensemble Kalman inversion. *SIAM Journal on Numerical Analysis*, 58(2):1263–1294, 2020.

[8] C. Cotter and S. Reich. Ensemble filter techniques for intermittent data assimilation. *Radon Ser. Comput. Appl. Math.*, 13:91–134, 2013. doi: 10.1515/9783110282269.91.

[9] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[10] S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge, 2017.

[11] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, Cambridge, 2016. doi: 10.1017/CBO9781107337862.

[12] M. Iglesias and Y. Yang. Adaptive regularisation for ensemble Kalman inversion with applications to non-destructive testing and imaging. *arXiv preprint arXiv:2006.14980*, 2020.

[13] M. A. Iglesias. A regularizing iterative ensemble Kalman method for pde-constrained inverse problems. *Inverse Problems*, 32(2):025002, 2016.

[14] M. A. Iglesias, K. J. Law, and A. M. Stuart. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

[15] I. M. Johnstone. *Gaussian estimation: Sequence and wavelet models, draft of a monograph*. 2017.

[16] B. Knapik, B. Szabó, A. van der Vaart, and H. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Relat. Fields*, 164:771–813, 2016. doi: 10.1007/s00440-015-0619-7.

[17] B. T. Knapik, A. W. van der Vaart, and J. H. van Zanten. Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.

[18] F. Parzer and O. Scherzer. On convergence rates of adaptive ensemble kalman inversion for linear ill-posed problems. *Numerical Mathematics*, 158(2):371–409,

oct 2022. doi: 10.1007/s00211-022-01314-y. URL `https://doi.org/10.1007/s00211-022-01314-y`.

[19] S. Reich. A dynamical systems framework for intermittent data assimilation. *BIT*, 51(1):235–249, 2011. doi: 10.1007/s10543-010-0302-4.

[20] S. Reich. Frequentist perspective on robust parameter estimation using the ensemble Kalman filter, 2022. arXiv:2201.00611.

[21] S. Reich and C. Cotter. *Probabilistic forecasting and Bayesian data assimilation*. Cambridge University Press, New York, 2015. ISBN 978-1-107-66391-6; 978-1-107-06939-8. doi: 10.1017/CBO9781107706804.

[22] S. Reich and P. Rozdeba. Posterior contraction rates for non-parametric state and drift estimation. *Foundation of Data Science*, 2:333–349, 2020. doi: 10.3934/fods.2020016.

[23] B. Stankewitz. Smoothed residual stopping for statistical inverse problems via truncated svd estimation. *Electronic Journal of Statistics*, 14(2):3396–3428, 2020. doi: 10.1214/20EJS1747.

[24] A. Stuart. Inverse problems: A bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061. URL `http://journals.cambridge.org/abstract_S0962492910000061`.

[25] B. Szabo, A. van der Vaart, and H. van Zanten. Empirical bayes scaling of gaussian priors in the white noise model. *Electron. J. Statist*, 7:991 – 101, 2013. doi: 10.1214/13-EJS798. URL `https://doi.org/10.1214/13-EJS798`.

[26] S. Weissmann, N. K. Chada, C. Schillings, and X. T. Tong. Adaptive Tikhonov strategies for stochastic ensemble Kalman inversion. *Inverse Problems*, 2022.