# Consistency and efficiency of Bayesian estimators in generalised linear inverse problems

# Consistency and efficiency of Bayesian estimators in generalised linear inverse problems

Natalia Bochkina*
University of Edinburgh

Peter J. Green†
University of Bristol

October 24, 2011



**Abstract**

Formulating a statistical inverse problem as one of inference in a Bayesian model has great appeal, notably for what this brings in terms of coherence, the interpretability of regularisation penalties, the integration of all uncertainties, and the principled way in which the set-up can be elaborated to encompass broader features of the context, such as measurement error, indirect observation, etc. The Bayesian formulation comes close to the way that most scientists intuitively regard the inferential task, and in principle allows the free use of subject knowledge in probabilistic model building. However, in some problems where the solution is not unique, for example in ill-posed inverse problems, it is important to understand the relationship between the chosen Bayesian model and the resulting solution.

Taking emission tomography as a canonical example for study, we present results about consistency of the posterior distribution of the reconstruction, and a general method to study convergence of posterior distributions. To study efficiency of Bayesian inference for ill-posed linear inverse problems with constraint, we prove a version of the Bernstein–von Mises theorem for nonregular Bayesian models.

*Some key words:* Ky Fan metric, rates of convergence, SPECT, inverse problems, Bayesian inference, nonregular likelihood, tomography.

*School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, UK. Email: `N.Bochkina@ed.ac.uk`

†School of Mathematics, University of Bristol, Bristol BS8 1TW, UK. Email: `P.J.Green@bristol.ac.uk`.

# 1 Introduction

Inverse problems are almost ubiquitous in applied science and technology, and because of the need for rigorous analysis to characterise such problems, derive numerical solutions and assess their performance – not to mention intrinsic mathematical interest, they have long been the subject of intense mathematical study. In the corresponding 'direct problem', (macroscopic, global) observational data are predicted from the (microscopic, local) model parameters of the system. The inverse problem aims to draw conclusions about model parameters from data: it is the home ground of statistical inference in the context of stochastic modelling.

This paper is a contribution to the theory of inverse problems from a statistical, indeed Bayesian, perspective. Motivated by important problems in tomographic reconstruction, taken as a canonical example, we consider the asymptotic performance of Bayesian procedures in the small-noise limit, for a new class of models that we call generalised linear inverse problems, and discuss further opportunities for theoretical analysis.

In the remainder of this Introductory section, we develop further background for our approach, by setting out our perspective on linear/Gaussian inverse problems.

## 1.1 Ill-posed problems and regularisation

Inverse problems encountered in nature are commonly ill-posed: their solutions fail to satisfy at least one of the three desiderata of existing, being unique, and being stable. Thus, in the case of linear inverse problems, the focus is not on a unique solution $x$ of

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

for given matrix $A$ and data vector $y$, but rather on the corresponding space of solutions.

Even when the solution $x$ to (1) exists and is unique for each possible $y$, lack of stability means that the solution can be extremely sensitive to small errors, either in the observed $y$ or in numerical computations for solving the equations. This has obvious deleterious consequences for the practical value of solutions. To circumvent this, the inverse problem is typically regularised, that is, re-formulated to include additional criteria, such as smoothness of the solution:

$$x = \text{argmin}_{y=Ax} \text{pen}(x),$$

where $\text{pen}(x)$ is a suitable scalar penalty functional.

If the data is observed with error

$$y = Ax + \text{error},$$

then, allowing for the possibility of lack of existence or uniqueness, we might replace the natural least-squares formulation

$$x = \text{argmin}||y - Ax||^2$$

of the inverse problem by

$$x = \text{argmin} ||y - Ax||^2 + \nu \, \text{pen}(x) \tag{2}$$

where $\nu$ a positive constant determining the trade-off between accuracy and smoothness.

Such solutions make sense, and are commonly used, whether we regard the error in the data used as deterministic or stochastic in nature. The least-squares set up is rather natural, but from a statistical perspective corresponds to a Gaussian likelihood, and, as we shall see below, this may be replaced by certain other distributions without material change to the subsequent analysis.

## 1.2 Inverse problems from a Bayesian perspective

Smoothness, or other 'regular' behaviour of the solution to an inverse problem, is a prior assumption on the unknown $x$, information about the model parameters known or assumed before the data are observed. To use such information is thus to accept that the required solution must combine data with prior information. In a statistical context the best-established principle for doing this is the Bayesian paradigm, in which all sources of variation, uncertainty and error are quantified using probability.

From this perspective, the solution to (2) is immediately recognisable – it is the maximum a posteriori (MAP) estimate of $x$, the mode of its posterior distribution in a Bayesian model in which the data $y$ are modelled with a Gaussian distribution with expectation $Ax$, with constant-variance uncorrelated errors, and in which the prior distribution of $x$ has negative log-density proportional to $\text{pen}(x)$.

However, the Bayesian perspective brings more than merely a different characterisation of a familiar numerical solution. Formulating a statistical inverse problem as one of inference in a Bayesian model has great appeal, notably for what this brings in terms of coherence, the interpretability of regularisation penalties, the integration of all uncertainties, and the principled way in which the set-up can be elaborated to encompass broader features of the context, such as measurement error, indirect observation, etc. The Bayesian formulation comes close to the way that most scientists intuitively regard the inferential task, and in principle allows the free use of subject knowledge in probabilistic model building. For an interesting philosophical view on inverse problems, falsification, and the role of Bayesian argument, see Tarantola (2006).

## 1.3 Convergence of the posterior distribution

Mathematical analysis of inverse problems usually takes the form of asymptotic arguments concerning how well the true solution (the value of $x$ assumed to generate the data) can be recovered in the presence of noise, as the size of that noise goes to zero. In a statistical setting, the noise is a random variable, it size might be the variance, and we are concerned with convergence of random variables or their distributions – in the case of a Bayesian analysis, the focus is on the posterior distribution of $x$.

Convergence of the posterior distribution on a finite-dimensional parameter space, with identifiable likelihood and with the true parameter being the interior point of the parameter space, follows from the Doob's martingale convergence theorem (see Doob (1949), and van der Vaart (1998), for the case the sample size grows to infinity). The rate of convergence follows from the Bernstein–von Mises theorem (van der Vaart 1998) which in fact states a stronger result, that the posterior distribution centred at the true parameter and rescaled by $\sqrt{n}$ converges to the Gaussian distribution as the sample size $n$ grows to infinity, provided the likelihood is identifiable with finite Fisher's information matrix, the prior is continuous at the true point and the true value of the parameter is an interior point of the parameter space. Moreover, the limit is independent of the choice of the prior distribution.

However, to the best of our knowledge, the rates of convergence of the posterior distribution on a finite-dimensional parameter space in the case of non-identifiable likelihood and with the true parameter lying on the boundary of the parameter space, studied here, have not been considered previously. The particular example of the lack of identifiability of the likelihood considered here is the ill-posedness of the linear inverse problem. As we shall see, in the case of non-identifiable likelihood, the choice of the prior distribution strongly influences the limit of the posterior distribution as well as the rate of convergence on the subspace where the likelihood is not identified. Also, we will show that the rate of convergence may change if the limiting point lies on the boundary of the parameter space. We shall identify the assumptions on the posterior distribution necessary for convergence which can be used as a guidance to narrow down the set of potential prior distributions.

There are different approaches to quantify the convergence rates. One of them is to consider the concentration rate of the almost sure convergence of the posterior distribution which is the smallest $\varepsilon_\sigma$ such that

$$\mathbb{P}(d(x, x^\star) > \varepsilon_\sigma \mid Y) \to 0 \quad \text{almost surely}$$

considered by Ghosal *et al.* (2000), Walker (2004), van der Vaart and van Zanten (2008) and Rousseau (2010) in the context of nonparametric models.

Another approach, considered by Hofinger and Pikkarainen (2007) in the context of linear inverse problems, is to metrise weak convergence of the posterior distribution as a random variable $\mu_{\text{post}}(\omega) = p(x|Y(\omega))$ using the Ky Fan metric (Fan 1944); see Section 4.2. This type of convergence is weaker than almost sure convergence, and the convergence rates in this metric are slower than the parametric rate with the mean square error loss. In particular, there is an extra logarithm in the rate which is unavoidable.

The setting for Hofinger and Pikkarainen (2007) is the Gaussian linear inverse problem in the form (2), with a particular quadratic penalty (Gaussian prior). Their main result (Theorem 11) provides an upper bound on the Ky Fan metric between the posterior distribution and its (degenerate) limit, as an explicit function of the size of the noise, the parameters of the model and prior, and quantities relating the prior mean to the null space of the matrix $A$. This result is used to prove a limit theorem (Theorem 13) on the convergence of this Ky Fan metric to 0, in a small-noise, high-prior-precision limit, and to give the rate of this convergence (Theorem 15).

We adopt the Hofinger and Pikkarainen (2007) paradigm in the present paper, which extends their results to a broader class of assumed probability distributions for the data, to linear constraints on the solution, and to more general prior distributions, and to the case of the solution of the exact linear inverse problem being on the boundary.

We motivate our study by presenting in Section 2 a nonlinear inverse problem important in medical imaging, and in Section 3 a geometrical view of the results in the linear/Gaussian case. Section 4 establishes the class of models we study and in Section 5 we formulate our theorems on rates of convergence of the posterior distribution. In Section 6 we study local behaviour of the posterior distribution in a neighbourhood of the limit that is formulated as a version of Bernstein–von Mises theorem. Their proofs are deferred to the Appendix.

# 2 Motivation

A general formulation for nonlinear inverse problems would replace $Ax = y$ by $\mathcal{A}(x) = y$, for some suitably smooth transformation or operator $\mathcal{A}$, together with an assumed probability distribution for $Y(\omega)$. Mathematical analysis of such problems is typically far more difficult and technical than for the linear case. However, a more modest generalisation is enough to formulate and analyse a broad range of nonlinear statistical inverse problems of considerable practical importance. The model class we consider is formally defined in Section 4.1; here we consider an important motivating example.

## 2.1 Single photon emission computed tomography

Single photon emission computed tomography (SPECT) is a medical imaging technique in which a radioactively-labelled substance, known to concentrate in the tissue to be imaged, is introduced into the subject. Emitted particles are detected in a device called a gamma camera, forming an array of counts. Tomographic reconstruction is the process of inferring the spatial pattern of concentration of the radioactive isotope in the tissue from these counts.

The Poisson linear model

$$y_t \sim \text{Poisson}(A_t x) \tag{3}$$

independently for different $t$, is close to reality for the SPECT problem (there are some dead-time effects and other artifacts in recording). Here $x$ represents the spatial distribution of the isotope, typically discretised on a grid, $x = \{x_s\}$, and $y$ the array of detected photons, also discretised $y = \{y_t\}$ by the recording process. The array $A = (a_{ts})$ quantifies the emission, transmission, attenuation, decay and recording process; $a_{ts}$ is the mean number of photons recorded at $t$ per unit concentration at pixel/voxel $s$.

See Green (1990) for further detail about the model, and an approach based on EM estimation for MAP reconstruction of $x$, in a Bayesian formulation in which spatial smoothness of the solution is promoted by using a pairwise difference Markov random field prior. Later, Weir (1997) proposed fully Bayesian reconstruction.

Since Poisson distributions form an exponential family, this model can be seen as a generalised linear model (Nelder and Wedderburn 1972), with identity link function, and since $A$ is ill-posed we can call this a *generalised linear inverse problem*.

We formalise the notion of 'small-noise limit' for this Poisson model in a practically-relevant way, by supposing that the exposure time for photon detection is extended by a factor $\mathcal{T}$, and then consider the *rate* of detection of photons, letting $\mathcal{T} \to \infty$. Thus the data-generation model becomes

$$Y_t | x_{\text{true}} \sim \text{Poisson}(\mathcal{T} A_t x_{\text{true}})/\mathcal{T},$$

independently, for $t = 1, 2, \ldots, n$. To preserve the appearance of results from Hofinger and Pikkarainen (200
as far as possible, we write $\mathcal{T}^{-1} = \sigma^2 \to 0$.

## 2.2 Prior distributions

From the beginning of Bayesian image analysis (Geman and Geman 1984; Besag 1986), use has been made of prior distributions for image scenes that express generic, qualitative beliefs about smoothness, yet do not rule out abrupt changes for real discontinuities (for example, at tissue type boundaries in the case of medical imaging).

In common with much of the literature, we will concentrate here on Markov random field prior distributions. The 'true image' $x_{\text{true}}$ in emission tomography corresponds to a physical reality, the discretised spatial distribution of concentration of a radioactive isotope. Of course, this is non-negative, so we impose a constraint, written $x_{\text{true}} \in \mathcal{X} \subset \mathbb{R}^p$ in general.

The first prior model we consider is Gaussian, apart from possible truncation by the constraint,

$$p(x) \propto \exp\left\{-\frac{1}{2\gamma^2}||x - x_0||_B^2\right\}, \quad x \in \mathcal{X},$$

where $||u||_B^2 = u^T B u$ and $B$ is a non-negative definite matrix. An important special case is where $x_0 = 0$ and $B$ satisfies $B_{ss'} = 1$ if $s$ and $s'$ are neighbouring pixels (written $s \sim s'$), otherwise $B_{ss'} = 0$. Then we have $||x - x_0||_B^2 = \sum_{i \sim j} (x_s - x_{s'})^2$, a *pairwise-interaction* model. In this and other important cases $B$ is singular.

A second prior model is a log cosh pairwise-interaction Markov random field (Green 1990):

$$p(x) \propto \exp\left(-\frac{\delta(1+\delta)}{2\gamma^2} \sum_{s \sim s'} \log\cosh((x_s - x_{s'})/\delta)\right), \quad x \in \mathcal{X}.$$

Here the parameter $\delta$ is considered to be fixed.

This model has some attractive properties. While giving less penalty to large abrupt changes in $x$ compared to the Gaussian, it remains log-concave. It bridges the extremes $\delta \to \infty$, the Gaussian model just mentioned, and $\delta = 0$, the corresponding Laplace pairwise-interaction model, sometimes called the 'median prior'.

These distributions are improper since they are invariant to perturbing $x$ by an arbitrary additive constant, but lead to proper posterior distributions, save in exceptional pathological circumstances.

# 3 Geometrical perspective

In this paper, we study inference for $x$ given observed $y$, in the limit as a noise parameter $\sigma^2$ (in the SPECT example, $1/\mathcal{T}$) goes to 0. We generally assume an identity link function, so that $y$ becomes concentrated on $Ax_{\text{true}}$ as $\sigma^2 \to 0$.

Because of the ill-posed/ill-conditioned character of the problem, we cannot expect consistency in inference about $x_{\text{true}}$ based on the likelihood alone. Even as $\sigma^2 \to 0$, so that $y$ converges to 'exact data' $y_{\text{exact}} = Ax_{\text{true}}$, we will not be able to distinguish between $\{x : Ax = Ax_{\text{true}}\}$.

One of the roles of the prior in the Bayesian approach is to resolve this ambiguity (as well as generally improve reconstruction through 'regularisation', even without $\sigma^2 \to 0$). We recall the 'physical' constraint in the SPECT problem, that $x$ is component-wise non-negative, that is, $x \in \mathcal{X} \subset \mathbb{R}^p$, since it quantifies the isotope concentration.

Insight into the interplay between the possibly ill-posed likelihood and the possibly degenerate prior, and the role of the constraint $x \in \mathcal{X}$ can be obtained from a geometrical view of the problem.

## 3.1 Gaussian likelihood and prior

Here we focus on the Gaussian prior $p(x) \propto \exp(-1/(2\gamma^2)||x - x_0||_B^2)$ and Gaussian likelihood $y|x \sim \mathcal{N}(Ax, \sigma^2 I)$. This is the setting of Hofinger and Pikkarainen (2007), except that we will allow $B$ to differ from the identity and even be singular.

In the limit as $\sigma^2 \to 0$, we are interested in solutions of $Ax = y_{\text{exact}}$, where $y_{\text{exact}} = Ax_{\text{true}}$, under the influence of the prior $p(x) \propto \exp(-1/(2\gamma^2)||x - x_0||_B^2)$. To obtain convergence to a degenerate limit, we will need $\gamma^2 \to 0$ as well (though, as shown by Hofinger and Pikkarainen (2007) for the case $B = I$, at a slower rate than $\sigma^2$).

Thus the posterior is proportional to

$$\exp(-1/(2\sigma^2)||y - Ax||^2 - 1/(2\gamma^2)||x - x_0||_B^2) \text{ subject to } x \in \mathcal{X}.$$

Let us first ignore any constraint on $x$, or equivalently assume $\mathcal{X} = \mathbb{R}^p$. By standard manipulations, we can write this posterior as

$$x|y \sim \mathcal{N}\left((A^TA + \nu B)^{-1}(A^Ty + \nu Bx_0), \sigma^2(A^TA + \nu B)^{-1}\right), \tag{4}$$

assuming the inverse matrix exists. As $\sigma^2 \to 0$ and $\gamma^2 \to 0$ in such a way that $\nu = \sigma^2/\gamma^2 \to 0$, the posterior converges to the point

$$x^\star = \text{argmin}_{x \in \mathcal{X}: Ax = y_{\text{exact}}} ||x - x_0||_B^2 \tag{5}$$

Suppose that $A$ is a real $n \times p$ matrix, and $B$ a real symmetric non-negative definite $p \times p$ matrix, both possibly of deficient rank. A rank condition is needed to ensure that the information from the likelihood and prior together define a proper posterior, and determine $x^\star$ uniquely.

**Proposition 1.** *Suppose that $A$ is a real $n \times p$ matrix, and $B$ a real symmetric non-negative definite $p \times p$ matrix, both possibly of deficient rank. Suppose also that the $p \times 2p$ block matrix $[B : A^TA]$ has full rank $p$ (or equivalently, the rows are linearly independent). Then for all $\nu > 0$, $A^TA + \nu B$ is nonsingular.*

*It follows that there exists a nonsingular real matrix $P$, not necessarily orthogonal, such that $P^TBP$, $P^TA^TAP$, and $P^T(A^TA + \nu B)P$ (for all $\nu > 0$) are all diagonal.*

*Furthermore, the limit as $\nu \to 0$ of $\nu(A^TA + \nu B)^{-1}$ is a well-defined finite non-negative definite matrix $C$, and $\nu(A^TA + \nu B)^{-1} - C = O(\nu)$.*

The proof is in Appendix A.1.

This last result gives us a full description of the posterior variance matrix as $\sigma^2 \to 0$, $\gamma^2 \to 0$ while $\nu = \sigma^2/\gamma^2 \to 0$: recalling that the posterior variance (in the Gaussian case) is $\sigma^2(A^TA + \nu B)^{-1}$, we see that in the limit, those components corresponding to $\alpha_i = +\infty$ scale as $\gamma^2$ and the remaining ones as $\sigma^2$. (This is before transformation by $P$, which scales and skews the result, but in a way independent of $\gamma^2$ and $\sigma^2$.) We see from the fact that $P^TA^TAP = \text{diag}(1/(1 + \alpha_i\nu_0))$ that the number of $\alpha_i$ not equal to $+\infty$ is just the rank of $A^TA$.

In summary, the posterior distribution is Gaussian, with variance scaling differently in different directions. If $q$ is the rank of $A$, then asymptotically the variance has $q$ eigenvalues scaling like $\sigma^2$ and the remaining $(p-q)$ like the (larger) $\gamma^2$. Geometrically, contours of equal posterior density are concentric ellipsoids in $\mathbb{R}^p$.

### 3.2 Constrained case, and KKT theory

When $\mathcal{X}$ is a proper subset of $\mathbb{R}^p$, the concentric ellipsoids are truncated by the constraints $x \in \mathcal{X}$. In the case of interest in SPECT, where we have simply componentwise non-negativity contraints, the ellipsoids are truncated into the non-negative orthant. As $\sigma^2$ and $\gamma^2$ become small, there are clear qualitative differences in the impact of this truncation according to whether the centre $(A^TA + \nu B)^{-1}(A^Ty + \nu Bx_0)$ of the ellipsoid lies in the interior of the orthant, on its boundary, or outside it. Since $y$ becomes close to $y_{\text{exact}}$ as $\sigma^2 \to 0$, in the limit this is the same as the question of where does $x^\star$ lie.

Equation (5) is a quadratic programming problem, and could be solved numerically by standard software.

We can get a theoretical handle on the solution through Karush–Kuhn–Tucker theory (Kuhn and Tucker 1951). In the non-negativity constrained case, $\mathcal{X} = \mathbb{R}^p_+$, to minimise $||x - x_0||^2_B$ subject to $x \geq 0$ and $Ax = y_{\text{exact}}$ it is necessary and sufficient to find $(x^\star, \mu, \lambda) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^n$ such that

$$
\begin{gathered}
B(x^\star - x_0) - \mu + A^T\lambda = 0 \\
x^\star \geq 0 \\
Ax^\star = y_{\text{exact}} \\
\mu \geq 0 \\
\text{for all } s, \mu_s = 0 \text{ or } x^\star_s = 0
\end{gathered}
$$

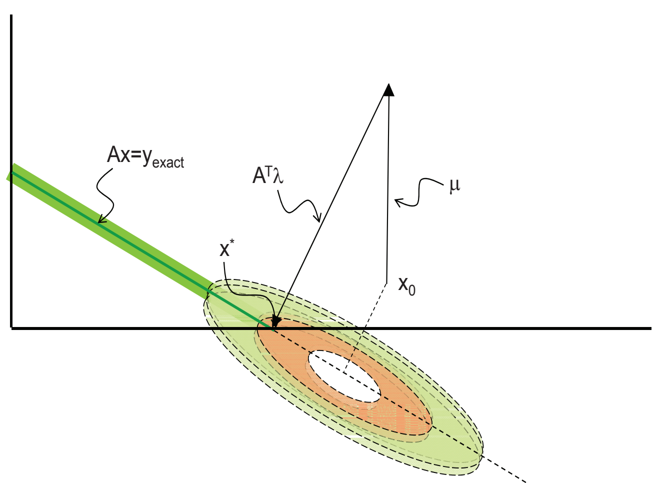


Figure 1: Illustrating the geometry in the case $p = 2$, $n = 1$, with $B = I$. Contours of posterior when $\gamma^2 > \sigma^2 > 0$.

The feasible set $\mathcal{X}^\star = \{x \in \mathcal{X} : Ax = y_{\text{exact}}\}$ is a closed convex set, and $x^\star$ may be an interior point, or satisfy one or more of the constraints $x_s \geq 0$. In the case where all entries of $A$ are non-negative (in accordance with physical reality), and for each $s$ there is at least one $t$ with $A_{ts} > 0$ (and if not, then $x_s$ is unidentifiable, so might as well be omitted from the model), $\mathcal{X}^\star$ is a bounded polyhedron (or polytope). Otherwise, $\mathcal{X}^\star$ may be unbounded.

If $\gamma^2$ remains bounded away from 0 as $\sigma^2 \to 0$, then, in the limit, the posterior has support $\mathcal{X}^\star$.

### 3.3 Geometry in a more general case

The form of (5) strongly suggests that analogous properties for the limit of the posterior should hold in a much broader class of models. Provided that $\sigma^2 \to 0$ and $\gamma^2 \to 0$ in such a way that $\nu = \sigma^2/\gamma^2 \to 0$, we would expect similar limiting behaviour so long as (a) the likelihood is maximised on $\{x : Ax = y_{\text{exact}}\}$ as $\sigma \to 0$, and (b) the prior becomes close to Gaussian as $\gamma \to 0$; subject to these the precise form of the likelihood and prior should be irrelevant. These observations motivate the model formulation of the next section.

In a general setting, more delicate, analytic, arguments will be needed to quantify the convergence precisely, and these are given in the following sections. However, the broad qualitative features of the solution for the Gaussian–Gaussian case (Section 3.1) continue to hold: the posterior becomes increasingly concentrated near the hyperplane $\{x : Ax = y_{\text{exact}}\}$, with $\sigma^2$ dominating its squared variation about this hyperplane, while the variance parallel to the hyperplane is of order $\gamma^2$. The effect of the truncation onto $x \in \mathcal{X}$ depends on whether in the absence of the constraint, the maximum of the posterior would lie in the interior of $\mathcal{X}$, on its boundary, or outside it.

## 4 Model formulation and preliminaries

### 4.1 General Bayesian model

We assume that the joint density of the observable responses $Y$ taking values in $\mathcal{Y} \subset \mathbb{R}^n$ (with respect to Lebesgue or counting measure) takes the form

$$p\,(y|\,x) = F(y, Ax, \tau) = C_{y,\tau} \exp\left\{\frac{1}{\tau}\tilde{f}_y(Ax)\right\}, \quad y \in \mathcal{Y}, \tag{6}$$

that is, that the distribution depends on $x \in \mathcal{X}$ only via $Ax$, where $\tau$ is a scalar dispersion parameter; in the Gaussian model, $\tau$ is the variance $\sigma^2$. The observed data are generated from this distribution, with $x = x_{\text{true}}$, and we aim to recover $x_{\text{true}}$ as $\tau \to 0$.

We assume a continuous bijective link function $G : \mathcal{Y} \to \mathbb{R}^n$ and write $G(y_{\text{exact}}) = Ax_{\text{true}}$. (In generalised linear models – see Example 3 below – commonly $G$ has identical component functions.)

We adopt a Bayesian paradigm, using a prior distribution with density given by

$$p\,(x) \propto \exp(g(x)/\gamma^2), \quad x \in \mathcal{X} \subset \mathbb{R}^p, \tag{7}$$

where $\gamma^2$ is a scalar dispersion parameter for the prior that may depend on $\tau$; we relate this to the data dispersion parameter $\tau$ by $\gamma^2 = \tau/\nu$, and express most of our results below in terms of $\tau$ and $\nu$. Thus the posterior distribution satisfies

$$p\,(x|\,y) \propto \exp([\tilde{f}_y(Ax) + \nu\, g(x)]/\tau), \quad x \in \mathcal{X}, \tag{8}$$

Denote $f_y(x) = \tilde{f}_y(Ax)$ and $h_y(x) = -f_y(x) - \nu\, g(x)$, so that $p\,(x|\,y) \propto e^{-h_y(x)/\tau}$.

We make the following assumptions about the error distribution:

1. If $Y \sim F(y, G(y_{\text{exact}}), \tau)$, then $Y \xrightarrow{P} y_{\text{exact}}$ as $\tau \to 0$.
2. For all $\mu_0 \in G^{-1}(A\mathcal{X})$, $\tilde{f}_{\mu_0}(\eta)$ has a unique maximum over $A\mathcal{X}$ at $\eta = G(\mu_0)$, $\nabla_\eta \tilde{f}_{\mu_0}(G(\mu_0)) = 0$ and $\nabla^2_\eta \tilde{f}_{\mu_0}(G(\mu_0))$ is of full rank.

(Throughout, we use $\nabla_i = \frac{\partial}{\partial x_i}$ as the differentiating operator, and $\nabla = (\nabla_1, \ldots, \nabla_p)^T$ as the gradient. Similarly, $\nabla_{ij}$ and $\nabla_{ijk}$ are operators of the second and third derivatives, with $\nabla^2 = (\nabla_{ij})$ being the matrix of second derivatives.)

Assumption 2 implies that the likelihood is regular in $Ax$. Various conditions are sufficient for a distribution to satisfy this assumption, for example, the following:

2†. $G(\mathbb{E}Y) = Ax$ and $\exists\, \alpha > 0$ and $v(\tau)$: $\forall\, i$, $\mathbb{E}\left(|Y_i - \mathbb{E}Y_i|^\alpha\right) \leqslant v(\tau)$ such that $v(\tau) \to 0$ as $\tau \to 0$.

Then, a variant of Chebyshev's inequality implies convergence.

**Example 1.** *Assumption 2 is satisfied even for a location-scale Cauchy distribution $t_1(\mu, \sigma)$ with, say, $\alpha = 1/2$:*

$$\mathbb{E}|Y - \mu|^{1/2} = \sqrt{\sigma} \int_0^\infty \frac{2\sqrt{x}}{\pi(1 + x^2)} dx,$$

*which is finite and goes to 0 as $\sigma \to 0$. However, assumption 1 is not satisfied for the Cauchy distribution (or indeed any rescaled/recentered distribution with polynomial decay) since the density cannot be cast in the form (6) for any choice of $\tau$.*

**Example 2.** *Both assumptions are satisfied for the power exponential (Subbotin) distributions $F(y, \mu, \sigma) = C_{\sigma,\beta} \exp\{-[(y-\mu)^2]^{\beta/2}/\sigma^\beta\}$ $(\beta > 0)$, with $\tau = \sigma^\beta$ and $\tilde{f}_y(\mu) = [(y-\mu)^2]^{\beta/2}$.*

**Example 3.** *In the generalised linear models of Nelder and Wedderburn (1972), an important class of nonlinear statistical regression problems, responses $y_t$, $t = 1, 2, \ldots, n$ are drawn* *independently from a one-parameter exponential family of distributions in canonical form, with density or probability function*

$$p(y_t; \mu_t, \tau) = \exp\left(\frac{y_t b(\mu_t) - c(\mu_t)}{\tau} + d(y_t, \tau)\right),$$

*using the mean parameterisation, for appropriate functions $b$, $c$ and $d$ characterising the particular distribution family. The parameter $\tau$ is a common dispersion parameter shared by all responses. The expectation of this distribution is $\mathbb{E}(y_t; \mu_t, \tau) = \mu_t = c'(\mu_t)/b'(\mu_t)$. Both assumptions are satisfied for this example.*

As the link function $G$ is continuous and monotonic, we could consider a linear inverse problem $Ax = \tilde{y}_{\text{exact}}$ where $\tilde{y}_{\text{exact}} = G(y_{\text{exact}})$, $\tilde{Y} = G(Y)$ and $\tilde{\mathcal{Y}} = G(\mathcal{Y})$. The expressions with respect to $x$ do not change, however, the Ky Fan distance $\rho_{\text{K}}(Y, y_{\text{exact}})$ is replaced with $\rho_{\text{K}}(\tilde{Y}, \tilde{y}_{\text{exact}})$. Hence, to simplify the notation, we assume below that the link function is the identity.

We will assume that $\mathcal{X} = [0, \infty)^p$. We could assume that parameter $x$ is restricted to an arbitrary convex polyhedron; this could be reduced to $[0, \infty)^p$ by a linear change of variables. In fact, the results below apply to $\mathcal{X}$ such that for any $x \in \mathcal{X}$, $[B(x, \delta) \cap \mathcal{X} - x]/\tau \to \mathbb{R}^k \times [0, \infty)^m \times (-\infty, 0]^{p-k-m}$ as $\tau, \delta \to 0$ and $\delta/\tau \to \infty$.

## 4.2 Metrics for quantifying convergence

**Definition 1.** *The Ky Fan metric between two random variables $\xi_1$ and $\xi_2$ in a metric space $(\mathcal{Y}, d_{\mathcal{Y}})$ is defined by*

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

Convergence in this metric is equivalent to convergence in probability (Dudley 2003). Hence, weak convergence of the posterior distribution $\mu_{\text{post}}$ (as a random variable) to $\delta_{x^\star}$, the point mass at $x^\star$, is equivalent to its convergence in the Ky Fan metric, where the metric space $(\mathcal{Y}, d_{\mathcal{Y}})$ is a space of probability distributions equipped with the Prokhorov metric.

**Definition 2.** *The Prokhorov metric between two measures on a metric space $(\mathcal{X}, d_{\mathcal{X}})$ is defined by*

$$\rho_{\text{P}}(\mu_1, \mu_2) = \inf\{\varepsilon > 0 : \mu_1(B) \leq \mu_2(B^\varepsilon) + \varepsilon \; \forall \text{ Borel } B\}$$

*where $B^\varepsilon = \{x : \inf_{z \in B} d_{\mathcal{X}}(x, z) < \varepsilon\}$.*

In particular, convergence in this metric is equivalent to convergence in distribution (Dudley 2003), and so weak convergence of the posterior distribution can be studied as convergence of the Ky Fan metric $\rho_{\text{K}}(\mu_{\text{post}}, \delta_{x^\star})$.

### 4.3 Boundary and local geometry

Now we describe the local geometry of the posterior distribution around the point $x^\star$ where

$$x^\star \quad = \quad \text{argmax}_{x \in \mathcal{X} \colon Ax = y_{\text{exact}}} g(x).$$

We assume that the prior distribution is such that $x^\star$ is a unique solution. We relax the assumption that $x^\star$ is a regular point, by allowing it to lie on the boundary of $\mathcal{X}$.

The definition above implies that if $x^\star$ is an interior point of $\{x \in \mathcal{X} : Ax = y_{\text{exact}}\}$, then

$$0 = \left( \frac{\partial}{\partial z_i} g(x^\star + (I - P_{A^T}) z)|_{z=0} \right)_{i=1}^{p} = (I - P_{A^T}) \nabla g(x^\star), \tag{9}$$

where $P_{A^T}$ is the projection on the range of $A$. However, if $x^\star$ is on the boundary, the gradient $\nabla g(x^\star)$ may not be zero. In case of a truncated distribution of $X$ this corresponds to the maximum lying outside of $\mathcal{X}$ (see Section 3 for further insight into the geometry of the boundary). Denote the set of coordinates where this vector is non-zero by

$$S = \{i : [\nabla g(x^\star)]_i \neq 0\},$$

and the projection on the $S$ coordinates by $P_S$, i.e. $(P_S)_{ii} = 1$ if $i \in S$ and $(P_S)_{ij} = 0$ for all other $i, j$. We assume that

$$\text{rank}\left( \left[ A^T : P_S \right] \right) = \text{rank}(A) + \text{rank}(P_S)$$

(where $[A^T : P_S]$ is the block matrix putting $A^T$ and $P_S$ side by side); this simply prevents degeneracy.

We consider the spaces $\mathcal{A} = \{v : Av = 0\}$ and $\mathcal{C} = \{v : P_S v = 0\}$ (the null spaces of $A$ and $P_S$), and their intersection. Let $r_0 = \text{rank}(A)$, $r_2 = \text{rank}(C)$ and set $r_1 = p - r_0 - r_2$ (so that $r_0 + r_1 + r_2 = p$). By the rank–nullity theorem, and using the rank assumption above, the dimensions of $\mathcal{A}$, $\mathcal{C}$ and $\mathcal{A} \cap \mathcal{C}$ are respectively $r_1 + r_2$, $r_0 + r_1$ and $r_1$.

Define three vector spaces by $\mathcal{V}_0 = \mathcal{C} \cap (\mathcal{A} \cap \mathcal{C})^\perp$, $\mathcal{V}_1 = \mathcal{A} \cap \mathcal{C}$, $\mathcal{V}_2 = \mathcal{A} \cap (\mathcal{A} \cap \mathcal{C})^\perp$; here, the orthogonal complements in the definitions of $\mathcal{V}_0$ and $\mathcal{V}_2$ are with respect to the same fixed choice of inner product as the projection operators defined earlier. It follows that $\mathcal{V}_i$ has dimension $p_i, i = 0, 1, 2$, $\mathbb{R}^p = \mathcal{V}_0 \oplus \mathcal{V}_1 \oplus \mathcal{V}_2$, and we can write any $v = x - x^\star$ in $\mathbb{R}^p$ uniquely in the form $v = v_0 + v_1 + v_2$, where $v_i \in \mathcal{V}_i, i = 0, 1, 2$.

In the regular, non-boundary case where $x^\star$ is in the interior of $\mathcal{X}$, then $S$ is empty, $P_S$ has rank 0, $r_0 = \text{rank}(A)$, $r_1 = p - \text{rank}(A)$ and $r_2 = 0$.

Now consider the following projections and compositions of projections:

$$Q_0 = (I - P_S) P_{A^T : P_S}, \quad Q_1 = (I - P_{A^T : P_S}), \quad Q_2 = (I - P_{A^T}) P_{[A^T : P_S]}.$$

For each $i = 0, 1, 2$, $Q_i$ is diagonalisable, and we can let $U_i$ denote a $p \times p_i$ matrix of orthogonal eigenvectors corresponding to the non-zero eigenvalues where $p_i = \text{rank}(Q_i)$.

Note that $p_2 = r_2$ but $p_0 \leq r_0$ (and hence $p_1 \geq r_1$); $p_0 = r_0$ only if matrix $A^T V_{y_{\text{exact}}}(x^\star)A$ is of the same rank as matrix $A^T A$.

Now let $U$ be the block matrix

$$U = [U_0 : U_1 : U_2] \,.$$

Note that $U$ is not orthogonal – the centre block is orthogonal to the first and last block, but the first and last blocks are not orthogonal to each other. These three blocks form bases for $\mathcal{V}_i$, $i = 0, 1, 2$, and $U$ itself is a basis for $\mathbb{R}^p$. Any $v = x - x^\star$ in $\mathbb{R}^p$ can be expressed uniquely as $v = Uw$, and $U^{-1}$ and $w$ can be partitioned conformably as

$$U^{-1} = \begin{pmatrix} (U^{-1})_0 \\ (U^{-1})_1 \\ (U^{-1})_2 \end{pmatrix}, \quad \text{and} \quad w = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$$

where $(U^{-1})_i$ is a $p_i \times p$ matrix, $w_i \in \mathbb{R}^{p_i}$ and $U_i w_i = v_i \in \mathcal{V}_i$ for each $i$. We will also denote $(U^{-1})_{01} = \begin{pmatrix} (U^{-1})_0 \\ (U^{-1})_1 \end{pmatrix}$, $U_{01} = [U_0 : U_1]$ and $p_{01} = p_0 + p_1$. In particular, $(U^{-1})_k U_k = I_{p_k}$ for $k = 0, 1, 2$ and $(U^{-1})_{01} U_{01} = I_{p_0 + p_1}$.

Introducing the matrices of second derivatives:

$$\begin{aligned} V_y(x) &= -\nabla^2 \tilde{f}_y(Ax), \\ B(x) &= -\nabla^2 g(x), \\ H_y(x) &= \nabla^2 h_y(x) = A^T V_y(x) A + \nu B(x), \end{aligned}$$

the following quantities will be used in approximating the posterior distribution:

$$HH_{01,01} = U_{01} H_y(x^\star) U_{01}^T, \tag{10}$$

$$x_0 = HH_{01,01}^{-1} \nabla h_y(x^\star), \tag{11}$$

$$b = |U_2 \nabla g(x^\star)|, \tag{12}$$

$$b_{\min} = \min_i b_i. \tag{13}$$

## 4.4 Quadratic approximation of $\log p(x \mid y)$

We approximate $h_y(x) = -\tau \log p(x \mid y)$ by a quadratic function of $x$ on $B(x^\star, \delta)$ for an appropriate $\delta$ using Taylor expansion:

$$h_y(x) = h_y(x^\star) + (x - x^\star)^T \nabla h_y(x^\star) + \frac{1}{2}(x - x^\star)^T \nabla^2 h_y(x^\star)(x - x^\star) + \Delta_h(\delta).$$

We will need the following assumptions. Introduce the following neighbourhood of $y_{\text{exact}}$ in $\mathcal{Y}$:

$$\mathcal{Y}_{\text{loc}} = \{y \in \mathcal{Y} : \quad ||y - y_{\text{exact}}|| \leq \rho_{\text{K}}(Y, y_{\text{exact}})\}.$$

By the definition of $\rho_{\text{K}}(Y, y_{\text{exact}})$, $\mathbb{P}(\mathcal{Y}_{\text{loc}}) \geq 1 - \rho_{\text{K}}(Y, y_{\text{exact}})$.

Assume that

1. $\exists \tau_0 > 0$: $\quad \forall \tau \leqslant \tau_0, \int_{\mathcal{X}} e^{-h_y(x)/\tau} dx < \infty$ for all $y \in \mathcal{Y}$;
2. $f_y$, $g \in \mathbb{C}^3(B(x^\star, \delta))$ for all $y \in \mathcal{Y}$;
3. $\exists C_{f,3}$, $C_{g,3} < \infty$ such that for all $x \in B(x^\star, \delta)$, for all $y \in \mathcal{Y}_{\text{loc}}$ and all $1 \leqslant i, j, k \leqslant p$,
$$|\nabla_{ijk} f_y(x)| \leqslant C_{f,3}, \tag{14}$$
$$|\nabla_{ijk} g(x)| \leqslant C_{g,3}; \tag{15}$$
4. (a) $\exists M_{f,0}, M_{f,1}, M_{f,2} < \infty$ such that for all $1 \leqslant j_1, \ldots, j_d \leqslant p$ with $d = 0, 1, 2$, and for all $y \in \mathcal{Y}_{\text{loc}}$,
$$|\nabla_{j_1,\ldots,j_d} f_y(x^\star) - \nabla_{j_1,\ldots,j_d} f_{y_{\text{exact}}}(x^\star)| \leqslant M_{f,d} ||y - y_{\text{exact}}||; \tag{16}$$
   (b) $\exists M_{f,3} < \infty$ such that for all $x \in B(x^\star, \delta)$ and all $1 \leqslant i, j, k \leqslant p$ and for $y \in \mathcal{Y}_{\text{loc}}$,
$$|\nabla_{ijk} f_y(x) - \nabla_{ijk} f_{y_{\text{exact}}}(x)| \leqslant M_{f,3} ||y - y_{\text{exact}}||. \tag{17}$$

The last two assumptions are satisfied if $\nabla^d f_{\mu_0}(x)$ is differentiable in $\mu_0$ and this derivative is bounded on $\mathcal{Y}_{\text{loc}}$, with

$$\begin{aligned} M_{f,d} &= \sup_{y \in \mathcal{Y}_{\text{loc}}} |\nabla_y \nabla_x^d f_y(x^\star)| \quad \text{for} \quad d = 0, 1, 2, \\ M_{f,3} &= \sup_{y \in \mathcal{Y}_{\text{loc}}} \sup_{x \in B(x^\star, \delta)} |\nabla_y \nabla_x^3 f_y(x)|. \end{aligned}$$

We choose $\delta$ such that the approximation error $\Delta_h(\delta)$ of $h_y(x)$ goes to 0, and that the integral of $e^{-h_y(x)/\tau}$ over $\mathcal{X} \setminus B(x^\star, \delta)$ is negligible compared to the integral over $B(x^\star, \delta)$. Assume that $\delta > 0$ satisfies the following conditions as $\tau \to 0$:

$$\delta > ||x_0||, \quad \delta^2 \ll ||x_0||, \quad \frac{\lambda_{\min}(HH_{01,01})\delta^2}{\tau} \to \infty \quad \text{with high probability}, \tag{18}$$
$$\delta \to 0, \quad \frac{\delta^4}{\tau} \to 0, \quad \frac{\delta \nu}{\tau} \to \infty.$$

## 4.5 Choice of $\delta$

Consider the integral of $e^{-h_y(x)/\tau}$ over $B(x^\star, \delta)$.

**Lemma 1.** *Under the assumptions on $f_y$, $g$ and $\delta$, and assuming that $HH_{01,01}^{-1}$ exists,*

$$\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau} dx = \frac{\tau^{(p+p_2)/2} e^{-h_y(x^\star)/\tau + x_0^T HH_{01,01}^{-1} x_0/(2\tau)} (2\pi)^{p_{01}/2}}{\nu^{p_2} [\det(HH_{01,01})]^{1/2} \prod_{i=1}^{p_2} b_i} \det(U)[1 + o_P(1)].$$

See Proposition 2, in the Appendix, for further details and the proof.

Now we need to choose $\delta$ satisfying the conditions of the lemma such that the integral over the remaining space $\mathcal{X} \setminus B(x^\star, \delta)$ is negligibly small compared to the integral over $B(x^\star, \delta)$ for small $\tau$, i.e. that

$$\begin{aligned} \Delta_0(B(0,\delta)) &= \frac{\int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx}{\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx} \qquad (19) \\ &= \frac{\int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx}{\tau^{(p+p_2)/2} \nu^{-p_2} e^{x_0^T HH_{01,01} x_0/(2\tau)} (2\pi)^{p_{01}/2} [\det(HH_{01,01})]^{-1/2} \prod_{i=1}^{p_2} b_i^{-1}} [1+o(1)] \\ &= o(1) \quad as \quad \tau \to 0 \end{aligned}$$

under the assumptions of Lemma 1. This condition (19) is satisfied, for instance, under the following assumptions on $h_y$.

Assume that there exists a function $q > 0$ such that for $x \in \mathcal{X} \setminus B(x^\star, \delta)$, $h_y(x) - h_y(x^\star) \geq q(||x - x^\star||)$, and $q(r) \geq cr^\alpha$ for $r > \delta$ and some $c = c(\delta) > 0, \alpha \in (0, 3)$.

Then, it is sufficient to choose $\delta$ satisfying

$$\int_{c\delta^\alpha/\tau}^{\infty} z^{p/\alpha-1} e^{-z} dz \ll \alpha \tau^{-p/\alpha} c^{p/\alpha} \tau^{(p+p_2)/2} \nu^{-p_2} \prod b_i^{-1} (2\pi)^{p_{01}/2} [\det(HH_{01,01})]^{-1/2}, \qquad (20)$$

which is satisfied, if $c(\delta) \asymp const$, e.g. with $\delta = [-\tau \log \tau]^{1/((1+a)\alpha)}$ for some $a > 0$. In this case, it is possible to choose $\delta$ satisfying conditions (18) for appropriate $a$ and $\nu$ if $\alpha < 3$.

Therefore, we can use the following lemma.

**Lemma 2.** *Assume that there exists a function $q : \mathbb{R}_+ \to \mathbb{R}_+$ such that for $||x - x^\star|| > \delta$, $h_y(x) - h_y(x^\star) \geq q(||x - x^\star||)$, and $q(r) \geq cr^\alpha$ for some $c = c(\delta) > 0$ and $\alpha \in (0, 3)$.*

*Then, for $\delta$ satisfying (18) and (20),*

$$\int_{\mathcal{X}} e^{-(h_y(x)-h_y(x^\star))/\tau} dx = [1 + o(1)] \int_{B(x^\star,\delta)} e^{-(h_y(x)-h_y(x^\star))/\tau} dx$$

*as $\tau \to 0$.*

*In particular, if $c(\delta) \asymp const$, the above condition is satisfied with $\delta = [-\tau \log \tau]^{1/((1+a)\alpha)}$ for any $a > 0$.*

If we assume that $q(r) \geq c \log r$, $c \geq \tau(p+1)$, then, to have $\int_{B(x^\star,\delta)} e^{-(h_y(x)-h_y(x^\star))/\tau} dx \gg \int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-(h_y(x)-h_y(x^\star))/\tau} dx$, $\delta$ must satisfy

$$\delta^{c/\tau - p}(c/\tau - p) \gg \tau^{(p+p_2)/2} \nu^{-p_2} \prod [b_i]^{-1} (2\pi)^{p_{01}/2} [\det(HH_{01,01})]^{-1/2},$$

or, equivalently,

$$\delta \gg \exp\left\{ \tau \frac{\log C}{c} + \tau \log \tau \frac{p(1 + p_2/p)}{2c} \right\},$$

where $C = \prod [b_i]^{-1} (2\pi)^{p_{01}/2} [\det(HH_{01,01})]^{-1/2}$. If $c(\delta)$ is a constant independent of $\delta$, the fourth condition of (18), that $\delta \to 0$, is not satisfied.

# 5 Rates of convergence of posterior distribution in Ky Fan metric

As we have seen in Section 3, the rate of contraction of the posterior distribution (in terms of Ky Fan distance) varies between $P_{A^T}\mathcal{X}$ and $(I - P_{A^T})\mathcal{X}$ and is determined by the second order behaviour of the logarithm of the posterior density. We shall also show below that, if $x^\star$ is on the boundary, the contraction rate in $P_S(I - P_{A^T})\mathcal{X}$ is different and is determined by the first order asymptotics.

Denote by $\mu_{\text{post}}(\omega)$ the posterior distribution of $X$ given $y = Y(\omega)$. We consider the metric space $(\mathcal{X}, \ell_2)$ equipped with the Euclidean metric $||x-z|| = \sqrt{\sum_{i=1}^p (x_i - z_i)^2}$, $\mathcal{X} \subset \mathbb{R}^p$. Then, the posterior measure $\mu_{\text{post}}(\omega)$ can be viewed as a measure on the metric space $(\mathcal{X}, \ell_2)$. The corresponding metric space for the observations is $(\mathcal{Y}, \ell_2)$, $\mathcal{Y} \subset \mathbb{R}^n$ equipped with metric generated by $\ell_2$ norm.

In the next section we evaluate the level of concentration of the posterior distribution $\mu_{\text{post}}$ around $x^\star$. We start with the concentration of the posterior distribution $\mu_{\text{post}}(\omega)$ for a fixed $\omega$ (i.e. for a particular data set) in the Prokhorov metric, and then, using the lifting theorem (Theorem 3), we use bounds thus obtained to derived a bound on the Ky Fan distance between the posterior distribution and the limit over all $\omega$. We consider separately the cases where $x^\star$ is an interior point of $\mathcal{X}$ and where it is on the boundary of $\mathcal{X}$. In the results below, it is assumed that the dimension $p$ is fixed and is independent of $\tau$.

Throughout this section we use the error $\Delta_0(B(0,\delta))$ defined by (19), and constants $\kappa_p$ and $C_p$ defined by (42) that feature in the upper bound on the Ky Fan metric between the Gaussian distribution and its mean (Lemma 3 in the Appendix).

## 5.1 Prokhorov distance, fixed $\omega$

Define $\lambda_{\min\text{pos}}(M)$ to be the minimum positive eigenvalue of a matrix $M$, and $\lambda_{\min, P}(M) = \min_{||v||=1,\, Pv=v} ||Mv||$ to be the smallest eigenvalue of a matrix $M$ on the range of a projection matrix $P$.

**Theorem 1.** *Suppose we have a Bayesian model given in Section 4.1, and let the assumptions stated in Section 4.4 hold.*

*Assume that $(I - P_{A^T})\nabla g(x^\star) = 0$ and $[A^T V_{Y(\omega)}(x^\star)A : B(x^\star)]$ is of full rank.*

*Then, $\exists \tau_0 > 0 > 0$ such that for $\forall\, \tau \in (0, \tau_0]$,*

$$\rho_{\text{P}}(\mu_{\text{post}}(\omega), \delta_{x^\star}) \leqslant \max\left\{\frac{2\Delta_0}{1+\Delta_0}, \quad \frac{M_{f1}||Y(\omega) - y_{\text{exact}}|| + \nu||P_{A^T}\nabla g(x^\star)||}{\lambda_{\min,\text{pos}}(A^T V_{Y(\omega)}(x^\star)A)} \right.$$
$$\left. + \sqrt{-\frac{\tau}{\lambda_{\min}(\omega)}\left(C_p \log\left(\frac{\tau}{\lambda_{\min}(\omega)}\right)^{\kappa_p}\right)(1 + \Delta_\star(\delta, 0, Y(\omega)))}\right\}, \quad (21)$$

*where* $\lambda_{\min}(\omega) = \lambda_{\min}(H_{Y(\omega)}(x^\star))$ *and* $\Delta_\star$ *is defined by (37).*

The first term in the sum represents the bias of the posterior distribution, and the second term is the Prokhorov distance between $\mathcal{N}(0, \tau H_{Y(\omega)}(x^\star)^{-1})$ and the point mass at zero. The maximum reflects the fact that there are two "competing" tails: Gaussian on the ball $B(x^\star, \delta)$ and the tail of the posterior distribution outside the ball.

This theorem implies that to have convergence of the posterior distribution to $\delta_{x^\star}$, we must have (a) convergence of the data so that $||Y - y_{\text{exact}}|| \stackrel{\mathbb{P}_{x_{\text{true}}}}{\to} 0$, (b) $\nu = \tau/\gamma^2 \to 0$, i.e. the prior distribution needs to be rescaled in a way dependent on the scale of the likelihood, and (c) $\tau/\lambda_{\min}(H_{Y(\omega)}(x^\star)) \to 0$. If the matrix $A^T V_{Y(\omega)}(x^\star)A$ is of full rank, then, for small $\tau$, $\lambda_{\min}(H_{Y(\omega)}(x^\star))$ is close to the constant $\lambda_{\min}(A^T V_{y_{\text{exact}}}(x^\star)A)$ with high probability, hence the latter condition is satisfied as $\tau \to 0$. However, if $A^T V_{Y(\omega)}(x^\star)A$ is not of full rank, then, for small enough $\nu$ and $\tau$, $\lambda_{\min}(H_{Y(\omega)}(x^\star)) = \nu\lambda_{\min,\mathrm{I}-\mathrm{P}_{A^T}}(B(x^\star))$; hence, we must have $\tau/\nu = \gamma^2 \to 0$.

This is summarised in the following corollary.

**Corollary 1.** *For weak convergence of the posterior distribution to the point mass at* $x^\star$ *as* $\tau \to 0$ *for a fixed* $\omega$*, we must have* $\nu = \tau/\gamma^2 \to 0$.

*1. If the matrix* $A^T V_{Y(\omega)}(x^\star)A$ *is not of full rank, then we must also have* $\gamma \to 0$.

*2. If the matrix* $A^T V_{Y(\omega)}(x^\star)A$ *is of full rank, however, the scale of the prior distribution* $\gamma$ *may be taken a positive constant.*

Now we consider the case where $x^\star$ is a boundary point of $\mathcal{X}$ and $(I - P_{A^T})\nabla g(x^\star) \neq 0$.

**Theorem 2.** *Suppose we assume the Bayesian model defined in Section 4.1, and let the assumptions on* $f_y$*,* $g$ *and* $\delta$ *stated in Section 4.4 hold.*

*Assume that* $U_{01}[A^T V_{Y(\omega)}(x^\star)A : B(x^\star)]U_{01}^T$ *is of full rank.*

*Then,* $\exists \tau_0 > 0$ *such that for* $\forall\, \tau \in (0, \tau_0]$ *and small enough* $\tau/\nu$ *and for any* $a \in (0, 1)$*,*

$$\begin{aligned}
\rho_{\mathrm{P}}(\mu_{\text{post}}(\omega), \delta_{x^\star}) \;\leqslant\; \max\Bigg\{ & \frac{2\Delta_0}{1+\Delta_0}, \quad \frac{M_{f1}||Y(\omega) - y_{\text{exact}}|| + \nu||P_{A^T}\,\nabla g(x^\star)||}{\lambda_{\min,\text{pos}}(U_{01}A^T V_{Y(\omega)}(x^\star)AU_{01}^T)}[a + \sqrt{1-a^2}]^{-1} \\
& + \sqrt{-\frac{\tau}{\lambda_{\min,\,01}(\omega)} \log\left( C_{p_{01}} \left( \frac{\tau\, a^2}{\lambda_{\min,\,01}(\omega)} \right)^{\kappa_{p_{01}}} \right) (1 + \Delta_\star(\delta, p_2, Y(\omega)))} \\
& - \frac{\tau}{\nu b_{\min}} \log\left( \frac{\tau p_2}{\nu b_{\min}\sqrt{1-a^2}} \right) (1 + \Delta_{\star\star}(\delta, p_2, Y(\omega))) \Bigg\},
\end{aligned} \tag{22}$$

*where* $\lambda_{\min,\,01}(\omega) = \lambda_{\min,\,\mathrm{U}_{01}}(H_{Y(\omega)}(x^\star))$ *and* $\Delta_\star$ *and* $\Delta_{\star\star}$ *are defined by (37).*

To have convergence of the posterior distribution to $\delta_{x^\star}$ for a fixed $\omega$ when $x^\star$ is on the boundary, a similar argument as in the case when $x^\star$ is an interior point applies (with

$A^T V_{y_{\text{exact}}}(x^\star)A$ and $H_{Y(\omega)}(x^\star)$ replaced by $U_{01}A^T V_{y_{\text{exact}}}(x^\star)AU_{01}^T$ and $U_{01}H_{Y(\omega)}(x^\star)U_{01}^T$ respectively), but also we must have $\tau/\nu = \gamma^2 \to 0$. Hence, in this case to have the convergence we must assume that $\nu = \tau/\gamma^2 \to 0$ and $\gamma \to 0$ as $\tau \to 0$.

The value of $a$ balances the parts of the Prokhorov distance attributable to Gaussian and exponential tails on $B(x^\star, \delta)$ (see the proof of the theorem for details). In the ill-posed case, the Gaussian Prokhorov rate is slower than the exponential one, hence we can choose $a$ in such a way that $a \approx 1$ and $\log\left(-\gamma^2/\sqrt{1-a^2}\right) \to \infty$ as $\gamma \to 0$, e.g. we can choose $1 - a^2 = \gamma^2$, i.e. $a = \sqrt{1-\gamma^2}$. If $U_{01}A^T V_y(x^\star)AU_{01}^T$ is of full rank, then the exponential rate $-\gamma^2 \log \gamma$ is slower than the Gaussian one $\tau\sqrt{-\log \tau}$, hence we can choose $a \approx 0$ such that the rate $\tau\sqrt{-\log(a^2\tau)}$ remains slower, e.g. $a = \tau^{1/4}$.

These theorems give an upper bound on the Prokhorov distance between the posterior distribution and the limit for any particular instance of observed data $Y(\omega)$. To "lift" the result obtained to a bound on the Ky Fan distance over all $\omega$, we use the following generalisation of the lifting theorem of Hofinger and Pikkarainen (2007) to the case of different bounds for different outcomes $\omega$.

**Theorem 3.** *Let random variables $X_1$, $X_2$ and $Y_1$, $Y_2$ be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in metric spaces $(X, d_x)$ and $(Y, d_y)$, respectively, and suppose the sample space $\Omega$ is partitioned into two parts, $\Omega = \Omega_1 \cup \Omega_2$, $\Omega_1 \cap \Omega_2 = \emptyset$.*

*Assume that there exist positive nondecreasing functions $\Phi_1$ and $\Phi_2$:*

$$\forall \omega \in \Omega_k, \quad d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_k(d_y(Y_1(\omega), Y_2(\omega))), \quad k = 1, 2$$

*i.e. we have different upper bounds on $\Omega_1$ and $\Omega_2$.*

*Then, the following inequalities hold:*

$$\begin{aligned}
\rho_{\mathrm{K}}(X_1, X_2) &\leqslant \max\{\rho_{\mathrm{K}}(Y_1, Y_2) + P(\Omega_2), \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2))\}, \\
\rho_{\mathrm{K}}(X_1, X_2) &\leqslant \max\{\rho_{\mathrm{K}}(Y_1, Y_2), \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2)), \Phi_2(\rho_{\mathrm{K}}(Y_1, Y_2))\}.
\end{aligned}$$

In our case, $(X, d_x)$ is the space of all distributions equipped with the Prokhorov metric, and $(Y, d_y)$ is the metric space $\mathcal{Y}$ with the $\ell_2$ metric. Theorems 1 and 2 provide an upper bound $\Phi_1$ on the event $\Omega_1$ where a random matrix $H_{Y(\omega)}(x^\star)$ (or $U_{01}H_{Y(\omega)}(x^\star)U_{01}^T$) is of full rank, and the first statement of the theorem is applied to obtain the Ky Fan rate of convergence.

## 5.2 Consistency of the posterior distribution

Applying the lifting inequalities given in Theorem 3 together with Theorem 1, we obtain the following bound on the Ky Fan distance.

Denote

$$v_{\min} = \min_{t:\, V_{y_{\text{exact}}\, tt}(x^\star)>0} V_{y_{\text{exact}}\, tt}(x^\star), \tag{23}$$

$$c_1 = \frac{M_{f1}}{v_{\min}\lambda_{\min,\text{pos}}(U_{01}A^TAU_{01}^T)} \tag{24}$$

$$c_2 = \frac{||P_{A^T}\nabla g(x^\star)||}{v_{\min}\lambda_{\min,\text{pos}}(U_{01}A^TAU_{01}^T)},$$

and, for small enough $\rho_{\text{K}}(Y, y_{\text{exact}})$,

$$\tilde{c}_k = c_k \left[1 - \frac{M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})}{\lambda_{\min\ \text{pos}}(U_{01}A^TAU_{01}^T)v_{\min}}\right]^{-1}, \quad k = 1, 2. \tag{25}$$

**Theorem 4.** *Suppose we assume the Bayesian model defined in Section 4.1, and that the assumptions on $f_y$, $g$ and $\delta$ stated in Section 4.4 hold.*

*Assume that $x^\star$ is an interior point of $\mathcal{X}$ and that $[A^TV_{y_{\text{exact}}}(x^\star)A : B(x^\star)]$ is of full rank.*

*Assume that as $\tau \to 0$, $\lambda_{\min}(H_\nu) \gg \rho_{\text{K}}(Y, y_{\text{exact}})$, where $H_\nu = A^TV_{y_{\text{exact}}}(x^\star)A + \nu B(x^\star)$ and that*

$$\Delta_0^\star \ll \max\left\{\nu,\ \rho_{\text{K}}(Y, y_{\text{exact}}),\ \sqrt{-\frac{\tau}{\lambda_{\min}(H_\nu)}\log\left(\frac{\tau}{\lambda_{\min}(H_\nu)}\right)}\right\}, \tag{26}$$

*where $\Delta_0^\star$ is defined by (39).*

*Then, $\exists \tau_0 > 0$ such that for $\forall\, \tau \in (0, \tau_0]$, and small enough $\nu$ and $\tau/\nu$,*

$$\begin{aligned} \rho_{\text{K}}(\mu_{\text{post}}, \delta_{x^\star}) \leqslant\ & \max\{2\rho_{\text{K}}(Y, y_{\text{exact}}), \tilde{c}_1\rho_{\text{K}}(Y, y_{\text{exact}}) + \tilde{c}_2\nu \\ +\ & \left[-\frac{\tau}{\lambda_{\min}(H_\nu)}\log\left(C_p\left(\frac{\tau}{\lambda_{\min}(H_\nu)}\right)^{\kappa_p}\right)\right]^{1/2}(1 + \Delta_{\star,K}(\delta, 0))\right\}, \end{aligned} \tag{27}$$

*where $\tilde{c}_1$ and $\tilde{c}_2$ defined by (25) with $U_{01} = I$, $\Delta_{\star,K}(\delta, p_2)$ is defined by (41).*

*Under the assumptions on $\tau$, $\nu$ and $\delta$ given in Section 4.4, $\Delta_{\star,K}(\delta, 0) = o(1)$ as $\tau \to 0$.*

Assumption $\lambda_{\min}(H_\nu) \gg \rho_{\text{K}}(Y, y_{\text{exact}})$ is necessary so that $\lambda_{\min}(H_{Y(\omega)}(x^\star))$ can be bounded from below by a positive value with high probability. In the well-posed case, this holds with high probability; in the ill-posed case, this holds if $\nu \gg \rho_{\text{K}}(Y, y_{\text{exact}})$. In the Gaussian or Poisson cases, where $\rho_{\text{K}}(Y, y_{\text{exact}}) = \sqrt{\tau}$, this means that we assume that $\sqrt{\tau}/\gamma^2 \to \infty$.

**Theorem 5.** *Suppose we assume the Bayesian model defined in Section 4.1, and that the assumptions on $f_y$ and $g$ stated in Section 4.4 hold.*

*Denote $\lambda_{\min,01} = \lambda_{\min,\,\text{U}_{01}}(H_\nu)$, where $H_\nu = A^TV_{y_{\text{exact}}}(x^\star)A + \nu B(x^\star)$.*

*Assume that*

- $U_{01}[A^TV_{y_{\text{exact}}}(x^\star)A : B(x^\star)]U_{01}^T$ *is of full rank,*

- *prior dispersion $\gamma^2$ satisfies $\frac{\tau^3}{\gamma^8} \to 0$,*
- *for any $a \in (0,1)$ that may depend on $\tau$ and $\gamma$,*

$$\Delta_0^\star \ll \max\left\{\nu,\ \sqrt{-\frac{\tau}{\lambda_{\min,01}}\log\left(\frac{\tau\, a^2}{\lambda_{\min,01}}\right)},\ -\frac{\tau}{\nu}\log\left(\frac{\tau}{\nu\sqrt{1-a^2}}\right)\right\},$$

*where $\Delta_0^\star$ is defined by (39).*

*Then, for small enough $\tau$,*

$$\begin{aligned}\rho_{\mathrm{K}}(\mu_{\mathrm{post}}, \delta_{x^\star}) \ &\leqslant\ \max\left\{2\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}),\quad \frac{\tilde{c}_1\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \tilde{c}_2\nu}{\lambda_{\min,\mathrm{pos}}(U_{01}A^T V_{y_{\mathrm{exact}}}(x^\star) A U_{01}^T)}\right. \\ &+\ \sqrt{-\frac{\tau}{\lambda_{\min,01}}\left(\kappa_p \log\left(\frac{\tau\, a^2}{\lambda_{\min,01}}\right) + \log C_p\right)}(1+\Delta_{\star,K}(\delta, p_2)) \\ &-\ \left.\frac{\gamma^2}{b_{\min}}\log\left(\frac{\gamma^2 p_2}{b_{\min}\sqrt{1-a^2}}\right)(1+\Delta_{\star\star,K}(\delta, p_2))\right\}, \qquad (28)\end{aligned}$$

*which holds for any $a \in (0,1)$ where $\Delta_{\star,K}(\delta, p_2)$ and $\Delta_{\star\star,K}(\delta, p_2)$ are defined by (41), $\tilde{c_1}$ and $\tilde{c_2}$ are defined by (25).*

*Under the assumptions on $\tau$, $\nu$ and $\delta$ given in Section 4.4, $\Delta_{\star,K}(\delta, p_2) = o(1)$ and $\Delta_{\star\star,K}(\delta, p_2) = o(1)$ as $\tau \to 0$.*

Hence, in the case that the solution is on the boundary, the competing rates of convergence are the Ky Fan distance for the data $\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})$ and the rate of convergence of the posterior distribution.

Recall that in the ill-posed case (if $U_{01}^T A^T V_{y_{\mathrm{exact}}}(x^\star) A U_{01}$ is not of full rank), $\lambda_{\min,01} \asymp \nu \cdot const$, and in the well-posed case $\lambda_{\min,01} \asymp const$.

## 5.3 Convergence of the data in Ky Fan metric

### 5.3.1 Examples

Now we consider some examples.

**Corollary 2.** *Let the assumptions of Theorem 4 on the prior distribution hold, and $Y_t$ be independent rescaled Poisson random variables, with $\nu = \tau/\gamma^2$, $\tau = \sigma^2$.*

1. *If $x^\star$ is an interior point of $\mathcal{X}$, then, for small enough $\sigma$ and $\gamma$,*

$$\begin{aligned}\rho_{\mathrm{K}}(\mu_{\mathrm{post}}, \delta_{x^\star}) \ &\leqslant\ \left[C_1\sqrt{-\tau\log\tau} + C_2\frac{\tau}{\gamma^2}\right. \\ &+\ \left. C_{3,\alpha}\tau^{(1-\alpha)/2}\gamma^\alpha\sqrt{-\log\left(\tau^{(1-\alpha)/2}\gamma^\alpha\right)}\right](1+o(1)),\end{aligned}$$

*where $\alpha = 0$ if $A^T V_{y_{\text{exact}}}(x^\star)A$ is of full rank and $\alpha = 1$ otherwise, and the constants are given by*

$$
\begin{aligned}
C_1 &= 2||y_{\text{exact}}||_1^{1/2} \max\left(1, \frac{M_{f1}||y_{\text{exact}}||_\infty}{\lambda_{\text{min,pos}}(A^TA)}\right), \\
C_2 &= \frac{||y_{\text{exact}}||_\infty}{\lambda_{\text{min,pos}}(A^TA)}||P_A \nabla g(x^\star)||, \\
C_{3,\alpha} &= \left(\kappa_p[(1-\alpha)\lambda_{\min}(A^TA) + \alpha\lambda_{\min,I-P_A}(B(x^\star))]\right)^{1/2}.
\end{aligned}
$$

*If $\alpha = 0$, the fastest rate is $\sigma\sqrt{-\log\sigma}$, with $\gamma = \sigma^{1/2}[-\log\sigma]^{-1/4}$.*

*If $\alpha = 1$ and $\tau = \sigma^2$, the fastest rate is $\sigma^{2/3}\sqrt{-\log\sigma}$, with $\gamma = \sigma^{2/3}[-\log\sigma]^{-1/3}$.*

2. *If $A^TA$ is not of full rank and $x^\star$ is on the boundary of $\mathcal{X}$, we have an additional term of order $-\gamma^2\log(c_3\gamma^2)$.*

### 5.3.2 General inequalities

First we give upper bounds on the Ky Fan distance in terms of the moments of $||Y-\mu||$, and then consider particular cases with independent observations.

**Theorem 6.** *1. If $\exists\alpha > 0$: $\mathbb{E}e^{\alpha||Y-\mu||} < \infty$, then*

$$\rho_{\text{K}}(Y,\mu) \le \mathbb{E}e^{\alpha||Y-\mu||}.$$

2. *If $\exists\alpha > 0$: $\mathbb{E}\left[||Y-\mu||^\alpha\right] < \infty$, then*

$$\rho_{\text{K}}(Y,\mu) \le \left[\mathbb{E}||Y-\mu||^\alpha\right]^{\frac{1}{1+\alpha}}.$$

The proof is obvious, applying the Markov inequality to the corresponding function of $||Y-\mu||$.

Now we evaluate the Ky Fan distance in the following two particular cases: the rescaled Poisson distribution corresponding to the tomography case (Section 2.1), and $Y_t = \mu_t + \sigma Z_t$ where the distribution of $Z_t$ is independent of $\sigma$.

**Example 4.** *A vector of rescaled independent Poisson random variables: $Y_t/\tau \sim Pois(\mu_t/\tau)$.*

*Apply the Chernoff-Cramer bound to obtain that for all $t$ and all $x,\varepsilon > 0$,*

$$\mathbb{P}(||Y-\mu|| > \varepsilon) \leqslant e^{-\varepsilon x}\mathbb{E}e^{x||Y-\mu||} \leqslant e^{-\varepsilon x}\mathbb{E}e^{x||Y-\mu||_1} = e^{-\varepsilon x}\prod_t \mathbb{E}e^{x|Y_t-\mu_t|}$$

*Now, $\mathbb{E}e^{x|Y_t-\mu_t|} \leqslant \mathbb{E}e^{x(Y_t-\mu_t)} + \mathbb{E}e^{-x(Y_t-\mu_t)}$. The cumulant function of a Poisson random variable $Z$ with parameter $\lambda$ is $\log\mathbb{E}e^{\varepsilon Z} = \lambda[e^\varepsilon - 1]$; hence, for $Y_t = \sigma^2\tau Z$ and $\lambda = \mu_t/\tau$, the cumulant function of $Y_t - \mu_t$ is*

$$c_t(x) \;=\; \log\mathbb{E}e^{x(Y_t-\mu_t)} = \log\mathbb{E}e^{x\tau Z} - x\mu_t = \frac{\mu_t}{\tau}[e^{x\tau} - 1 - x\tau].$$

*Hence, the cumulants of the rescaled Poisson distribution are* $\kappa_k = \mu_t \sigma^{2(k-1)}$*. Similarly,*

$$\log \mathbb{E} e^{-x(Y_t-\mu_t)} = \frac{\mu_t}{\tau}[e^{-x\tau} - 1 + x\tau] \leqslant c_t(x) \quad \forall x > 0.$$

*Hence, denoting* $M = 2\sum_t \mu_t$*, we have*

$$\mathbb{P}(||Y - \mu|| > \varepsilon) \leqslant e^{-\varepsilon x} e^{2\sum_t c_t(x)} = \exp\{-\varepsilon x + M[e^{x\tau} - 1 - x\tau]/\tau\}.$$

*Since* $x > 0$ *is arbitrary, we can take* $x$ *corresponding to the minimum of the upper bound, which is achieved at* $x = \tau^{-1}\log(1 + \varepsilon/M)$*, implying*

$$\mathbb{P}(||Y - \mu|| > \varepsilon) \leqslant \exp\left\{-\frac{\varepsilon + M}{\tau}\log\left(1 + \frac{\varepsilon}{M}\right) + \frac{\varepsilon}{\tau}\right\} \leqslant \exp\left\{-\frac{\varepsilon^2}{2M\tau}\left(1 - \frac{\varepsilon}{3M}\right)\right\},$$

*due to the inequality* $(1 + x)\log(1 + x) - x \geqslant -\frac{x^2}{2}(1 - \frac{x}{3})$ *for small enough* $x > 0$*. For* $\varepsilon \leqslant 3M/2$ *we have*

$$\mathbb{P}(||Y - \mu|| > \varepsilon) \leqslant \exp\left\{-\frac{\varepsilon^2}{4M\tau}\right\}.$$

*Using Lemma 4, for* $\tau \leqslant 1/(2eM)$*, the solution of* $\exp\{-\varepsilon^2/(4M\tau)\} = \varepsilon$ *satisfies*

$$\varepsilon = \sqrt{-2\tau M \log(2\tau M)}(1 + \omega),$$

*where* $\omega = o(1)$ *as* $\sigma \to 0$ *and* $\omega \leqslant 0$*.*

**Theorem 7.** *Assume that* $Y_t$ *are independent,* $\mathbb{E}Y_t = \mu_t$ *and* $\mathrm{Var}(\mathrm{Y_t}) = \mathrm{w_t}\tau$*.*

1. *Assume that* $\exists C_t \geqslant 1$ *such that* $\kappa_{t,k}$*, the kth cumulant of* $Y_t$*, is bounded by* $|\kappa_{t,k}| \leqslant C_t w_t \tau^{k-1} \quad \forall k > 2$ *and* $C_t$ *and* $w_t$ *are independent of* $\tau$*. Denote* $M = 2\sum_t C_t w_t$*.*

   *Then, for* $\tau \leqslant 1/(2eM)$*,*

$$\rho_{\mathrm{K}}(Y, \mu) \leqslant 2\sqrt{-\tau M \log(2\tau M)/2}.$$

2. *Assume that* $\exists K \geqslant 2$*:* $\mathbb{E}|Y_t|^K < \infty$ *and* $\forall \delta > 0$ $\mathbb{E}|Y_t|^{K+\delta} = \infty$*. Assume that* $\mathbb{E}|Y_t - \mu_t|^K \leqslant \tau^{m(K)} L_K$ *for some* $L_K > 0$ *that may depend on* $\mu_t$ *or* $w_t$ *but not on* $\tau$*, for some* $m(K) > 0$*.*

   *Then, for small enough* $\tau$*,*

$$\rho_{\mathrm{K}}(Y, \mu) \leqslant [n\tau^{m(K)/2} L_K]^{1/(K+1)}.$$

*Proof.* 1. Following the rescaled Poisson example, we have that the cumulant function for $Y_t$ is bounded by

$$\begin{aligned} c_t(x) = \log \mathbb{E} e^{xY_t} &= x\mu_t + \frac{x^2}{2} w_t \tau + \sum_{i=3}^{\infty} \frac{x^k}{k!} \kappa_k \leqslant x\mu_t + \frac{x^2}{2} w_t \tau + \frac{1}{\tau} \sum_{i=3}^{\infty} \frac{(x\tau)^k}{k!} C_t w_t \\ &= x\mu_t + \frac{x^2}{2} w_t \tau + \frac{C_t w_t}{\tau} [e^{x\tau} - 1 - x\tau - (x\tau)^2/2] \\ &\leqslant x\mu_t + \frac{C_t w_t}{\tau} [e^{x\tau} - 1 - x\tau], \end{aligned}$$

since $C_t \geqslant 1$. Similarly, $\log \mathbb{E} e^{xY_t}$ can be bounded in the same way. Hence, we have

$$\mathbb{P}(||Y - \mu|| > \varepsilon) \leqslant e^{-\varepsilon x} e^{2 \sum_t c_t(x)} = \exp\{-\varepsilon x + \frac{M}{\tau} [e^{x\tau} - 1 - x\tau]\}.$$

where $M = 2 \sum_t C_t w_t$. Now, this is the same upper bound as for the rescaled Poisson distribution. Hence, we have the same inequality for the Ky Fan distance.

2. Apply the Markov inequality to the random variable $||Y - \mu||^K$:

$$\mathbb{P}(||Y - \mu|| > z) \leqslant \frac{\mathbb{E}||Y - \mu||^K}{z^K} \leqslant \frac{\mathbb{E}||Y - \mu||_K^K}{z^K} \leqslant \frac{n\tau^{m(K)/2} L_K}{z^K}.$$

Hence, an upper bound on the Ky Fan distance satisfies $n\tau L_K / z^K = z$, i.e. $z = [n\tau^{m(K)/2} L_K]^{1/(K+1)}$.

□

The conditions in the first case are satisfied, for example, for the binomial distribution $Y_t \sim Bin(n_t, p_t)$, independently, since $c_t(x) = n_t \log(p_t e^x + q_t) \leqslant n_t p_t (e^x - 1)$.

Here is an example for the second case.

**Example 5.** *Suppose we have $Y_t$ following a $t$ distribution with $\nu$ degrees of freedom, means $\mu_t$ and scales $\sqrt{\tau} w_t$. Then we can take $K = \nu - 2 - \delta$ for some $\delta > 0$;*

$$\mathbb{E}|Y_t - \mu_t|^K = [\sqrt{\tau} w_t]^K \nu_K,$$

*where $\nu_K$ is the $K$th moment of the standard $t_\nu$ distribution, i.e. $m(K) = K/2$ and $L_K = w_t^K \nu_K$. Hence,*

$$\rho_{\mathrm{K}}(Y, \mu) \leqslant \tau^{1/2 - 1/2(K+1)} [n w_t^K \nu_K]^{1/(K+1)}.$$

*Note that this bound holds if $Y_t$ can be written as $Y_t = \mu_t + \sigma w_t Z_t$ where $Z_t$ are iid and whose distribution is independent of $\tau$.*

# 6 Approximation of the posterior distribution

For completeness, we shall also show how the posterior distribution can be rescaled so that it converges to a finite limit. This can be used to approximate the posterior distribution in practice, for small values of $\tau$.

For a differentiable identifiable likelihood and prior distribution positive and continuous at the "true" value of the parameter, the posterior distribution is asymptotically Gaussian in the case where the "true" parameter is an interior point of the parameter space. This result is known as the Bernstein–von Mises theorem. Van der Vaart (1998) gives a total variation distance version of the theorem under mild additional assumptions on the error model, adapted from Le Cam (1953) and Le Cam and Yang (1990). The theorem in fact implies that, under the above conditions, the prior distribution has no influence on the asymptotic distribution.

We extend the Bernstein–von Mises theorem in two directions. Firstly, the assumption of identifiability of the likelihood is relaxed; a consequence is that the limit of the posterior distribution, as well as the rate of convergence, depend on the choice of the prior distribution. Secondly, the assumption that the "true" value of the parameter is an interior point of the parameter space is relaxed, by assuming that it can lie on the boundary. In the latter case, we show that the limiting distribution changes from Gaussian to a product of Gaussian and exponential in different directions.

Now we make use of the three-part transformation of the re-centered variable $x - x^\star$ from Section 4.3: $w = U^{-1}v = U^{-1}(x - x^\star)$. Define the following scaling transform $\mathcal{S} = \mathcal{S}_{\tau,\gamma}$: $\mathcal{X} - x^\star \to \mathbb{R}^{p_0+p_1} \times \mathbb{R}_+^{p_2}$: $\mathcal{S} = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2)$, with

$$\begin{aligned} \mathcal{S}_0 &= (U^{-1})_0(x - x^\star)/\sqrt{\tau}, \\ \mathcal{S}_1 &= (U^{-1})_1(x - x^\star)/\gamma, \\ \mathcal{S}_2 &= (U^{-1})_2(x - x^\star)/\gamma^2. \end{aligned} \tag{29}$$

Denote the posterior distribution of $\theta = \mathcal{S}(x - x^\star)$ given $Y(\omega)$ by $\tilde{\mu}_{\text{post}}(\omega)$.

The limiting distribution is defined in terms of the following parameters:

$$\begin{aligned} \Omega_{00} &= U_0 \nabla^2 f_{y_{\text{exact}}}(x^\star) U_0^T, \\ B &= \nabla^2 g(x^\star), \end{aligned}$$

and $B_{ij} = U_i^T B U_j$, $i, j = 0, 1, 2$.

Now we can formulate a version of the Bernstein – von Mises theorem for this problem.

**Theorem 8.** *Consider the Bayesian model defined in Section 4.1, and let the assumptions on $f_y$ and $g$ stated in Section 4.4 hold. Assume that condition (19) holds.*

*Assume that matrices $\Omega_{00}$ and $B_{11}$ are of full rank, that $B_{00} - B_{01}B_{11}^{-1}B_{10} \geq 0$, and that*

$$\frac{\tau}{\gamma^2} \to 0, \quad \gamma^2 = o(\tau^{1/3}),$$

$$c = \lim_{\tau \to 0} \frac{\sqrt{\tau}}{\gamma^2} < \infty, \quad \lim_{\tau \to 0} \frac{\text{Var(Y)}}{\tau} < \infty.$$

*Assume that the following limit exists for all* $\omega$*:* $\lim_{\tau\to 0}[U_0^T \nabla f_{Y(\omega)}(x^\star)/\sqrt{\tau}] < \infty$*, and denote*

$$a_0(\omega) = \Omega_{00}^{-1}\left[\lim_{\tau\to 0}[U_0^T \nabla f_{Y(\omega)}(x^\star)/\sqrt{\tau}] + cU_0^T \nabla g(x^\star)\right].$$

*Denote by* $\mu^\star$ *the following measure on* $\mathbb{R}^{p_0+p_1} \times \mathbb{R}_+^{p_2}$*:*

$$\mu^\star(\omega) = \mathcal{N}_{p_0}\left(a_0(\omega), \Omega_{00}^{-1}\right) \times \mathcal{N}_{p_1}\left(0, B_{11}^{-1}\right) \times Exp_{p_2}(b). \tag{30}$$

*Then, as* $\tau \to 0$*,*

$$||\mathbb{P}_{\mathcal{S}(x-x^\star)|Y} - \mu^\star||_{TV} \overset{P_{x_{\text{true}}}}{\to} 0 \quad as \quad \tau \to 0.$$

An upper bound on the total variation distance is given in the proof of the theorem.

**Remark 1.** *We assumed that* $\nabla_\eta \tilde{f}_{y_{\text{exact}}}(\eta)|_{G(\eta)=y_{\text{exact}}} = 0$*, i.e. that the likelihood is regular with respect to the "natural" parameter. If this assumption does not hold, we have an additional linear subspace determined by* $A$ *and* $\nabla_\eta \tilde{f}_{y_{\text{exact}}}(G^{-1}(y_{\text{exact}}))$ *where the limit is exponential.*

# A Appendix: proofs

## A.1 Proof of Proposition 1 in Section 3

*Proof.* For arbitrary $\nu > 0$, suppose $x \in \mathbb{R}^p$ is such that $(A^TA + \nu B)x = 0_p$, the zero vector in $\mathbb{R}^p$. We have to show that $x = 0_p$. But $(A^TA + \nu B)x = 0_p$ implies $x^T(A^TA + \nu B)x = 0$, and so by non-negative-definiteness of $B$ and $A^TA$, $x^TBx = 0 = x^TA^TAx$. But then $Bx = 0_p = A^TAx$, and so $x^T[B : A^TA] = 0_{2p}^T$. By the assumed full rank of this matrix, $x$ must be $0_p$.

Now fix $\nu_0 > 0$. By Theorem 2 of **?**, page 313, (with his $A$ replaced by $A^TA+\nu_0 B$), there exists a nonsingular real matrix $P$, not necessarily orthogonal, such that $P^T(A^TA+\nu_0 B)P = I$ and $P^TBP$ is the diagonal matrix $\Lambda$ of the solutions for $\lambda$ to $|B - \lambda(A^TA + \nu_0 B)| = 0$ (which all satisfy $0 \leq \lambda \leq \nu_0^{-1}$). But then $P^TA^TAP = I - \nu_0\Lambda$ and for any $\nu > 0$, $P^T(A^TA + \nu B)P = I + (\nu - \nu_0)\Lambda$, both of which are of course also diagonal.

The matrix $P$ can depend on the choice of $\nu_0$, but evidently always diagonalises $A^TA$, $B$ and any linear combination. Also, $\Lambda$ depends on $\nu_0$, but since $|B - \lambda(A^TA + \nu_0 B)| = (1 - \lambda\nu_0)^p|B - \alpha A^TA|$ where $\alpha = \lambda/(1 - \lambda\nu_0)$, the (diagonal) elements in $\Lambda$ are $\lambda_i = \alpha_i/(1 + \alpha_i\nu_0)$ where $\{\alpha_i\}$ are the solutions to $|B - \alpha A^TA| = 0$ (possibly some $\alpha_i = +\infty$). So $P^TBP = \text{diag}(\alpha_i/(1 + \alpha_i\nu_0))$, $P^TA^TAP = \text{diag}(1/(1 + \alpha_i\nu_0))$ and for any $\nu$, $P^T(A^TA + \nu B)P = \text{diag}((1 + \alpha_i\nu)/(1 + \alpha_i\nu_0))$. For the final assertions, note that $\nu(A^TA + \nu B)^{-1} = P\text{diag}(\nu(1 + \alpha_i\nu_0)/(1 + \alpha_i\nu))P^T$, which converges to $P\text{diag}(\delta_i)P^T = C$, say, where $\delta_i = \nu_0$ if $\alpha_i = +\infty$, and 0 otherwise. Further, we can estimate the difference: $\nu(A^TA + \nu B)^{-1} - C = P\text{diag}(\nu(1+\alpha_i\nu_0)/(1+\alpha_i\nu) - \delta_i)P^T = \nu P\text{diag}(\phi_i)P^T + O(\nu^2)$, where $\phi_i = 0$ if $\alpha_i = +\infty$ and otherwise $\phi_i = 1 + \alpha_i\nu_0$. □

## A.2 Proofs of the results in Section 5

**Proposition 2.** *Let assumptions on $f_y$, $g$ in Section 4.4 and assumptions (18) on $\delta$ hold.*

*Assume that $U_{01}[A^T V_{y_{\text{exact}}}(x^\star)A : B(x^\star)]U_{01}^T$ is of full rank, and that $\tau/\nu \to 0$ and $\nu \to 0$ as $\tau \to 0$.*

*Then, for any $\varepsilon \in (c_1\rho_{\mathrm{K}}(Y, y_{\text{exact}}) + c_2\nu, \delta)$ such that $\sqrt{1-\beta}\,\varepsilon\nu/\tau \to \infty$, for any $\beta \in [0,1]$*

$$\begin{aligned} &\frac{\int_{\mathcal{X}\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{\mathcal{X}} e^{-h_y(x)/\tau}dx} \\ \leqslant\ & \left[1-\Gamma\left(\frac{(\varepsilon\sqrt{\beta}-||\bar{H}H_{01,01}^{-1}HH_{01,01}x_0||)^2\lambda_{\min}(\bar{H}H_{01,01})}{2\tau} \mid \frac{p_{01}}{2}\right)\right. \\ & \left. + p_2\exp\left\{-\frac{\varepsilon\sqrt{1-\beta}\nu b_{\min}}{\tau}\right\}(1+\Delta_1)\right]\frac{1+\Delta_2}{1+\Delta_0} + \frac{\Delta_0}{1+\Delta_0}, \end{aligned}$$

*and, in particular,*

$$\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx \geq \frac{\tau^{(p+p_2)/2}}{\nu^{p_2}[\det(HH_{01,01})]^{1/2}}\frac{[2\pi]^{p_{01}/2}\det(U)}{\prod_i b_i}\exp\left\{\frac{x_0^T HH_{01,01}x_0 - 2h_y(x^\star)}{2\tau}\right\}[1+\Delta_3],$$

*where $\Delta_1$ and $\Delta_2$ are defined by*

$$\Delta_1(\varepsilon,\delta) = \frac{1-\prod_{i=1}^{p_2}(1-e^{-(b_i-\delta\lambda_{\max}(P_1B(x^\star)))\sqrt{1-\beta}\,\varepsilon\nu/\tau})}{p_2 e^{-(b_{\min}-\delta\lambda_{\max}(U_2B(x^\star)))\sqrt{1-\beta}\,\varepsilon\nu/\tau})} I(p_2\neq 0), \tag{31}$$

$$\Delta_2(\delta,p_2,y) = -1+\prod_i\left[\frac{1+b_i^{-1}\delta[\lambda_{\max}(U_2B(x^\star))+2\delta\sqrt{p\,p_2}C_{g3}/3]}{1-b_i^{-1}\delta[\lambda_{\max}(U_2B(x^\star))+2\delta\sqrt{p\,p_2}C_{g3}/3]}\right] \tag{32}$$

$$\times\ \left[\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}H_{01,01})[\delta-||\widetilde{H}H_{01,01}^{-1}HH_{01,01}x_0||]^2}{2(1-\gamma)\tau}\mid\frac{p_{01}}{2}\right)\right]^{-1}\left[\frac{\det(HH_{01,01}+\delta\sqrt{p}D)}{\det(HH_{01,01}-\delta\sqrt{p}D)}\right]^{1/2} \tag{33}$$

$$\times\ \exp\left\{\delta\sqrt{p}x_0^T HH_{01,01}\bar{H}H_{01,01}^{-1}D\tilde{H}H_{01,01}^{-1}HH_{01,01}x_0/\tau\right\}\prod_{i=1}^{p_2}\left[1-e^{-\bar{b}_i\delta\nu/(\tau\sqrt{\gamma p_2})}\right]^{-1},$$

*where $D = diag(p_0(C_{f3}+2\nu C_{g3})I_{p_0}, 2\nu p_1 C_{g3}I_{p_1})/3$ and $\Delta_3$ is defined in the proof of the proposition.*

An upper bound on $\Delta_2$ is given by:

$$\Delta_2(\delta,p_2,y) \leq -1+\left[\frac{1+b_{\min}^{-1}\delta[\lambda_{\max}(U_2B(x^\star))+2\delta\sqrt{p\,p_2}C_{g3}/3]}{1-b_{\min}^{-1}\delta[\lambda_{\max}(U_2B(x^\star))+2\delta\sqrt{p\,p_2}C_{g3}/3]}\right]^{p_2} \tag{34}$$

$$\times\ \left[\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}H_{01,01})[\delta-||\widetilde{H}H_{01,01}^{-1}HH_{01,01}x_0||]^2}{2(1-\gamma)\tau}\mid\frac{p_{01}}{2}\right)\right]^{-1}\left[\frac{1+\delta\sqrt{p}\lambda_{\max}(DHH_{01,01}^{-1})}{1-\delta\sqrt{p}\lambda_{\max}(DHH_{01,01}^{-1})}\right]^{p_{01}/2} \tag{35}$$

$$\times\ \exp\left\{\frac{\delta\sqrt{p}||D^{1/2}x_0||_2^2}{\tau[1-\delta^2 p\lambda_{\max}^2(DHH_{01,01}^{-1})]}\right\}\left[1-\exp\left\{-\frac{\widetilde{b}_{\min}\delta\nu}{\tau\sqrt{\gamma p_2}}\right\}\right]^{-p_2}. \tag{36}$$

Here $\lambda_{\min}(HH_{01,01}D^{-1}) = \min\left\{\frac{3\lambda_{\min,P_{A,V}}(A^TV_y(x^\star)A+\nu B(x^\star))}{C_{f3}+2\nu C_{g3}}, \frac{3\lambda_{\min,I-P_{A,V}}(B(x^\star))}{2C_{g3}}\right\}$.

*Proof of Proposition 2.* Taylor decomposition of $h_y(x)$ at $x^\star$ for $x \in B(x^\star, \delta)$ gives

$$h_y(x) \quad = \quad h_y(x^\star) + (x - x^\star)^T \nabla h_y(x^\star) + \frac{1}{2}(x - x^\star)^T H_y(x^\star)(x - x^\star) + \Delta_{00}(\delta),$$

where

$$\begin{aligned}
|\Delta_{00}(\delta)| &= \frac{1}{6}\left|\sum_{ijk} \nabla_{ijk} h_y(x_c)(x_i - x_i^\star)(x_j - x_j^\star)(x_k - x_k^\star)\right| \\
&\leq \frac{||x - x^\star||_1}{6}\left[C_{f3}|| \left(U^{-1}\right)_0 (x - x^\star)||_1^2 + \nu C_{g3} \sum_{k=0}^{2} || \left(U^{-1}\right)_k (x - x^\star)||_1\right]^2 \\
&\leq \frac{\delta\sqrt{p}}{6}\left[p_0(C_{f3} + 2\nu C_{g3})|| \left(U^{-1}\right)_0 (x - x^\star)||_2^2 + 4\nu p_1 C_{g3}|| \left(U^{-1}\right)_1 (x - x^\star)||_2^2\right. \\
&\quad \left. +4\nu C_{g3}|| \left(U^{-1}\right)_2 (x - x^\star)||_1^2\right]
\end{aligned}$$

for some $x_c \in \langle x, x^\star \rangle$.

Denote $b_0 = HH_{01,01} x_0 = \nabla h_y(x^\star)$. Make a change of variables $v_0 = \left(U^{-1}\right)_{01}(x - x^\star)/\sqrt{\tau}$ and $v_2 = \nu\left(U^{-1}\right)_2 (x - x^\star)/\tau \geq 0$, the Jacobian is $J = \tau^{(p+p_2)/2}\nu^{-p_2}\det(U)$ and we have

$$\begin{aligned}
\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau} dx &\geq J \exp\left\{-\frac{1}{\tau}\left[h_y(x^\star) - ||HH_{01,01}^{1/2} x_0||^2/2\right]\right\} \\
&\times \int_{\tau||v_0||^2 + \tau^2||v_2||^2/\nu^2 \leq \delta^2} \exp\left\{-\frac{1}{2}||HH_{01,01}^{1/2}(v_0 - \tau^{-1/2}x_0)||^2 - b^T v_2\right\} \\
&\times \exp\left\{-\delta\sqrt{p} v_0^T \mathrm{diag}\left(p_0(C_{f3} + 2\nu C_{g3})I_{p_0}, 4\nu p_1 C_{g3} I_{p_1}\right) v_0/6\right\} \\
&\times \exp\left\{-\sqrt{\tau} v_0^T U_{01}^T B(x^\star) U_2 v_2 - \frac{\tau}{2\nu} v_2^T U_2^T B(x^\star) U_2 v_2 - \tau\delta\sqrt{p}\frac{2C_{g3}}{3\nu}||v_2||_1^2\right\} dv_2 dv_0.
\end{aligned}$$

Note that $U_i^T H_y(x^\star) U_2 = \nu U_i^T B U_2$ for any $i$.

For $v_0, v_2$ such that $\tau||v_0||^2 + \tau^2||v_2||^2/\nu^2 \leq \delta^2$,

$$\begin{aligned}
&\sqrt{\tau} v_2^T U_2^T B(x^\star) U_0 v_0 + \frac{\tau}{2\nu} v_2^T U_2^T B(x^\star) U_2 v_2 \leq \lambda_{\max}(U_2 B(x^\star))||v_2|| \left[\frac{\tau^2}{4\nu^2}||v_2||^2 + \tau||v_0||^2\right]^{1/2} \\
&\leq \lambda_{\max}(U_2 B(x^\star))||v_2||_1 \delta.
\end{aligned}$$

Denoting

$$\begin{aligned}
\widetilde{H}H_{01,01} &= HH_{01,01} + \delta\sqrt{p} D, \\
\widetilde{b} &= b + \delta\lambda_{\max}(U_2 B(x^\star)) + \delta^2\sqrt{pp_2}\frac{2C_{g3}}{3},
\end{aligned}$$

we can bound the integral above by

$$\begin{aligned}
\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau} dx &\geq J \exp\left\{-\frac{1}{\tau}\left[h_y(x^\star) - ||\widetilde{H}H_{01,01}^{-1/2} b_0||^2/2\right]\right\} \\
&\times \int_{\tau||v_0||^2 + \tau^2||v_2||^2/\nu^2 \leq \delta^2} \exp\left\{-\frac{1}{2}||\widetilde{H}H_{01,01}^{1/2}(v_0 - \tau^{-1/2}\widetilde{H}H_{01,01}^{-1} b_0)||^2 - \widetilde{b}^T v_2\right\} dv_2 dv_0.
\end{aligned}$$

Hence,

$$
\begin{aligned}
\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx &\geq J\exp\left\{-\frac{1}{2\tau}\left[2h_y(x^\star) - ||\widetilde{H}H_{01,01}^{-1/2}b_0||^2\right]\right\}\\
&\times \int_{\tau||v_0||^2+\tau^2||v_2||^2/\nu^2\leq\delta^2} \exp\left\{-\widetilde{b}^T v_2\right\}\\
&\times \exp\left\{-\frac{1}{2}(v_0 - \tau^{-1/2}\widetilde{H}H_{01,01}^{-1}b_0)^T\widetilde{H}H_{01,01}(v_0 - \tau^{-1/2}\widetilde{H}H_{01,01}^{-1}b_0)\right\}dv_2dv_0\\
&\geq J\exp\left\{-\frac{2h_y(x^\star) - ||\widetilde{H}H_{01,01}^{-1/2}b_0||^2}{2\tau}\right\}\prod_i \widetilde{b}_i^{-1}[2\pi]^{p_{01}/2}[\det(\widetilde{H}H_{01,01})]^{-1/2}\\
&\times \prod_{i=1}^{p_2}\left[1 - e^{-\widetilde{b}_i\delta\nu/(\tau\sqrt{\alpha p_2})}\right]\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}H_{01,01})[\delta - ||\widetilde{H}H_{01,01}^{-1}b_0||]^2}{2(1-\alpha)\tau}\mid \frac{p_{01}}{2}\right)\\
&= J\exp\left\{-\frac{2h_y(x^\star) - x_0^T HH_{01,01}x_0}{2\tau}\right\}\prod_i b_i^{-1}[2\pi]^{p_{01}/2}[\det(HH_{01,01})]^{-1/2}[1+\Delta_3]
\end{aligned}
$$

for some $\alpha \in [0,1]$, since $\delta\nu/\tau \to \infty$, $\lambda_{\min}(HH_{01,01})\delta^2/\tau \to \infty$, $\delta \gg ||x_0||$ with high $\mathbb{P}_{y_{\text{exact}}}$ probability. Here $\Delta_3$ is defined by

$$
\begin{aligned}
\Delta_3(\delta, p_2, y) &= -1 + \prod_{i=1}^{p_2}\left[1 + \delta\frac{\lambda_{\max}(U_2B(x^\star)) + 2\delta\sqrt{p\,p_2}C_{g3}/3}{b_i}\right]^{-1}\\
&\times \exp\left\{-\delta\sqrt{p}\,b_0^T HH_{01,01}^{-1}D\widetilde{H}H_{01,01}^{-1}b_0/(2\tau)\right\}\prod_{i=1}^{p_2}\left[1 - e^{-\widetilde{b}_i\delta\nu/(\tau\sqrt{\alpha p_2})}\right]^{-1}\\
&\times \left[\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}H_{01,01})[\delta - ||\widetilde{H}H_{01,01}^{-1}b_0||]^2}{2(1-\alpha)\tau}\mid\frac{p_{01}}{2}\right)\right]^{-1}\left[\det(I + \delta\sqrt{p}HH_{01,01}^{-1}D)\right]^{-1/2}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} & e^{-h_y(x)/\tau}dx \leq J\exp\left\{-h_y(x^\star)/\tau + ||HH_{01,01}^{1/2}x_0||^2/(2\tau)\right\}\\
\times\quad & \int_{\varepsilon^2\leq\tau||v_0||^2+\tau^2||v_2||^2/\nu^2\leq\delta^2}\exp\left\{-\frac{1}{2}||HH_{01,01}^{1/2}(v_0 - \tau^{-1/2}x_0)||^2 - b^T v_2\right\}\\
\times\quad & \exp\left\{\frac{\delta p_{01}\sqrt{p}v_0^T D v_0}{2} - \sqrt{\tau}v_0^T U_{01}^T B(x^\star)U_2v_2 - \frac{\tau}{2\nu}v_2^T U_2^T B(x^\star)U_2 v_2\right\}\\
\times\quad & \exp\left\{\tau\delta\sqrt{p}\frac{2C_{g3}}{3\nu}||v_2||_1^2\right\}dv_2dv_0\\
\leq\quad & J\exp\left\{-\frac{1}{\tau}\left[h_y(x^\star) - ||\bar{H}H_{01,01}^{-1/2}b_0||^2/2\right]\right\}\\
\times\quad & \int_{\varepsilon^2\leq\tau||v_0||^2+\tau^2||v_2||^2/\nu^2\leq\delta^2}\exp\left\{-\frac{1}{2}||\bar{H}H_{01,01}^{1/2}(v_0 - \tau^{-1/2}\bar{H}H_{01,01}^{-1}b_0)||^2 - \bar{b}^T v_2\right\}dv_2dv_0,
\end{aligned}
$$

where

$$\begin{aligned}
\bar{H}H_{01,01} &= HH_{01,01} - \delta\sqrt{p}D, \\
\bar{b} &= b - \delta\lambda_{\max}(U_2 B(x^\star)) - \delta^2\sqrt{p\,p_2}\frac{2C_{g3}}{3}.
\end{aligned}$$

Assume that $\delta$ is small enough so that $\bar{b}_i > 0$ for all $i$ and $\bar{H}H_{01,01}$ is positive definite.

The ring $\{\varepsilon^2 \leq \tau||v_0||^2 + \tau^2||v_2||^2/\nu^2 \leq \delta^2\}$ is a subset of

$$\{||v_0||^2 \leq \beta\varepsilon^2/\tau \,\&\, ||v_2||^2 \geq (\varepsilon^2 - \tau||v_0||^2)\nu^2/\tau^2\} \cup \{\beta\varepsilon^2/\tau \leq ||v_0||^2\}$$

for any $\beta \in (0,1)$, therefore

$$\begin{aligned}
& \int_{\varepsilon^2 \leq \tau||v_0||^2 + \tau^2||v_2||^2/\nu^2 \leq \delta^2} \exp\left\{-\bar{b}^T v_2 - ||\bar{H}H_{01,01}^{1/2}(v_0 - \tau^{-1/2}\bar{H}H_{01,01}^{-1}b_0)||^2/2\right\} dv_2 dv_0 \\
\leq\ & [2\pi]^{p_0/2}[\det(\bar{H}H_{01,01})]^{-1/2} \int_{||v_2||^2 \geq \varepsilon^2(1-\beta)\nu^2/\tau^2} \exp\left\{-\bar{b}^T v_2\right\} dv_2 \\
+\ & \prod_i [\bar{b}_i]^{-1} \int_{||v_0||^2 \geq \beta\varepsilon^2/\tau} \exp\left\{-\frac{1}{2}||\bar{H}H_{01,01}^{1/2}(v_0 - \tau^{-1/2}\bar{H}H_{01,01}^{-1}b_0)||^2\right\} dv_0.
\end{aligned}$$

The integral over the exponential density can be bounded by

$$\begin{aligned}
& \int_{||v_2||^2 \geq \varepsilon^2(1-\beta)\nu^2/\tau^2} \exp\left\{-\bar{b}^T v_2\right\} dv_2 \leq \int_{||v_2||_\infty \geq \varepsilon\sqrt{1-\beta}\nu/\tau} \exp\left\{-\bar{b}^T v_2\right\} dv_2 \\
=\ & \prod_i [\bar{b}_i]^{-1}[1 - \prod_{i=1}^{p_2}(1 - e^{-\bar{b}_i\sqrt{1-\beta}\,\varepsilon\nu/\tau})] = \prod_i [\bar{b}_i]^{-1} p_2 e^{-b_{\min}\sqrt{1-\beta}\varepsilon\nu/\tau}(1+\Delta_1),
\end{aligned}$$

since $\sqrt{1-\beta}\,\varepsilon\nu/\tau \to \infty$ as $\tau \to 0$, where

$$\begin{aligned}
\Delta_1(\varepsilon,\delta,p_2) &= -1 + \frac{1 - \prod_{i=1}^{p_2}(1 - e^{-\bar{b}_i\sqrt{1-\beta}\,\varepsilon\nu/\tau})}{p_2 e^{-b_{\min}\sqrt{1-\beta}\,\varepsilon\nu/\tau}} \\
&\leq -1 + \frac{1 - (1 - e^{-\bar{b}_{\min}\sqrt{1-\beta}\,\varepsilon\nu/\tau})^{p_2}}{p_2 e^{-b_{\min}\sqrt{1-\beta}\,\varepsilon\nu/\tau}} \\
&= \exp\left\{\delta\sqrt{1-\beta}\,\varepsilon\nu/\tau\left[\lambda_{\max}(U_2 B(x^\star)) + \delta\sqrt{p\,p_2}\frac{2C_{g3}}{3}\right]\right\} - 1 \\
&+ \sum_{k=2}^{p_2}\frac{1}{p_2}\binom{p_2}{k}(-1)^{k-1}\exp(-[k\bar{b}_{\min} - b_{\min}]\sqrt{1-\beta}\,\varepsilon\nu/\tau).
\end{aligned}$$

Hence, $\Delta_1$ is small if $\sqrt{1-\beta}\,\varepsilon\nu/\tau \to \infty$ and $\delta\sqrt{1-\beta}\,\varepsilon\nu/\tau \to 0$ as $\tau \to 0$.

Combining these results together, we have

$$\begin{aligned}
\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} &\le \Big[p_2 e^{-b_{\min}\sqrt{1-\beta}\,\varepsilon\nu/\tau}(1+\Delta_1) \\
&+ 1-\Gamma\left(\frac{\lambda_{\min}(\bar{H}H_{01,01})[\sqrt{\beta}\varepsilon - ||\bar{H}H_{01,01}^{-1}b_0||]^2}{2\tau} \mid \frac{p_0}{2}\right)\Bigg] \\
&\times \prod_i \left[\frac{1+b_i^{-1}\delta[\lambda_{\max}(U_2 B(x^\star)) + 2\delta\sqrt{p\,p_2}C_{g3}/3]}{1-b_i^{-1}\delta[\lambda_{\max}(U_2 B(x^\star)) + 2\delta\sqrt{p\,p_2}C_{g3}/3]}\right] \\
&\times \left[\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}H_{01,01})[\delta - ||\widetilde{H}H_{01,01}^{-1}b_0||]^2}{2(1-\alpha)\tau} \mid \frac{p_0}{2}\right)\right]^{-1}\left[\frac{\det(\widetilde{H}H_{01,01})}{\det(\bar{H}H_{01,01})}\right]^{1/2} \\
&\times \exp\left\{\delta\sqrt{p}\,b_0^T\bar{H}H_{01,01}^{-1}D\widetilde{H}H_{01,01}^{-1}b_0/\tau\right\}\prod_{i=1}^{p_2}\left[1-e^{-\bar{b}_i\delta\nu/(\tau\sqrt{\alpha p_2})}\right]^{-1}.
\end{aligned}$$

Now we take into account the error of approximating the integral over $\mathcal{X}$ by the integral over $B(x^\star,\varepsilon)$:

$$\begin{aligned}
\frac{\int_{\mathcal{X}\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{\mathcal{X}} e^{-h_y(x)/\tau}dx} &= \frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx + \int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx + \int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} \\
&= \frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{(1+\Delta_0)\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} + \frac{\Delta_0}{1+\Delta_0}.
\end{aligned}$$

Choosing some fixed $\alpha$, e.g. $\alpha = 1/2$, we have the required statement.

□

*Proof of Theorems 1 and 2.* By Strassen's theorem, for any $x$, $\rho_{\mathrm{P}}(\mu_{\mathrm{post}}(\omega),\delta_x) = \rho_{\mathrm{K}}(\xi,x)$ where $\xi \sim \mu_{\mathrm{post}}(\omega)$. Hence, we find an upper bound on the Ky Fan distance between $X \mid Y$ and $x^\star$.

Denote $\varepsilon_0 = \sqrt{\beta}\varepsilon$ and $\varepsilon_1 = \sqrt{1-\beta}\,\varepsilon$ for some $\beta \in (0,1)$. Then, $\varepsilon_0+\varepsilon_1 = \varepsilon(\sqrt{\beta}+\sqrt{1-\beta})$.

Take $\varepsilon_0 > ||x_0||$. Using Proposition 2, we have that an upper bound on $\varepsilon$ satisfies

$$\begin{aligned}
\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} &\leqslant \Bigg[p_2\exp\left\{-\frac{\varepsilon_1\nu b_{\min}}{\tau}\right\}(1+\Delta_1) \\
&\quad +1-\Gamma\left(\frac{(\varepsilon_0 - ||\bar{H}H_{01,01}^{-1}b_0||)^2\lambda_{\min}(\bar{H}H_{01,01})}{2\tau} \mid \frac{p_{01}}{2}\right)\Bigg](1+\widetilde{\Delta}_2)+\widetilde{\Delta}_0 \\
&\le \varepsilon,
\end{aligned}$$

where $\widetilde{\Delta}_0 = \Delta_0/(1+\Delta_0)$ and $\widetilde{\Delta}_2 = (1+\Delta_2)/(1+\Delta_0)-1$. By an assumption of the Theorems, $\Delta_0 \ll c_1\rho_K(Y, y_{\mathrm{exact}}) + c_2\nu$.

In particular, for some $\alpha \in (0,1)$, with $a = \alpha/\sqrt{\beta}$ and $\widetilde{a} = (1-\alpha)/\sqrt{1-\beta}$, take the smallest $\varepsilon > 0$ such that

$$\begin{aligned}
\widetilde{\Delta}_0 &\leq \varepsilon, \\
1 - \Gamma\left(\frac{(\varepsilon_0 - ||\bar{H}H_{01,01}^{-1}b_0||)_1^2 \lambda_{\min}(\bar{H}H_{01,01})}{2\tau} \mid \frac{p_0}{2}\right) &\leq \frac{a\varepsilon_0}{1+\widetilde{\Delta}_2}, \\
p_1 \exp\left\{-\frac{\varepsilon_1 \nu b_{\min}}{\tau}\right\}(1+\Delta_1)(1+\widetilde{\Delta}_2) &\leq \widetilde{a}\varepsilon_1.
\end{aligned}$$

The last two inequalities imply that as $\tau/\lambda_{\min}(HH_{01,01}) \to 0$, $\varepsilon_0 \to 0$ and $\varepsilon_0^2 \lambda_{\min}(HH_{01,01})/\tau \to \infty$. Similarly, as $\nu/\tau \to \infty$, $\varepsilon_1 \to 0$. Hence, using Lemmas 3 and 4, we have that

$$\begin{aligned}
\varepsilon_0 &\leq ||\bar{H}H_{01,01}^{-1} HH_{01,01} x_0|| + \sqrt{-\frac{\tau}{(1+\widetilde{\Delta}_2)^2 \lambda_{\min}(\bar{H}H_{01,01})} \log\left(C_{p_{01}}\left(\frac{a^2\tau}{\lambda_{\min}(\bar{H}H_{01,01})(1+\widetilde{\Delta}_2)^2}\right)^{\kappa_{p_{01}}}\right)}, \\
\varepsilon_1 &\leq -\frac{\tau}{\nu b_{\min}}\left[\log\left(\frac{\tau p_2}{\widetilde{a}\nu b_{\min}}\right) + \log\left((1+\Delta_1)(1+\widetilde{\Delta}_2)\right)\right].
\end{aligned}$$

By Lemma 7,

$$||x_0|| \leq \frac{M_{f1}||y - y_{\text{exact}}|| + \nu||P_{A^T}\nabla g(x^\star)||}{\lambda_{\min \text{pos}}(A^T V_y(x^\star) A)},$$

which we can substitute into the upper bound for $\varepsilon_0$, and

$$||\bar{H}H_{01,01}^{-1} HH_{01,01}|| = ||(I - \delta\sqrt{p} HH_{01,01}^{-1} D)^{-1}|| = \left[1 - \delta\sqrt{p}\lambda_{\max}(DHH_{01,01}^{-1})\right]^{-1}.$$

Now, the smallest $\varepsilon > 0$ such that $\varepsilon \geq \widetilde{\Delta}_0$ and satisfies the obtained upper bound on $\varepsilon = (\varepsilon_0 + \varepsilon_1)/(\sqrt{\beta} + \sqrt{1-\beta})$ is the upper bound if it is greater than $\widetilde{\Delta}_0$. Otherwise, $\varepsilon \leq 2\widetilde{\Delta}_0$.

The bound on $\varepsilon_0$ with $U_{01} = I$ and $a = 1$ gives the statement of Theorem 1, and adding up the bounds on $\varepsilon_0$ and $\varepsilon_1$ divided by $\sqrt{\beta} + \sqrt{1-\beta}$ gives the statement of Theorem 2, with the errors denoted by

$$\begin{aligned}
\Delta_\star(\delta, p_2, y) &= \frac{1+\Delta_0}{(1+\Delta_2)(a+\sqrt{1-a^2})}\left[1 + 2\frac{\log(1+\Delta_2) - \log(1+\Delta_0)}{\log\left(\lambda_{\min}(\bar{H}H_{01,01})/(a^2\tau)\right)}\right]^{1/2} - 1, \quad (37) \\
\Delta_{\star\star}(\delta, p_2, y) &= \frac{\log\left((1+\Delta_1)(1+\Delta_2)/(1+\Delta_0)\right)}{\log\left(\tau p_2/(\nu b_{\min}\sqrt{1-a^2})\right)}.
\end{aligned}$$

To simplify the expressions, the results are stated with $\alpha = \beta$, making $a = \sqrt{\beta}$ and $\widetilde{a} = \sqrt{1-\beta}$, i.e. $a^2 + \widetilde{a}^2 = 1$.

We assumed that $\varepsilon < \delta$; in particular, the upper bound on $\varepsilon$ is less than $\delta$ if $\nu^4/\tau = \tau^3/\gamma^8 \to 0$ and $-\gamma^4 \log\gamma/\tau \to 0$ (the latter – in the ill-posed case).

□

*Proof of Theorems 4 and 5.* Now we prove Theorems 4 and 5 in the notation defined in the proof of Theorems 1 and 2.

We apply Theorem 3 with $\Omega_1 = \{\omega : ||Y(\omega) - y_{\text{exact}}|| \leq \rho_{\text{K}}(Y, y_{\text{exact}})\}$ and $\Omega_2 = \Omega \setminus \Omega_1$ with $\mathbb{P}(\Omega_2) \leq \rho_{\text{K}}(Y, y_{\text{exact}})$ by the definition of Ky Fan distance, with the bounds given in Theorems 4 and 5 which we modify to depend on $y$ only via $||y - y_{\text{exact}}||$. For small enough $\tau$, the assumption of the theorems that $U_{01}[A^T V_y(x^\star)A : B(x^\star)]U_{01}^T$ is of full rank holds on $\Omega_1$, as we shall show below.

The upper bound depends on $y$ via $||y - y_{\text{exact}}||$, $\lambda_{\min}(U_{01}H_y(x^\star)U_{01}^T)$, $\lambda_{\text{min pos}}(A^T V_y(x^\star)A)$, $\Delta_0$ and $\Delta_2$.

We start bounding the eigenvalues from below. Denote $HH^\star_{01,01} = U_{01}H_{y_{\text{exact}}}(x^\star)U_{01}^T$. Since $[U_0^T(H_y(x^\star) - H_{y_{\text{exact}}}(x^\star)U_0)]_{ij} = [\nabla^2 f_{y_{\text{exact}}}(x^\star) - \nabla^2 f_y(x^\star)]_{ij} \geq -M_{f2}||y - y_{\text{exact}}||$, on $\Omega_1$ we have, by Lemma 8,

$$\lambda_{\min}(\tilde{H}H_{01,01}) \geq \lambda_{\min U_{01}}(H_{y_{\text{exact}}}(x^\star)) - M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})\mathbb{I}(0) - \delta\sqrt{p}\lambda_{\max}(DH^{\star -1}_{01,01}),$$

where $\mathbb{I}(0) = 1$ if the minimum eigenvalue is achieved on the subspace $U_0$, and is zero otherwise. For small enough $\tau$ and $\delta$, on $\Omega_1$ the lower bound is positive.

Similarly, since $A^T V_y(x)A = -\nabla^2 f_y(x)$,

$$||[A^T(V_Y(x^\star) - V_{y_{\text{exact}}}(x^\star))A]_{ij}| \leqslant M_{f2}||Y - y_{\text{exact}}|| \quad \text{for all} \quad i, j,$$

hence $\lambda_{\text{min pos}}(A^T V_Y(x^\star)A) \geq \lambda_{\text{min pos}}(A^T V_{y_{\text{exact}}}(x^\star)A) - M_{f2}||Y - y_{\text{exact}}||$. Hence, since $||\tilde{H}H^{-1}_{01,01}HH_{01,01}|| \leq 1$, we have that

$$\begin{aligned} ||\tilde{H}H^{-1}_{01,01}b_0|| &\leq ||HH^{-1}_{01,01}b_0|| \leq \frac{M_{f1}||Y - y_{\text{exact}}|| + \nu||P_{A^T}\nabla g(x^\star)||}{\lambda_{\text{min pos}}(A^T V_Y(x^\star)A)} \\ &\leq \frac{c_1||Y - y_{\text{exact}}|| + c_2\nu}{1 - M_{f2}||Y - y_{\text{exact}}||/\lambda_{\text{min pos}}(A^T V_{y_{\text{exact}}}(x^\star)A)}, \end{aligned}$$

where $c_1$ and $c_2$ are defined by (23).

To bound $\Delta_2$ on $\Omega_1$, we need to bound below the following expression for $\delta > ||\tilde{H}H^{-1}_{01,01}b_0||$ on $\Omega_1$:

$$\lambda_{\min}(\tilde{H}H_{01,01})[\delta - ||\tilde{H}H^{-1}_{01,01}b_0||]^2 \;\geq\; [\lambda_{\min}(HH^\star_{01,01}) - M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})]\,[\delta - \widetilde{c}_1||Y - y_{\text{exact}}|| - \widetilde{c}_2\nu]^2,$$

where $\widetilde{c}_k = c_k\,[1 - M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})/\lambda_{\text{min pos}}(A^T V_{y_{\text{exact}}}(x^\star)A)]^{-1}$ for $k = 1, 2$.

Note also, that on $\Omega_1$,

$$\begin{aligned} \lambda_{\min}(HH_{01,01}D^{-1}) &= \min\left\{\frac{3\lambda_{\min,P_{A,V}}(A^T V_y(x^\star)A + \nu B(x^\star))}{p_0(C_{f3} + 2\nu C_{g3})}, \frac{3\lambda_{\min,I-P_{A,V}}(B(x^\star))}{2p_1 C_{g3}}\right\} \\ &\geq 3\min\left\{\frac{\lambda_{\min,\text{pos}}(A^T V_{y_{\text{exact}}}(x^\star)A) - M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})}{p_0(C_{f3} + 2\nu C_{g3})}, \frac{\lambda_{\min,I-P^\star_{A,V}}(B(x^\star))}{2p_1 C_{g3}}\right\} \\ &\stackrel{def}{=} \lambda_{DH}. \end{aligned}$$

Hence, on $\Omega_1$,

$$\begin{aligned}
\widetilde{\Delta}_2(\delta, p_2, y) &\leq -1 + \left[\frac{1 + b_{\min}^{-1}\delta[\lambda_{\max}(U_2 B(x^\star)) + 2\delta\sqrt{p\,p_2}C_{g3}/3]}{1 - b_{\min}^{-1}\delta[\lambda_{\max}(U_2^T B(x^\star)) + 2\delta\sqrt{p\,p_2}C_{g3}/3]}\right]^{p_2} \\
&\times \left[\Gamma\left(\frac{[\lambda_{\min}(HH^\star_{01,01}) - M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})]\,[\delta - \widetilde{c}_1||Y - y_{\mathrm{exact}}|| - \widetilde{c}_2\nu]^2}{\tau} \mid \frac{p_{01}}{2}\right)\right]^{-1} \\
&\times \left[\frac{1 + \delta\sqrt{p}/\lambda_{DH})}{1 - \delta\sqrt{p}/\lambda_{DH})}\right]^{p_{01}/2} \left[1 - \exp\left\{-\frac{\widetilde{b}_{\min}\delta\nu}{\tau\sqrt{p_2/2}}\right\}\right]^{-p_2} \\
&\times \exp\left\{\frac{\delta\sqrt{p}(C_{f3} + 2\nu C_{g3})[\widetilde{c}_1\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \widetilde{c}_2\nu]^2}{3\tau[1 - \delta^2 p/\lambda_{DH}^2]}\right\}[1 + \Delta_0^\star]^{-1} \\
&\overset{def}{=} \Delta_2^\star. \qquad (38)
\end{aligned}$$

This gives an upper bound on $\widetilde{\Delta}_2$.

The upper bound increases in $\Delta_0$, hence we need to bound it from above. By Assumption 4a) in Section 4.4,

$$\int_{\mathcal{X}\setminus B(x^\star,\delta)} \exp\{-[h_y(x) - h_y(x^\star)]/\tau\}dx \leq e^{2M_{f,0}||y - y_{\mathrm{exact}}||/\tau} \int_{\mathcal{X}\setminus B(x^\star,\delta)} \exp\{-[h_{y_{\mathrm{exact}}}(x) - h_{y_{\mathrm{exact}}}(x^\star)]/\tau\}dx.$$

Using Lemma 7 and the lower bound on the minimum eigenvalue of $HH_{01,01}$, we have, on $\Omega_1$,

$$x_0^T HH_{01,01} x_0 \geq \frac{[\nu||P_{A^T}\nabla g(x^\star)|| - M_{f1}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})]^2}{\lambda_{\min\mathrm{pos}}(A^T V_{y_{\mathrm{exact}}}(x^\star)A) - M_{f\,2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})},$$

and, by Lemma 8, on $\Omega_1$,

$$\det(HH_{01,01}) \leq \det(HH^\star_{01,01})[1 + M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})]^{p_0}.$$

Hence, $\Delta_0$ is bounded on $\Omega_1$ from above by

$$\begin{aligned}
\Delta_0^\star(B(0,\delta)) &= \frac{\nu^{p_2}\tau^{(p+p_2)/2}\int_{\mathcal{X}\setminus B(x^\star,\delta)} \exp\left\{-\frac{h_{Y(\omega)}(x) - h_{Y(\omega)}(x^\star)}{\tau}\right\}dx}{\exp\left\{\frac{[\bar{c}_1\nu - \bar{c}_2\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})]^2}{2\tau}\right\}} \qquad (39) \\
&\times \frac{\prod_i (b_i)[\det(HH^\star_{01,01})]^{1/2}[1 + M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})]^{p_0/2}}{[2\pi]^{p_{01}/2}[1 + \Delta_3^\star]}, \qquad (40)
\end{aligned}$$

where $\Delta_3^\star$ is a lower bound on $\Delta_3$ on $\Omega_1$ derived in a similar way.

Using $\lambda_{\min 01} = \lambda_{\min U_{01}}(H_{y_{\mathrm{exact}}}(x^\star))$ and $\Delta_{22} = M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})/\lambda_{\min 01}$, we have that, on $\Omega_1$,

$$\begin{aligned}
\varepsilon_0 &\leq \frac{\widetilde{c}_1||Y - y_{\mathrm{exact}}|| + \widetilde{c}_2\nu}{1 - \delta\sqrt{p}/\lambda_{DH}} \\
&+ \sqrt{-\frac{\tau}{(1 + \Delta_2^\star)^2\lambda_{\min 01}(1 - \Delta_{22})}\log\left(C_{p_{01}}\left(\frac{a^2\tau}{\lambda_{\min 01}(1 - \Delta_{22})(1 + \Delta_2^\star)^2}\right)^{\kappa_{p_{01}}}\right)}, \\
\varepsilon_1 &\leq -\frac{\tau}{\nu b_{\min}}\left[\log\left(\frac{\tau p_2}{\nu b_{\min}\sqrt{1 - a^2}}\right) + \log\left((1 + \Delta_1)(1 + \Delta_2^\star)\right)\right],
\end{aligned}$$

since the function $-x \log x$ increases for $x < 1/e$.

The bound on $\varepsilon_0$ increases in $||y - y_{\text{exact}}||$, and the bound on $\varepsilon_1$ is independent of it. Assume that $\Delta_0^\star/(1 + \Delta_0^\star)$ is less than the upper bound on $(\varepsilon_0 + \varepsilon_1)/(a + \sqrt{1-a^2})$ on $\Omega_1$. Using the lifting Theorem 3, we have that, for small enough $\tau, \nu$,

$$\begin{aligned}\rho_{\mathrm{K}}(\mu_{\text{post}}, \delta_{x^\star}) &\leq \max\left\{2\rho_{\mathrm{K}}(Y, y_{\text{exact}}), \frac{\widetilde{c}_1 \rho_{\mathrm{K}}(Y, y_{\text{exact}}) + \widetilde{c}_2 \nu}{a + \sqrt{1-a^2}}\right.\\ &+ \frac{\sqrt{\tau}[a + \sqrt{1-a^2}]^{-1}}{\sqrt{(1+\Delta_2^\star)^2 \lambda_{\min 01}(1 - \Delta_{22})}} \sqrt{-\log\left(C_{p_{01}}\left(\frac{a^2\tau}{\lambda_{\min 01}(1-\Delta_{22})(1+\ \Delta_2^\star)^2}\right)^{\kappa_{p_{01}}}\right)}\\ &- \frac{\tau}{\nu b_{\min}}\left[\log\left(\frac{\tau p_2}{\nu b_{\min}\sqrt{1-a^2}}\right) + \log\left((1+\Delta_1)(1+\Delta_2^\star)\right)\right] [a + \sqrt{1-a^2}]^{-1}.\end{aligned}$$

Denoting

$$\begin{aligned}\Delta_{\star,K}(\delta, p_2) &= \frac{(a + \sqrt{1-a^2})^{-1}}{(1+\Delta_2^\star)(1-\Delta_{22})^{1/2}}\left[1 + 2\frac{\log(1+\Delta_2^\star) + 0.5\log(1-\Delta_{22})]}{\log\left(\lambda_{\min 01}/(a^2\tau)\right)}\right]^{1/2} - 1,\\ \Delta_{\star\star,K}(\delta, p_2) &= \frac{\log\left((1+\Delta_1)(1+\Delta_2^\star)\right)}{\log\left(\frac{\tau p_2}{\nu b_{\min}\sqrt{1-a^2}}\right)}, \qquad (41)\end{aligned}$$

we have the statement of Theorem 5.

□

*Proof of Theorem 3.* First we note that

$$\begin{aligned}&\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_1\}\\ +\ &\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_2(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_2\}\\ \geq\ &\mathbb{P}\{\, \Phi_1(d_y(Y_1(\omega), Y_2(\omega))) \leqslant \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_1\}\\ +\ &\mathbb{P}\{\, \Phi_2(d_y(Y_1(\omega), Y_2(\omega))) \leqslant \Phi_2(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_2\}\\ =\ &\mathbb{P}\{\, d_y(Y_1(\omega), Y_2(\omega)) \leqslant \rho_{\mathrm{K}}(Y_1, Y_2) \cap \Omega_1\}\\ +\ &\mathbb{P}\{\, d_y(Y_1(\omega), Y_2(\omega)) \leqslant \rho_{\mathrm{K}}(Y_1, Y_2) \cap \Omega_2\}\\ =\ &\mathbb{P}\{\, d_y(Y_1(\omega), Y_2(\omega)) \leqslant \rho_{\mathrm{K}}(Y_1, Y_2)\}\\ \geq\ &1 - \rho_{\mathrm{K}}(Y_1, Y_2).\end{aligned}$$

On the other hand,

$$\begin{aligned}&\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_1\}\\ +\ &\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_2(\rho_{\mathrm{K}}(Y_1, Y_2)) \cap \Omega_2\}\\ \leq\ &\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2))\} + \mathbb{P}\{\, \Omega_2\}.\end{aligned}$$

Putting these together implies

$$\mathbb{P}\{\, d_x(X_1(\omega), X_2(\omega)) > \Phi_1(\rho_{\mathrm{K}}(Y_1, Y_2))\} \leq \rho_{\mathrm{K}}(Y_1, Y_2) + \mathbb{P}\{\, \Omega_2\},$$

hence, using Lemma 5, we have

$$\rho_{\rm K}(X_1, X_2) \leqslant \max\left\{\Phi_1(\rho_{\rm K}(Y_1, Y_2)), \rho_{\rm K}(Y_1, Y_2) + \mathbb{P}(\Omega_2)\right\},$$

and we have the first statement. The second statement follows from the first inequality and

$$\begin{aligned} & \mathbb{P}\left\{ d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_1(\rho_{\rm K}(Y_1, Y_2)) \cap \Omega_1\right\} \\ + & \mathbb{P}\left\{ d_x(X_1(\omega), X_2(\omega)) \leqslant \Phi_2(\rho_{\rm K}(Y_1, Y_2)) \cap \Omega_2\right\} \\ \leq & \mathbb{P}\left\{ d_x(X_1(\omega), X_2(\omega)) \leqslant \max[\Phi_1(\rho_{\rm K}(Y_1, Y_2)), \Phi_2(\rho_{\rm K}(Y_1, Y_2))]\right\}. \end{aligned}$$

□

## A.3 Ky Fan distance inequalities

In this section we quote the result by Hofinger and Pikkarainen (2007).

**Lemma 3.** *(Lemma 7, Hofinger and Pikkarainen (2007)) Let $\xi \sim \mathcal{N}_p(\mu, \Sigma)$. Define*

$$C_p = \begin{cases} 2\pi/(p+1)^2 & \text{if } p \text{ is odd,} \\ 2^p/p^2 & \text{if } p \text{ is even.} \end{cases} \tag{42}$$

$$\kappa_p = \max\{1, p-2\} \tag{43}$$

*Then there exists a positive constant $\theta(p)$ such that for any $\Sigma$: $||\Sigma|| < \theta(p)$,*

$$\rho_{\rm P}(\mathcal{N}(\mu, \Sigma), \delta_\mu) \leq \left(-||\Sigma|| \log\{C_p ||\Sigma||^{\kappa_p}\}\right)^{1/2}. \tag{44}$$

In particular, we will use the following bound on the solution $z = z(p, \lambda)$ of

$$z(p, \alpha) = \inf_{z>0}\{z : \quad 1 - \Gamma\left(\frac{z^2}{2\alpha} \mid \frac{p}{2}, 1\right) < z\},$$

given in the proof of this lemma for sufficiently small $\alpha$:

$$z(p, \alpha) \leq \left[-\alpha \log(C_p \alpha^{\kappa_p})\right]^{1/2}. \tag{45}$$

Here $\Gamma(x|a, b)$ is the cumulative distribution function of the Gamma distribution $\Gamma(a, b)$ with probability density function $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$, $x > 0$.

**Lemma 4.** *Assume that $A \to 0$ and $A \leqslant e^{-1}$. Then the solution of*

$$\exp\{-z/A\} = z$$

*satisfies*

$$z = -A \log(A)(1 + \omega),$$

*where $\omega \leqslant 0$ and $\omega = o(1)$ as $A \to 0$.*

*Proof.* Taking the logarithm of the given expression, we have

$$-z/A = \log z$$

Since $A \to 0$, we must have $z/\log z \to 0$ which implies $z \to 0$. Denote $f = z/A$, i.e. $z = Af$. Hence, the equation above can be rewritten as

$$-f = \log A + \log f$$

implying that $f \to \infty$ as $A \to 0$ at the rate $f = -\log A(1 + o(1))$. Hence, the solution is $z = -A \log A(1 + o(1))$.

To show that $z \leqslant z_* = -A\log(A)$, we note that for $A \leqslant e^{-1}$,

$$\exp\{z_*/A\}z_* = \exp\{-\log(A)\}(-A\log(A)) = -\log(A) \geqslant 1 = \exp\{z/A\}z$$

implying the desired inequality.

□

The following lemma follows obviously from the definition of Ky Fan distance.

**Lemma 5.** *If $\mathbb{P}(d(X,Y) > \varepsilon_1) \leqslant \varepsilon_2$ for some $\varepsilon_1, \varepsilon_2 \in (0,1)$, then $\rho_{\mathrm{K}}(X,Y) \leqslant \max(\varepsilon_1, \varepsilon_2)$.*

## A.4 Auxiliary results

**Lemma 6.** *Assume that $\alpha, \beta > 0$ satisfy $\alpha\beta < 1/4$.*

*If $z < 1/(2\beta)$ and $z \leqslant \alpha + \beta z^2$, then $z \leqslant 2\alpha$.*

*If $\alpha \to 0$ and $\beta$ is bounded, the solution of $z = \alpha + \beta z^2$ such that $z < 1/(2\beta)$ satisfies $z = \alpha(1 + o(1))$.*

The proof is obvious.

Define the following projections

$$\begin{aligned} P_V &= V^\dagger V, \\ P_{A,V} &= (A^TVA)^\dagger A^TVA = A^\dagger P_V A. \end{aligned}$$

**Lemma 7.** *If $[A^TV_y(x)A : B(x)]$ is of full rank,*

$$||H_y^{-1}(x)|| = [\min(\lambda_{\min,\, P_{A,V}}(A^TV_y(x)A + \nu B(x)), \nu\lambda_{min,\, I-P_{A,V}}(B(x)))]^{-1},$$

*where $\lambda_{\min,\, P}(B(x)) = \min_{||v||=1,\, Pv=v} ||B(x)v||$ is the smallest eigenvalue of $B(x)$ on the range of $P$.*

*In particular, if $H_y(x)$ is of full rank and $\nabla^3 h_y(x)$ is uniformly bounded on $B(x^\star, \delta)$ for $y \in \mathcal{Y}_{\rm loc}$,*

$$
\begin{aligned}
||H_y^{-1}(x)|| &\leqslant \frac{1}{\min\left[\lambda_{min,pos}(A^T V_y(x) A) + \nu\lambda_{min,P_{A,V}}(B(x)),\ \nu\lambda_{min,\,I-P_{A,V}}(B(x))\right]}, \\
||H_y(x^\star)^{-1}\nabla h_y(x^\star)|| &\leqslant \frac{1}{\lambda_{\rm min,pos}(A^T V_y(x^\star)A) + \nu\lambda_{{\rm min},P_{A,V}}(B(x^\star))}\left[||P_{A,V}\,\nabla f_y(x^\star)|| + \nu||P_{A,V}\,\nabla g(x^\star)||\right] \\
&+ \frac{1}{\lambda_{min,\,I-P_{A,V}}(B(x^\star))}\left[\nu^{-1}||(I-P_{A,V})\,\nabla f_y(x^\star)|| + ||(I-P_{A,V})\,\nabla g(x^\star)||\right].
\end{aligned}
$$

*Proof.* of Lemma 7.

The norm of $H^{-1}$ is given by

$$
\begin{aligned}
||H^{-1}|| &= \sigma^2[\lambda_{min}(A^T V A + \nu B)]^{-1} = \sigma^2[\min_{||x||=1}||(A^T V A + \nu B)x||]^{-1} \\
&= \sigma^2[\min_{||x||=1}||(A^T V A + \nu B)P_{A^T}x + \nu B)(I - P_{A^T})x||]^{-1} \\
&= \sigma^2[\min(\min_{||x||=1, P_{A^T}x = x}||(A^T V A + \nu B)P_{A^T}x||, \min_{||x||=1,(I-P_{A^T})x=x}\nu||B(I-P_{A^T})x||)]^{-1} \\
&= \sigma^2[\min(\lambda_{min,\,P_{A^T}}(A^T V A + \nu B), \nu\lambda_{min,\,I-P_{A^T}}(B))]^{-1}.
\end{aligned}
$$

Weyl inequality implies that $\lambda_{min,\,P_{A^T}}(A^T V A + \nu B) \geqslant \lambda_{min,\,P_{A^T}}(A^T V A) + \nu\lambda_{min,\,P_{A^T}}(B)$.

Note that since we assumed that $V_{y_{\rm exact}}(x^\star)$ is of full rank, the projection on the range of $A^T$ coincides with the projection on the range of $A^T V_{y_{\rm exact}}(x^\star)A$.

Now we find an upper bound on $||H_{y_{\rm exact}}(x^\star)^{-1}\nabla h_y(x^\star)||$ using the first statement in Lemma 8:

$$
\begin{aligned}
||H_{y_{\rm exact}}(x^\star)^{-1}\nabla h_y(x^\star)|| &= \tau^{-1}||H_{y_{\rm exact}}(x^\star)^{-1}\left(\nabla f_y(x^\star) + \nu\nabla g(x^\star)\right)|| \\
&\leqslant \tau^{-1}||H_{y_{\rm exact}}(x^\star)^{-1}||_{P_{A^T}}||P_{A^T}(\nabla f_y(x^\star) + \nu\nabla g(x^\star))|| \\
&+ \tau^{-1}||H_{y_{\rm exact}}(x^\star)^{-1}||_{I-P_{A^T}}||(I-P_{A^T})[\nabla f_y(x^\star) + \nu\nabla g(x^\star)]|| \\
&\leqslant \frac{1}{\lambda_{min,pos}(A^T V_{y_{\rm exact}}(x^\star)A) + \nu\lambda_{min,P_{A^T}}(B(x^\star))}||P_{A^T}[\nabla f_y(x^\star) + \nu\nabla g(x^\star)]|| \\
&+ \frac{1}{\lambda_{min,I-P_{A^T}}(B)}\tau||(I-P_{A^T})[\nabla f_y(x^\star) + \nu\nabla g(x^\star)]|| \\
&\leqslant \frac{[||P_{A^T}\,\nabla f_y(x^\star)|| + \nu||P_{A^T}\,\nabla g(x^\star)||]}{\lambda_{min,pos}(A^T V_{y_{\rm exact}}(x^\star)A) + \nu\lambda_{min,P_{A^T}}(B(x^\star))} \\
&+ \frac{1}{\lambda_{min,I-P_{A^T}}(B(x^\star))}\left[\nu^{-1}||(I-P_{A^T})\,\nabla f_y(x^\star)|| + ||(I-P_{A^T})\,\nabla g(x^\star)]||\right].
\end{aligned}
$$

□

**Lemma 8.** *1.* $||(C + \delta I)^{-1}x|| \leqslant (\delta + \lambda_k(C))^{-1}||P_C x|| + \delta^{-1}||(I - P_C)x||$

*where $k = rank(C)$ and $\lambda_k(C)$ is the smallest positive eigenvalue of $C$, and $P_C = C^\dagger C$ is the projection matrix.*

2. *Cauchy's interlacing theorem (**?**): let $C = C^T$ be a $n \times n$ matrix, $L$ any $n-k$ dimensional linear subspace, and $C_L = P_L C P_L$. Then, for any $j = 1, \ldots, n-k$,*

$$\lambda_j(C) \geqslant \lambda_j(C_L) \geqslant \lambda_{j+k}(C).$$

3. $\lambda_{\min \mathrm{pos}}(A^T D A) \geqslant \min_{D_i > 0} D_i \lambda_{\min \mathrm{pos}}(A^T A)$ *where $D$ is a diagonal matrix with non-negative entries.*

*Proof.* of Lemma 8

3. $\lambda_j(A^T D A) = \lambda_j(D^{1/2} A A^T D^{1/2})$, and since $j \geqslant \mathrm{rank}(A^T D A) = \mathrm{rank}(D^{1/2} A A^T D^{1/2})$,

$$\lambda_j(D^{1/2} A A^T D^{1/2}) \geqslant \min_{D_i > 0} D_i \lambda_j(P_D A A^T P_D) \geqslant \min_{D_i > 0} D_i \lambda_{j+m}(A A^T)$$

by Cauchy's interlacing theorem, where $m = \mathrm{rank}(P_D)$, $n = \dim(D)$.

If $j = r = \mathrm{rank}(P_{A^T} P_D)$, $\lambda_r(A^T D A)$ is the smallest positive eigenvalue of $A^T D A$, and $j + m = \mathrm{rank}(P_D) + \mathrm{rank}(P_{A^T} P_D) \geqslant \mathrm{rank}(P_{A^T})$. Hence $\lambda_{r+m}(A^T A) \geqslant \lambda_{\mathrm{rank}(P_{A^T})}(A^T A)$, and the latter is the smallest positive eigenvalue of $A^T A$.

□

## A.5 Proof of Bernstein – von Mises theorem in a nonregular case

*Proof.* Denote $\sigma = \sqrt{\tau}$. An upper bound on the total variation distance between the rescaled posterior distribution and its limit is given by

$$\begin{aligned} ||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)} - \mu^\star||_{TV} &\leq ||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)} \mathbf{1}_{B_R} - \mu^\star \mathbf{1}_{B_R}||_{TV} \\ &+ ||\mu^\star - \mu^\star \mathbf{1}_{B_R}||_{TV} + ||\mathbb{P}_{\mathcal{S}(x-x^\star)|Y} \mathbf{1}_{B_R} - \mathbb{P}_{\mathcal{S}(x-x^\star)|Y}||_{TV}, \end{aligned}$$

where the balls $B_R$ are defined below. Here $\mu \mathbf{1}_{B_R}$ is a probability measure $\mu$ truncated to $B_R$ and normalised to be a probability measure. We start with the distance between the truncations of the rescaled posterior distribution and the limit on $B_R$.

Denote $w_k = (U^{-1})(x - x^\star)$, and rescale to $v_0 = w_0/\sigma$, $v_1 = w_1/\gamma$ and $v_2 = w_2/\gamma^2$, with the Jacobian of this change of variables being $J = \sigma^{p_0} \gamma^{p_1 + 2p_2} \det(U)$.

Consider a neighbourhood of $x^\star$, $B_\delta(x^\star) = x^\star + B_\delta$, where $B_\delta = B_2(0, \delta_0) \times B_2(0, \delta_1) \times B_\infty(0, \delta_2)$. For the rescaled parameter, we use the corresponding neighbourhood $B_R = B_2(0, R_0) \times B_2(0, R_1) \times B_\infty(0, R_2)$ where

$$R_0 = \delta_0/\sigma, \quad R_1 = \delta_1/\gamma, \quad R_2 = \delta_2/\gamma^2.$$

We assume that $\delta_k$ are such that $\delta_k \to 0$ and $R_k \to \infty$. In addition, we will need conditions $\delta_1^2 \ll \delta_0$, $\delta_1 \gg \delta_0$ and $\delta_0 \gg \nu = \frac{\sigma^2}{\gamma^2}$. In particular, we can take $\delta_0 = \nu^{1/3}$, $\delta_1 = \sqrt{\gamma}$ and $\delta_2 = \gamma$.

Approximate $h_y(x)$ by a quadratic function using Taylor decomposition in a neighbourhood of $x^\star$:

$$h_y(x) \;=\; h_y(x^\star) + [\nabla h_y(x^\star)]^T(x - x^\star) + \frac{1}{2}(x - x^\star)^T H(x - x^\star) + \Delta_{00}(x).$$

Note that

$$\nabla h_y(x^\star) \;=\; P_{A^T}(\nabla f_y(x^\star) + \nu \nabla g(x^\star)) + \nu(I - P_{A^T})\nabla g(x^\star).$$

Bound $\Delta_{00}$ on $w \in B_\delta$ using Taylor decomposition of $h_y(x)$: $\exists x_c \in \langle x, x^\star\rangle$:

$$\begin{aligned}
|\Delta_{00}(\delta)| &= \frac{1}{6}\left|\sum_{ijk} \nabla_{ijk} h_y(x_c)(x_i - x_i^\star)(x_j - x_j^\star)(x_k - x_k^\star)\right| \\
&= \frac{1}{6}\left|\sum_{i}(x_i - x_i^\star)\nabla_i \left[(x - x^\star)^T \nabla^2 h_y(x_c)(x - x^\star)\right]\right| \\
&\le \frac{1}{6}||x - x^\star||_1 \left[||w_0||_1^2 C_{f3} + \nu C_{g3}[\sum_{k=0}^{2} ||w_k||_1]^2\right] \\
&\le \frac{1}{6}||x - x^\star||_1 \left[||w_0||_1^2 (C_{f3} + 3\nu C_{g3}) + 3\nu C_{g3}[||w_1||_1^2 + ||w_2||_1^2]\right].
\end{aligned}$$

On $B_\delta$,

$$||x - x^\star||_1 = ||w_0||_1 + ||w_1||_1 + ||w_2||_1 \le \sqrt{p_0}\delta_0 + \sqrt{p_1}\delta_1 + p_2\delta_2 \stackrel{def}{=} \delta.$$

Denote

$$D_0 = p_0(C_{f3}/3 + \nu C_{g3}) I_{p_0}, \quad D_1 = p_1 C_{g3} I_{p_1}.$$

Then,

$$\begin{aligned}
[h_y(x) - h_y(x^\star)]/\sigma^2 &\le b^T v_2 + \frac{1}{2}(v_0 - x_0/\sigma)^T H_{00}(v_0 - x_0/\sigma) + \frac{1}{2} v_1^T B_{11} v_1 \\
&+ \sqrt{\nu}\, v_0^T B_{01} v_1 - ||H_{00}^{1/2} x_0||^2/(2\sigma^2) + \delta[v_0^T D_0 v_0 + v_1^T D_1 v_1]/2 \\
&+ [\sigma v_0^T U_0^T + \gamma v_1^T U_1^T]\, B U_2 v_2 + \frac{2C_{g3}\,\delta}{3}\gamma^2 ||v_2||_1^2 \\
&+ \frac{\gamma^2}{2} v_2^T B_{22} v_2,
\end{aligned}$$

where $H_{ij} = U_i^T H U_j$, $B_{ij} = U_i^T B U_j$, $i, j \in \{0, 1, 2\}$, and $H_{jk} = \nu U_j^T B U_k$ for $(j,k) \notin (0,0)$.

Therefore, we have that

$$\begin{aligned}
[h_y(x) - h_y(x^\star)]/\sigma^2 &\le \widetilde{b}_2^T v_2 + ||\widetilde{H}_{00}^{1/2}(v_0 - \widetilde{H}_{00}^{-1} H_{00} x_0/\sigma)||^2/2 + ||\widetilde{B}_{11}^{1/2}(v_1 + \sqrt{\nu}\widetilde{B}_{11}^{-1} B_{10} v_0)||_2^2/2 \\
&- ||\widetilde{H}_{00}^{-1/2} H_{00} x_0||^2/(2\sigma^2),
\end{aligned}$$

where $\widetilde{B}_{11} = B_{11} + \delta_1 D_1$, $\widetilde{H}_{00} = H_{00} - \nu B_{01}\widetilde{B}_{11}^{-1}B_{10} + \delta_0 D_0$, and

$$\widetilde{b} \;\; = \;\; b + \mathbf{1}\left[\delta_0||B_{02}||_{2,\infty} + \delta_1||B_{12}||_{2,\infty} + \frac{\delta_2}{2}||B_{22}||_{1,\infty} + \delta p_2\delta_2\frac{2C_{g3}}{3}\right],$$

since on $B_\delta$,

$$[\sigma v_0^T U_0^T B + \gamma v_1^T U_1^T B]\, U_2 v_2 + \frac{\gamma^2}{2}v_2^T B_2 v_2 \leq [\delta_0||B_{02}||_{2,\infty} + \delta_1||B_{12}||_{2,\infty}]||v_2||_1 + \frac{\delta_2}{2}||B_{22}||_{1,\infty}||v_2||_1.$$

Therefore,

$$\begin{aligned}
&\int_{x^\star + B_\delta} \exp\left\{-[h_y(x) - h_y(x^\star)]/\sigma^2\right\} dx \geq J\exp\{||\widetilde{H}_{00}^{-1/2}H_{00}x_0||^2/(2\sigma^2)\} \\
&\qquad \times \int_{B_R} \exp\left\{-\widetilde{b}^T v_2 - ||\widetilde{H}_{00}^{1/2}(v_0 - \widetilde{H}_{00}^{-1}H_{00}x_0/\sigma)||^2/2 - ||\widetilde{B}_{11}^{1/2}(v_1 + \sqrt{\nu}\widetilde{B}_{11}^{-1}B_{10}v_0)||_2^2/2\right\} dv \\
\geq\;\; & J\exp\left\{||\widetilde{H}_{00}^{-1/2}H_{00}x_0||^2/(2\sigma^2)\right\}\left[\det(\widetilde{H}_{00})\det(\widetilde{B}_{11})\right]^{-1/2} \\
&\qquad \times \prod_{i=1}^{p_2}\widetilde{b}_i^{-1}\left[\prod_{i=1}^{p_2}[1 - \exp\{-\widetilde{b}_i\delta_2/\gamma^2\}]\right](2\pi)^{(p_0+p_1)/2} \\
&\qquad \times\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}_{00})[\delta_0 - ||\widetilde{H}_{00}^{-1}H_{00}x_0||]^2}{2\sigma^2}|\frac{p_0}{2}\right)\times\Gamma\left(\frac{\lambda_{\min}(\widetilde{B}_{11})(\delta_1 - \delta_0||\widetilde{B}_{11}^{-1}B_{10}||)^2}{2\gamma^2}|\frac{p_1}{2}\right) \\
=\;\; & J\exp\{||\widetilde{H}_{00}^{-1/2}b_0||^2/(2\sigma^2)\}M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)C(\widetilde{H}, \widetilde{B}, \widetilde{b}),
\end{aligned}$$

since we have that $\delta_1 \gg \delta_0$.

Denote

$$\begin{aligned}
M_R(Q, A, b, \nu) &= \prod_{i=1}^{p_2}[1 - \exp\{-b_i R_2\}]\,\Gamma\left(\frac{\lambda_{\min}(Q_{00})[R_0 - ||Q_{00}^{-1}H_{00}x_0||/\sigma]^2}{2}|\frac{p_0}{2}\right) \\
&\times\;\; \Gamma\left(\frac{\lambda_{\min}(A_{11})(R_1 - \sqrt{\nu}R_0||A_{11}^{-1}B_{10}||)^2}{2}|\frac{p_1}{2}\right), \\
C(Q, A, b) &= \prod_{i=1}^{p_2} b_i^{-1}\,[\det(Q_{00})\det(A_{11})]^{-1/2}.
\end{aligned}$$

Here $M_R(Q, A, b, \nu)$ is the measure of $B_R$ under the considered class of distributions, in particular, $\mu^\star(B_R) = M_R(\Omega, B, a, 0)$, and $C(Q, A, b)$ is the inverse of the normalising constant.

Similarly, we obtain an upper bound on $e^{-h_y(x)/\sigma^2}$ for any $x \in B_\delta(x^\star)$. Denote $\bar{B}_{11} = B_{11} - \delta_1 D_1$, $\bar{H}_{00} = H_{00} - \nu B_{01}\bar{B}_{11}^{-1}B_{10} - \delta_0 D_0$, and

$$\bar{b} \;\; = \;\; b - \mathbf{1}\left[\delta_0||B_{02}||_{2,\infty} + \delta_1||B_{12}||_{2,\infty} + \frac{\delta_2}{2}||B_{22}||_{1,\infty} + \delta p_2\delta_2\frac{2C_{g3}}{3}\right],$$

for $\delta_k$ small enough so that $\bar{H}_{00}$ and $\bar{B}_{11}$ are positive definite and $\bar{b}_2 > 0$.

Then, for any $x \in B_\delta(x^\star)$,

$$\exp\left\{-[h_y(x) - h_y(x^\star)]/\sigma^2\right\} dx \le J \exp\left\{||\bar{H}_{00}^{-1/2} H_{00} x_0||^2/(2\sigma^2)\right\}$$
$$\times \exp\left\{-\bar{b}^T v_2 - ||\bar{H}_{00}^{1/2}(v_0 - \bar{H}_{00}^{-1} H_{00} x_0/\sigma)||^2/2 - ||\bar{B}_{11}^{1/2}(v_1 + \sqrt{\nu}\bar{B}_{00}^{-1} B_{01} v_0)||^2/2\right\} dv.$$

The posterior density normalised by the posterior measure of $B_R$ can be written as

$$\begin{aligned}
\frac{p(\mathcal{S}(x - x^\star) \mid Y)dx}{p(B_R \mid Y)} &\le \frac{\exp\left\{-||\bar{H}_{00}^{1/2}(v_0 - \bar{H}_{00}^{-1} H_{00} x_0/\sigma)||^2/2 - ||\bar{B}_{11}^{1/2}(v_1 + \sqrt{\nu}\bar{B}_{00}^{-1} B_{01} v_0)||^2/2\right\} dv}{M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)\, C(\widetilde{H}, \widetilde{B}, \widetilde{b})} \\
&\times \exp\left\{-\bar{b}^T v_2\right\}(1 + \Delta_{2,0}(R)),
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{2,0}(R)) &= \frac{\exp\left\{||\bar{H}_{00}^{-1/2} H_{00} x_0||^2/(2\sigma^2)\right\}}{\exp\left\{||\widetilde{H}_{00}^{-1/2} H_{00} x_0||^2/(2\sigma^2)\right\}} - 1 \\
&= \exp\{x_0^T H_{00} \bar{H}_{00}^{-1}[\delta_0 D_0 + \nu\delta_1 B_{01}\tilde{B}_{11}^{-1} D_1 \bar{B}_{11}^{-1} B_{10}]\widetilde{H}_{00}^{-1} H_{00} x_0/\sigma^2\} - 1 \\
&\le \exp\{[\delta_0 p_0(C_{f3}/3 + \nu C_{g3}) + \nu\delta_1 p_1 C_{g3}||B_{01}\tilde{B}_{11}^{-1}\bar{B}_{11}^{-1} B_{10}||]\, ||H_{00}\bar{H}_{00}^{-1}||\, ||\widetilde{H}_{00}^{-1} H_{00}||\, ||x_0||^2/\sigma^2\} - 1.
\end{aligned}$$

The total variation distance between the rescaled posterior distribution and its limit, both truncated to $B_R$, is bounded by

$$||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)}\mathbf{1}_{B_R} - \mu^\star\mathbf{1}_{B_R}||_{TV} \le 2\int_{B_R} \frac{f^\star(v)}{\mu^\star(B_R)}\left[\frac{p(v \mid Y)\mu^\star(B_R)}{f^\star(v)p(B_R \mid Y)} - 1\right]_+ dv$$

$$\begin{aligned}
&\le 2\int_{B_R} \frac{f^\star(v)}{\mu^\star(B_R)}\left[\frac{M(\Omega, B, a, 0)\, C(\Omega, B, a)}{M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)\, C(\widetilde{H}, \widetilde{B}, \widetilde{b})}(1 + \Delta_{2,0}(R))\right. \\
&\times \left.\frac{\exp\left\{-\bar{b}_2^T v_2 - ||\bar{H}_{00}^{1/2}(v_0 - \bar{H}_{00}^{-1} H_{00} x_0/\sigma)||^2/2 - ||\bar{B}_{11}^{1/2}(v_1 + \sqrt{\nu}\bar{B}_{00}^{-1} B_{01} v_0)||^2/2\right\}}{\exp\left\{-b^T v_2 - ||\Omega_{00}^{1/2}(v_0 - a_0)||^2/2 - v_1^T B_{11} v_1/2\right\}} - 1\right]_+ dv \\
&= 2\int_{B_R} \frac{f^\star(v)}{\mu^\star(B_R)}\left[\frac{M_R(\Omega, B, a, 0)\, C(\Omega, B, a)}{M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)\, C(\widetilde{H}, \widetilde{B}, \widetilde{b})}\exp\left\{-[x_0^T H_{00}\bar{H}_{00}^{-1} H_{00} x_0/\sigma^2 - a_0\Omega_{00} a_0]/2\right\}\right. \\
&\times \exp\left\{\tilde{\delta}_2^T v_2 + v_0^T \tilde{D}_0 v_0/2 + \delta_1(v_1 + \sqrt{\nu}\bar{B}_{00}^{-1} B_{01} v_0)^T D_1(v_1 + \sqrt{\nu}\bar{B}_{00}^{-1} B_{01} v_0)/2\right\} \\
&\times \left.\exp\left\{v_0^T[H_{00} x_0/\sigma - \Omega_{00} a_0]\right\} - 1\right]_+ dv,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{D}_0 &= \delta_0 D_0 - U_0[\nabla^2 f_y(x^\star) - \nabla^2 f_{y_{\text{exact}}}(x^\star)]U_0^T - \nu[B_{00} - B_{01}\bar{B}_{11}^{-1} B_{10}], \\
\tilde{\delta}_2 &= -U_2^T[\nabla f_y(x^\star) - \nabla f_{y_{\text{exact}}}(x^\star) + \nu\nabla g(x^\star)] \\
&+ \mathbf{1}\left[\delta_0||B_{02}||_{2,\infty} + \delta_1||B_{12}||_{2,\infty} + \frac{\delta_2}{2}||B_{22}||_{1,\infty} + \delta p_2\delta_2\frac{2C_{g3}}{3}\right].
\end{aligned}$$

Since $\delta_0 = \sqrt{\nu} \gg \nu$ and $\delta_0 \gg \sqrt{\tau} \geq C\sqrt{\mathrm{Var(Y)}}$, $\tilde{D}_0$ is positive definite and $\tilde{\delta}_2 > 0$ with high probability as $\sigma \to 0$.

Now we show that the ratio of the constants is greater than 1. Since function $a^{-1}(1-e^{-ax})$ decreases in $a$ for large $ax$, we have that

$$\prod_{i=1}^{p_2} \left[ \frac{1-\exp\{-b_i R_2\}}{1-\exp\{-[b_i + \tilde{\delta}_{2,i}]R_2\}} \frac{b_i^{-1}}{[b_i + \tilde{\delta}_{2,i}]^{-1}} \right] \geq 1.$$

Also, we have $\widetilde{B}_{11} - B_{11} = \delta_1 D_1 > 0$, hence, by Interlacing Theorem (Lemma 8), all eigenvalues of $\widetilde{B}_{11}$ are greater than the corresponding eigenvalues of $B_{11}$, and therefore $\det(\widetilde{B}_{11}) > \det(B_{11})$, and the difference of arguments of the Gamma functions is given by

$$\begin{aligned} & \frac{\lambda_{\min}(B_{11})R_1^2}{2} - \frac{\lambda_{\min}(\tilde{B}_{11})(R_1 - \sqrt{\nu}R_0||\tilde{B}_{11}^{-1}B_{10}||)^2}{2} \\ = \quad & -\delta\lambda(D_1)R_1^2/2 + \sqrt{\nu}R_0R_1\lambda_{\min}(\tilde{B}_{11})c - \nu R_0^2 \frac{\lambda_{\min}(\tilde{B}_{11})c^2}{2}, \end{aligned}$$

which is positive if $\delta R_1 \ll \sqrt{\nu}R_0$, i.e. if $\delta\delta_1 \ll \delta_0$ (recall that $R_1 \gg \sqrt{\nu}R_0$).

Since for large $\alpha x$, $\alpha^{-1/2}\Gamma(\alpha x \mid \beta)$ decreases in $\alpha$, and, with high probability $\lambda_{\min}(\Omega_{00}) \leq \lambda_{\min}(H_{00} - \nu B_{01}\widetilde{B}_{11}^{-1}B_{10} + \delta_0 D_0)$, we have

$$\frac{[\det(\Omega_{00})]^{-1/2}\Gamma\left(\frac{\lambda_{\min}(\Omega_{00})[R_0 - ||a_0||]^2}{2} | \frac{p_0}{2}\right)}{[\det(\widetilde{H}_{00})]^{-1/2}\Gamma\left(\frac{\lambda_{\min}(\widetilde{H}_{00})[R_0 - ||\widetilde{H}_{00}^{-1}H_{00}x_0||/\sigma]^2}{2} | \frac{p_0}{2}\right)} \geq 1$$

with high probability. Then,

$$\frac{M_R(\Omega, B, a, 0)\, C(\Omega, B, a)}{M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)\, C(\widetilde{H}, \widetilde{B}, \widetilde{b})} \geq 1$$

with high probability. Hence,

$$\begin{aligned} & ||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)}\mathbf{1}_{B_R} - \mu^\star \mathbf{1}_{B_R}||_{TV} \leq -2 + 2\frac{(1+\Delta_{2,0}(R))}{M_R(\widetilde{H}, \widetilde{B}, \widetilde{b}, \nu)\, C(\widetilde{H}, \widetilde{B}, \widetilde{b})} \\ \times \quad & \int_{B_R} \exp\left\{-\bar{b}^T||v_2||_1 - ||\bar{H}_{01,01}^{1/2}(v_{01} + \bar{H}_{01,01}^{-1}b_{01}/\sigma)||^2/2\right\} dv \\ = \quad & 2\frac{(1+\Delta_{2,0}(R))}{M(\widetilde{H}, \widetilde{a}, \delta)\, C(\widetilde{H}, \widetilde{a})} M(\bar{H}, \bar{a}, \delta)\, C(\bar{H}, \bar{a}) - 2 \\ = \quad & 2\Delta_2(R), \end{aligned}$$

where

$$\Delta_2(R) = \frac{M(\bar{H}, \bar{B}, \bar{a}, \delta)\, C(\bar{H}, \bar{B}, \bar{a})}{M(\widetilde{H}, \widetilde{B}, \widetilde{a}, \delta)\, C(\widetilde{H}, \widetilde{B}, \widetilde{a})} (1 + \Delta_{2,0}(R)) - 1.$$

Note that $M(\bar{H}, \bar{a}, \delta)/M(\widetilde{H}, \widetilde{B}, \widetilde{a}, \delta) \leq 1$ with probability $\to 1$ as $\sigma \to 0$, and

$$
\begin{aligned}
\frac{C(\bar{H}, \bar{B}, \bar{a})}{C(\widetilde{H}, \widetilde{B}, \widetilde{a})} &= \prod_{i=1}^{p_2} \left[1 + \frac{2\tilde{\delta}_2}{b_{2,i} - \tilde{\delta}_2}\right] \\
&\times \prod_{i=1}^{p_0} \left[\frac{\lambda_i(H_{00} - \nu B_{01} B_{11}^{-1} B_{10} + D_{00})}{\lambda_i(H_{00} - \nu B_{01} B_{11}^{-1} B_{10} - D_{00})}\right]^{1/2} \prod_{i=1}^{p_1} \left[\frac{\lambda_i(B_{11} + \delta_1 D_1)}{\lambda_i(B_{11} - \delta_1 D_1)}\right]^{1/2} \\
&\leq \left[1 + \frac{2\tilde{\delta}_2}{b_{\min} - \tilde{\delta}_2}\right]^{p_2} \left[1 + \frac{2\delta_1 D_{1,11})}{\lambda_{\min}(B_{11}) - \delta_1 D_{1,11})}\right]^{p_1/2} \\
&\times \left[1 + \frac{2\lambda_{\max}(D_{00}))}{\lambda_{\min}(H_{00} - \nu B_{01} B_{11}^{-1} B_{10}) - \lambda_{\max}(D_{00})}\right]^{p_0/2},
\end{aligned}
$$

where $D_{00} = \delta_0 D_0 + \nu \delta_1 B_{01} B_{11}^{-1} D_1 \widetilde{B}_{11}^{-1} B_{10}$ and $\lambda_i(M)$ is the $i$th largest eigenvalue of matrix $M$. Note that matrix $H_{00} - \nu B_{01} B_{11}^{-1} B_{10}$ is positive definite with high probability.

Therefore, with high probability,

$$
\Delta_2(R) \leq \Delta_2^\star(R) \stackrel{def}{=} \exp\left\{\frac{\delta_0 x_0^T H_{00} \bar{H}_{00}^{-1} D_0 \widetilde{H}_{00}^{-1} H_{00} x_0}{\sigma^2}\right\} \left[1 + \frac{2\tilde{\delta}_2}{b_{\min} - \tilde{\delta}_2}\right]^{p_2} \tag{46}
$$

$$
\times \left[1 + \frac{2\lambda_{\max}(D_{00}))}{\lambda_{\min}(H_{00} - \nu B_{01} B_{11}^{-1} B_{10}) - \lambda_{\max}(D_{00}))}\right]^{p_0/2} \left[1 + \frac{2\delta_1 D_{1,11}}{\lambda_{\min}(B_{11}) - \delta_1 D_{1,11}}\right]^{p_1/2} \tag{47}
$$

The total variation distance between the limit measure and its truncation to $B_R$ is bounded by

$$
\begin{aligned}
||\mu^\star - \mu^\star \mathbf{1}_{B_R}||_{TV} &\leq 2\mu^\star(B_R^c) \\
&\leq 2\left[1 - \prod_{i=1}^{p_2}[1 - \exp\{-b_i R_2\}] + 1 - \Gamma\left(\frac{\lambda_{\min}(B_{11}) R_1^2}{2} | \frac{p_1}{2}\right)\right. \\
&+ \left. 1 - \Gamma\left(\frac{\lambda_{\min}(\Omega_{00})(R_0 - ||a_0||)^2}{2} | \frac{p_0}{2}\right)\right].
\end{aligned}
$$

The total variation distance between the posterior distribution and its truncation to $B_R$ is bounded by

$$
\begin{aligned}
||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)} \mathbf{1}_{B_R} - \mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)}||_{TV} &\leq 2\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)}(B_R^c) \\
&= 2\frac{\int_{\mathcal{X} \setminus B_\delta(x^\star)} \exp\{-(h_y(x) - h_y(x^\star))/\sigma^2\} \, d\, x}{\int_{\mathcal{X}} \exp\{-(h_y(x) - h_y(x^\star))/\sigma^2\} \, d\, x} \\
&= 2\Delta_0(B_\delta) \leq 2\Delta_0^\star(B_\delta),
\end{aligned}
$$

where $\Delta_0^\star(B_\delta)$ is defined by (39).

Combining the three bounds, we have that

$$
||\mathbb{P}_{(\mathcal{S}(x-x^\star)|Y)} - \mu^\star||_{TV} \leq 2\Delta_2^\star(R) + 2\mu^\star(B_R^c) + 2\Delta_0^\star(B_\delta),
$$

which gives the statement of the theorem.

□

## References


Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. B*, **48**, 259–302.

Doob, J. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications*, pp. 23–27. Colloques Internationaux du Centre National de la Recherche Scientifique, Paris.

Dudley, R. M. (2003). *Real analysis and probability*, (second edn). Cambridge University Press, Cambridge.

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

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, (425), 721–41.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *The Annals of Statistics*, **28**, (2), 500–31.

Green, P. J. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, **9**, 84–93.

Hofinger, A. and Pikkarainen, H. K. (2007). Convergence rate for the Bayesian approach to linear inverse problems. *Inverse Problems*, **23**, (6), 2469–84.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, pp. 481–92. Berkeley: University of California Press.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, **135**, (3), 370–84.

Rousseau, J. (2010). Rates of convergence for the posterior distributions of mixtures of Betas and adaptive nonparametric estimation of the density. *Annals of Statistics*, **38**, 146–80.

Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, **2**, 492–4.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, **36**, (3), 1435–63.

Walker, S. (2004). New approaches to Bayesian consistency. *The Annals of Statistics*, **32**, (5), 2028–43.

Weir, I. S. (1997). Fully Bayesian reconstructions from single-photon emission computed tomography data. *Journal of the American Statistical Association*, **92**, (437), 49–60.