Consistency of the posterior distribution in generalized linear inverse problems

Citation for published version:
Bochkina, N 2013, 'Consistency of the posterior distribution in generalized linear inverse problems', *Inverse problems*, vol. 29, no. 9, 095010. https://doi.org/10.1088/0266-5611/29/9/095010

Digital Object Identifier (DOI):
10.1088/0266-5611/29/9/095010

Published In:
Inverse problems

Consistency of the posterior distribution in generalised linear inverse problems

Natalia Bochkina
University of Edinburgh, UK

Abstract

For ill-posed inverse problems, a regularised solution can be interpreted as a mode of the posterior distribution in a Bayesian framework. This framework enriches the set the solutions, as other posterior estimates can be used as a solution to the inverse problem, such as the posterior mean. Bayesian formulation of an ill-posed inverse problem is also natural for scientists as it uses a priori information in a rigourous probabilistic framework, and the posterior distribution can be viewed as a set of possible solutions to the considered ill-posed inverse problem, with a weight characterising how well it is supported by the data and the prior information.

In this paper we study properties of Bayesian solutions to ill-posed inverse problems, namely consistency and the rate of convergence in the Ky Fan metric. We consider the cases where the error distribution is not necessarily Gaussian, but belongs to a particular type of models we refer to as Generalised Linear Inverse Problems. This setting includes some models where the response depends on the unknown parameter nonlinearly. We also consider a particular case of the unknown parameter being on the boundary of the parameter set, and show that the rate of convergence in this case is faster than in the case the unknown parameter is an interior point.

Some key words: Ky Fan metric, consistency, rates of convergence, inverse problems, Bayesian inference, nonregular likelihood, boundary, constrained ill-posed inverse problem.

1 Introduction

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 function.

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 ν a positive constant determining the trade-off between accuracy and smoothness. For further details, see **?**.

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, in most cases 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 (e.g. **?**; **?**; **?**; **?**; **?**). For an interesting philosophical view on inverse problems, falsification, and the role of Bayesian argument, see **?**. Various Bayesian methods to solve inverse problems have been proposed (**?**; **?**; **?**; **?**; **?**).

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, its 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.

In this paper, we present the rates of convergence of the posterior distribution on a finite-dimensional parameter space for an ill-posed inverse problem where the distribution of errors is not necessarily Gaussian. We also consider a particular case where the regularised solution is on the boundary of the parameter space. As we shall see, in the case of an ill-posed inverse problem, 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 $x^\star$ lies on the boundary of the parameter space for a constrained inverse problem (for a Gaussian noise and a Gaussian prior, this problem has been studied by **?**). 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 of the posterior distribution. One of them is to consider the concentration rate of the almost sure convergence of the posterior distribution which is the smallest ε_σ such that

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

as the noise level σ goes to 0, considered by **?**.

Another approach, considered by **?**; **?** 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 (**?**); see Section 3. 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 factor in the rate which is unavoidable. In particular, the Ky Fan rate of convergence ε_σ satisfies, with probability at least $1 - \rho_{\text{K}}(Y, y_{\text{exact}})$,

$$\mathbb{P}(d(x, x^\star) \leq \varepsilon_\sigma \mid Y) \geq 1 - \varepsilon_\sigma \quad \text{on} \quad \{\omega : d(Y(\omega), y_{\text{exact}}) \leq \rho_{\text{K}}(Y, y_{\text{exact}})\},$$

where $\rho_{\text{K}}(Y, y_{\text{exact}})$ is the Ky Fan distance between the data Y and its small noise limit y_{exact}. This allows to have a non-asymptotic framework for the study of convergence of the posterior distribution.

The setting for **?** 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).

In this paper we consider two asymptotic properties of the posterior distribution in the small noise limit: we identify the limit of the posterior distribution and state the rate of convergence in Ky Fan metric. As an intermediate step in deriving the Ky Fan rate of convergence, we have an upper bound on the Prokhorov metric between the posterior distribution and its limit, that metrises weak convergence. This bound is very simple and it allows to make conclusions about the sufficient conditions for weak convergence. We consider a broad class of probability distributions for the data, that we call generalised linear inverse problems, allowing the likelihood to be unidentifiable, and a broad class of prior distributions.

We will also study the asymptotics of the posterior distribution in a particular case where the exact solution lies on the boundary of the parameter space. This is the case of so called nonregular likelihood since the error density has a jump when the value of the parameter coincides with the exact solution. Other examples of the behaviour of the posterior distribution for nonregular likelihoods, including densities with jumps as well as other nonregular models, were considered by **?** and **?** who extended the models studied by **?** in the frequentist setting. We consider a particular case where all coordinates of the exact solution are on the boundary, and show that the rate of convergence of the posterior distribution can be faster than for the regular models.

Section 2 establishes the class of models we study. In Section 3 we discuss the Ky Fan distance and present some examples of calculating the Ky Fan distance for various error distributions. In Section 4 we formulate our theorems on rates of convergence of the posterior distribution. In Section 5 we study an inverse problem where the limit of the posterior distribution (the regularised solution) is situated on the boundary. The proofs are deferred to the Appendix.

2 Model formulation

2.1 Generalised linear inverse problems (GLIP)

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

$$p\,(y|\,x) = F(y, Ax, \tau) = C_{y,\tau} \exp\left\{-\frac{1}{\tau}\tilde{f}_y(Ax)\right\}, \quad y \in \mathcal{Y}, \tag{3}$$

that is, that the distribution depends on $x \in \mathcal{X}$ only via Ax, where τ is a scalar dispersion parameter; in the Gaussian model, τ is the variance σ^2. The observed data y are generated from this distribution, with $x = x_{\text{true}}$, and we aim to recover x_{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 make the following assumptions about the error distribution:

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

Assumption (i) states that τ is not only the dispersion parameter in the model but also a scale parameter for the distribution of Y. Assumption (ii) establishes identifiability of the likelihood with respect to the link parameter $\eta = Ax$.

More generally, Assumption 1 is satisfied by generalised linear models **?**, 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

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

for appropriate functions b, c and d characterising the particular distribution family. The parameter τ 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.

2.2 Bayesian formulation of GLIP

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

$$p\,(x) \propto \exp(-g(x)/\gamma^2), \quad x \in \mathcal{X} \subset \mathbb{R}^p, \tag{4}$$

where γ^2 is a scalar dispersion parameter for the prior that may depend on τ; we relate this to the data dispersion parameter τ by $\gamma^2 = \tau/\nu$, and express most of our results below in terms of τ and ν. Set of possible values of the parameters $\mathcal{X}$ can be any subset of $\mathbb{R}^p$ that contains a nonempty neighbourhood of $x^\star$. Therefore, the posterior distribution satisfies

$$p\,(x|\,y) \propto \exp(-[\tilde{f}_y(Ax) + \nu\, g(x)]/\tau), \quad x \in \mathcal{X}, \tag{5}$$

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 will show that in the limit $\tau \to 0$, the posterior distribution concentrates at point $x^\star$ defined by

$$x^\star = \arg \min_{Ax = Ax_{\text{true}}} g(x).$$

Below we make further assumptions on the likelihood and the prior distribution that we apply to study convergence of the posterior distribution.

3 Types of convergence and corresponding distances

Convergence in distribution (weak convergence) can be metrised by Prokhorov metric (**?**).

Definition 1. *The Prokhorov metric between two measures on a metric space* $(\mathcal{X}, d_\mathcal{X})$ *is defined by*

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

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

This metric can be used to study the weak convergence of the posterior distribution $\mu_{\text{post}}(\omega) = \mathbb{P}_{X|Y(\omega)}$ as a measure on $\mathcal{X}$ to its limit *for a fixed* data set $Y(\omega)$. We consider the Euclidean metric $d(x,z) = ||x - z||$ on $\mathcal{X}$.

To study a weak convergence of the posterior distribution to its limit over all ω, we can use Ky Fan metric that metrised convergence in probability (**?**).

Definition 2. *The Ky Fan metric between two random variables ξ_1 and ξ_2 in a metric space $(\mathcal{W}, d_{\mathcal{W}})$ is defined by*

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

Hence, weak convergence of the posterior distribution μ_{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 $(\mathcal{W}, d_{\mathcal{W}})$ is a space of probability distributions on $\mathcal{X}$ equipped with the Prokhorov metric.

Now we give the Ky Fan distance or its upper bound for some distributions.

For the Gaussian distribution, we quote Lemma 7 from **?**.

Lemma 1. *Let $\xi \sim \mathcal{N}_p(\mu, \Sigma)$. Define*

$$C_p = \begin{cases} 2\pi/(p+1)^2 & \text{if } p \text{ is odd,} \\ 2^p/p^2 & \text{if } p \text{ is even.} \end{cases} \tag{6}$$

$$\kappa_p = \max\{1, p-2\} \tag{7}$$

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

$$\rho_{\mathrm{P}}(\mathcal{N}(\mu, \Sigma), \delta_\mu) \leq \left(-||\Sigma|| \log\{C_p ||\Sigma||^{\kappa_p}\}\right)^{1/2}. \tag{8}$$

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

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

given in the proof of this lemma for sufficiently small α:

$$z(p, \alpha) \leq \left[-\alpha \log(C_p \alpha^{\kappa_p})\right]^{1/2}. \tag{9}$$

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

Now we consider a rescaled Poisson distribution.

Lemma 2. *Consider independent random variables* $Y_t/\tau \sim Pois(\mu_t/\tau)$, $t = 1,\ldots,n$, $\mu_t > 0$. *Denote* $M = 4\sum_t \mu_t$.

Then,

$$\rho_{\mathrm{K}}(Y,\mu) = \sqrt{-\tau M \log(\tau M)}(1 + w_\tau),$$

where $w_\tau = o(1)$ *as* $\tau \to 0$ *and* $w_\tau \leqslant 0$.

Note that the Ky Fan distance has the same asymptotic order as for the Gaussian distribution with $\Sigma = \tau\Sigma_0$, Σ_0 is independent of τ, as $\tau \to 0$.

Now, if we consider the exponential distribution with variance proportional to τ, the order of the Ky Fan distance is different. Let $Y - \mu \sim \mathrm{Exp}(\lambda/\tau)$, then $\mathbb{E}Y = \mu + \tau/\lambda$, $\mathrm{Var(Y)} = \tau^2/\lambda^2$. As $\tau \to 0$, $Y \to \mu$ in probability. The Ky Fan distance is given by

$$\rho_{\mathrm{K}}(Y,\mu) = -\frac{\tau}{\lambda}\log\left(\frac{\tau}{\lambda}\right)(1 + w_\tau),$$

where $w_\tau \leqslant 0$ and $w_\tau = o(1)$ as $\tau \to 0$. This follows from Lemma 5.

Now we give some general statements on an upper bound on the Ky Fan distance for various distributions.

Proposition 1. *Assume that* Y_t *are independent,* $\mathbb{E}Y_t = \mu_t$ *and* $\mathrm{Var(Y_t)} = \mathrm{w_t}\tau$.

1. *Assume that* $\exists C_t \geqslant 1$ *such that* $\kappa_{t,k}$*, the* k*th 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* τ. *Denote* $M = 4\sum_t C_t w_t$.

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

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

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

   Then, for small enough τ,

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

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 1. *Suppose Y_t has a t distribution with ν degrees of freedom, means μ_t and scales $\sqrt{\tau}w_t$, $t = 1, \ldots, n$. Then we can take $K = \nu - 2 - \delta$ for some small $\delta > 0$. Then, using the second statement of Proposition 1,*

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

where ν_K is the Kth moment of the standard t_ν 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 τ.

4 Rates of convergence of posterior distribution in Ky Fan metric

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 ℓ_2 norm.

In the next section we evaluate the level of concentration of the posterior distribution μ_{post} around $x^\star$. We start with the concentration of the posterior distribution $\mu_{\text{post}}(\omega)$ for a fixed ω (i.e. for a particular data set) in the Prokhorov metric, and then, using the lifting theorem (Theorem 2), we use bounds thus obtained to derive a bound on the Ky Fan distance between the posterior distribution and the limit over all ω. In the results below, it is assumed that the dimension p is fixed and is independent of τ.

Throughout the section, we assume that $x^\star$ is an interior point of $\mathcal{X}$.

4.1 Assumptions on the likelihood and the prior

We will make the two main assumptions that the posterior distribution is proper and that the log likelihood and log prior density have bounded third order derivatives.

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, ∇_{ij} and ∇_{ijk} are operators of the second and third derivatives, with $\nabla^2 = (\nabla_{ij})$ being the matrix of second derivatives.

Assumptions on prior distribution.

We assume that the prior distribution is such that the posterior distribution is proper.

1. $\exists \tau_0 > 0$: $\quad \forall \tau \leqslant \tau_0, \quad \int_{\mathcal{X}} e^{-h_y(x)/\tau} dx < \infty$ for all $y \in \mathcal{Y}$.
2. $x^\star = \arg\min_{x \in \mathcal{X}\, Ax = Ax_{\text{true}}} g(x)$ is a unique solution of the minimisation problem.

Smoothness in x.

There exists $\delta > 0$ such that there exist bounded third order derivatives $\exists f_y'''$, $\exists g'''$ on $B(x^\star, \delta)$ for all $y \in \mathcal{Y}_{\text{loc}}$, i.e. $\exists C_{f,3}$, $C_{g,3} < \infty$ such that for all $x \in B(x^\star, \delta)$, for all $y \in \mathcal{Y}_{\text{loc}}$ and all $1 \leqslant i, j, k \leqslant p$,

$$|\nabla_{ijk} f_y(x)| \leqslant C_{f,3}, \quad |\nabla_{ijk} g(x)| \leqslant C_{g,3}, \tag{10}$$

where $\mathcal{Y}_{\text{loc}}$ is the following neighbourhood of y_{exact} in $\mathcal{Y}$:

$$\mathcal{Y}_{\text{loc}} = \{y \in \mathcal{Y} : \quad ||y - y_{\text{exact}}|| \leq \rho_{\text{K}}(Y, y_{\text{exact}})\} \tag{11}$$

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

Convergence in Y.

$\exists M_{f,1}, M_{f,2} < \infty$ such that for all $1 \leqslant j_1, \ldots, j_d \leqslant p$ with $d = 1, 2$, and for all $y \in \mathcal{Y}_{\text{loc}}$,

$$|\nabla_{j_1,\ldots,j_d} f_y(x^\star) - \nabla_{j_1,\ldots,j_d} f_{y_{\text{exact}}}(x^\star)| \leqslant M_{f,d} ||y - y_{\text{exact}}||; \tag{12}$$

These assumptions are satisfied if $\nabla^d f_{\mu_0}(x)$ is differentiable in μ_0 for $d = 1, 2$ and this derivative is bounded on $\mathcal{Y}_{\text{loc}}$, with

$$M_{f,d} \quad = \quad \sup_{y \in \mathcal{Y}_{\text{loc}}} |\nabla_y \nabla_x^d f_y(x^\star)| \quad \text{for} \quad d = 1, 2.$$

Assumptions on δ.

Assume that $\delta > 0$ satisfies the following conditions as $\tau \to 0$:

1.

$$\delta \to 0, \quad \frac{\delta}{\sqrt{\tau}} \to 0, \quad \delta \gg \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \nu, \quad \delta \frac{[\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \nu]^2}{\tau} \to 0, \tag{13}$$

$$\frac{\delta}{\gamma} \to \infty \quad \text{(not necessary if } A^T A \text{ is of full rank).}$$

2. With high probability,

$$\Delta_0(B(0,\delta)) \to 0 \quad \text{as} \quad \tau \to 0, \tag{14}$$

where

$$\Delta_0(B(0,\delta)) \;=\; \frac{\int_{\mathcal{X}\setminus B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx}{\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx}. \tag{15}$$

After the approximation to $e^{-[h_y(x)-h_y(x^\star)]/\tau}$ on $B(x^\star, \delta)$ is derived, condition (14) will be stated in a simplified form in Lemma 3.

Throughout this section we use the error $\Delta_0 = \Delta_0(B(0,\delta))$ defined by (15), and constants κ_p and C_p defined by (6) that feature in the upper bound on the Ky Fan metric between the Gaussian distribution and its mean (Lemma 1 in the Appendix).

4.2 Ky Fan distance

The limiting behaviour of the posterior distribution is characterised by the matrices of second derivatives:

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

and by the gradient:

$$x_0 \;=\; [H_y(x^\star)]^{-1} \nabla h_y(x^\star). \tag{16}$$

Denote a projection matrix on the image of A^T by P_{A^T}, and $P_{A,V} = (A^T V A)^\dagger A^T V A$.

Define $\lambda_{\min\mathrm{pos}}(M)$ 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.

For a fixed ω, we have the following upper bound on the Prokhorov distance between the posterior distribution and its limit.

Theorem 1. *Suppose we have a Bayesian model given in Section 2.1, and let the assumptions stated in Section 4.1 hold. Assume also that $x^\star$ is an interior point of $\mathcal{X}$, and that $H_{Y(\omega)}(x^\star) = A^T V_{Y(\omega)}(x^\star) A + \nu B(x^\star)$ is of full rank.*

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

$$\rho_{\mathrm{P}}(\mu_{\mathrm{post}}(\omega), \delta_{x^\star}) \leqslant \max\left\{ \frac{\Delta_0}{1+\Delta_0}, \quad \frac{M_{f1}||Y(\omega) - y_{\mathrm{exact}}|| + \nu ||P_{A^T}\nabla g(x^\star)||}{\lambda_{\mathrm{min,pos}}(A^T V_{Y(\omega)}(x^\star)A)} \right.$$
$$\left. + \sqrt{-\frac{\tau}{\lambda_{\mathrm{min}}(\omega)}\left(C_p \log\left(\frac{\tau}{\lambda_{\mathrm{min}}(\omega)}\right)^{\kappa_p}\right)}(1 + \Delta_\star(\delta, Y(\omega))) \right\}, \quad (17)$$

where $\lambda_{\mathrm{min}}(\omega) = \lambda_{\mathrm{min}}(H_{Y(\omega)}(x^\star))$, $\Delta_0 = \Delta_0(B(0,\delta))$ defined by (15) and $\Delta_\star$ is defined by (28).

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 $||Y - y_{\mathrm{exact}}|| \overset{\mathbb{P}_{y_{\mathrm{exact}}}}{\to} 0$, (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) $\tau/\lambda_{\mathrm{min}}(H_{Y(\omega)}(x^\star)) \to 0$. If the matrix $A^T V_{Y(\omega)}(x^\star) A$ is of full rank, then, for small τ, $\lambda_{\mathrm{min}}(H_{Y(\omega)}(x^\star))$ is close to the constant $\lambda_{\mathrm{min}}(A^T V_{y_{\mathrm{exact}}}(x^\star) A)$ 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 ν and τ, $\lambda_{\mathrm{min}}(H_{Y(\omega)}(x^\star)) = \nu \lambda_{\mathrm{min,I-P_{A^T}}}(B(x^\star))$; 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 ω, 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 γ may be taken a positive constant.

The theorem also implies that 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 derivative of the logarithm of the posterior density.

This theorem gives 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 ω, we use the following generalisation of the lifting theorem of **?** to the case of different bounds for different outcomes ω.

Theorem 2. *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 Ω is partitioned into two parts, $\Omega = \Omega_1 \cup \Omega_2$, $\Omega_1 \cap \Omega_2 = \emptyset$.*

Assume that there exist positive nondecreasing functions Φ_1 and Φ_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 Ω_1 and Ω_2.

Then, the following inequalities hold:

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

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 ℓ_2 metric. Theorem 1 provides an upper bound Φ_1 on the event Ω_1 where a random matrix $H_{Y(\omega)}(x^\star)$ is of full rank, and the first statement of the theorem is applied to obtain the Ky Fan rate of convergence. Note that we do not need an upper bound Φ_2 to bound the Ky Fan distance on Ω_2, as long as $\mathbb{P}(\Omega_2)$ is vanishingly small as $\tau \to 0$.

Denote

$$\begin{aligned} v_{\min} &= \min_{t:\, V_{y_{\mathrm{exact}}\, tt}(x^\star) > 0} V_{y_{\mathrm{exact}}\, tt}(x^\star), \\ c_1 &= \frac{M_{f1}}{v_{\min}\lambda_{\mathrm{min,pos}}(A^T A)} \qquad c_2 = \frac{||P_{A^T}\nabla g(x^\star)||}{v_{\min}\lambda_{\mathrm{min,pos}}(A^T A)}, \end{aligned} \tag{18}$$

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

$$\bar{c}_k = c_k \left[1 - \frac{M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})}{v_{\min}\lambda_{\mathrm{min\ pos}}(A^T A)}\right]^{-1} [1 - \delta\sqrt{p}/\lambda_{DH}]^{-1}, \quad k = 1, 2, \tag{19}$$

where λ_{DH} is a constant defined by (29).

Theorem 3. *Suppose we have the Bayesian model defined in Section 2.1, and that the assumptions on f_y, g and δ stated in Section 4.1 hold.*

Assume that

1. *$x^\star$ is an interior point of $\mathcal{X}$,*
2. *$H_\nu = A^T V_{y_{\text{exact}}}(x^\star)A + \nu B(x^\star)$ is of full rank,*

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

$$\begin{aligned}\rho_{\mathrm{K}}(\mu_{\text{post}}, \delta_{x^\star}) \;&\leqslant\; \max\Bigg\{ 2\rho_{\mathrm{K}}(Y, y_{\text{exact}}),\ \frac{\Delta_0}{1+\Delta_0},\ \bar{c}_1 \rho_{\mathrm{K}}(Y, y_{\text{exact}}) + \bar{c}_2 \nu \\ &+ \left[-\frac{\tau}{\lambda_{\min}(H_\nu)} \log\left(C_p \left(\frac{\tau}{\lambda_{\min}(H_\nu)} \right)^{\kappa_p} \right) \right]^{1/2} (1 + \Delta_{\star,K}(\delta)) \Bigg\},\end{aligned} \tag{20}$$

where $\bar{c}_1$ and $\bar{c}_2$ defined by (19), $\Delta_0 = \Delta_0(B(0,\delta))$ is given by (21), and $\Delta_{\star,K}(\delta)$ is defined by (34).

Under the assumptions on τ, ν and δ, $\Delta_{\star,K}(\delta) = o(1)$ as $\tau \to 0$.

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

Corollary 2. *Suppose that $\rho_{\mathrm{K}}(Y, y_{\text{exact}}) \leqslant C\sqrt{-\tau \log \tau}$ for some constant C, and that the assumptions of Theorem 3 are satisfied, and that $\frac{\Delta_0}{1+\Delta_0}$ is smaller than the other terms in the maximum.*

If $A^T V_{y_{\text{exact}}}(x^\star)A$ is of full rank (well-posed problem), the smallest upper bound on the Ky Fan distance is

$$\rho_{\mathrm{K}}(\mu_{\text{post}}, \delta_{x^\star}) \leqslant C_1 \left(-\tau \log \tau\right)^{1/2},$$

with $\gamma^2 \geq \tau^{1/2}[-\log \tau]^{-1/4}$.

If $A^T V_{y_{\text{exact}}}(x^\star)A$ is not of full rank (ill-posed problem), the smallest upper bound on the Ky Fan distance is

$$\rho_{\mathrm{K}}(\mu_{\text{post}}, \delta_{x^\star}) \leqslant C_2 \left(-\tau \log \tau\right)^{1/3},$$

with $\gamma^2 = \tau^{2/3}[-\log \tau]^{-1/6}$.

The assumption of corollary $\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) \leqslant C\sqrt{-\tau \log \tau}$ is satisfied for Gaussian random variables Y. In Section 3 we saw that it is also satisfied for such distributions as rescaled Poisson and binomial distributions.

Consider the case of the rescaled Poisson distribution, and a linear inverse problem with the identity link.

Example 2. *Rescaled Poisson random variables satisfy the assumptions of GLIP since it belongs to the exponential family,* $\mathbb{E}Y_i = \mu_i$, $\mathrm{Var(Y_i)} = \tau\mu_{\mathrm{i}}$, *and* $Y \to \mu$ *as* $\tau \to 0$.

Since $x^\star$ *is an interior point of* $\mathcal{X}$*, then, for small enough* σ *and* γ,

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

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

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

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

If $\alpha = 1$ *(ill-posed problem) and* $\tau = \sigma^2$*, the fastest rate is* $\sigma^{2/3}[-\log\sigma]^{1/3}$*, with* $\gamma = \sigma^{2/3}[-\log\sigma]^{-1/6}$.

4.3 Choice of δ

Now, we discuss how to choose δ in such a way that

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

with high probability as $\tau \to 0$, i.e. that the condition (14) $\Delta_0(B(0,\delta)) \to 0 \quad \text{as} \quad \tau \to 0$ is satisfied with high probability.

We introduce the following additional notation. Diagonalise the projection matrices P_{A^T} and $I - P_{A^T}$ simultaneously, so that $P_{A^T} = U^T \text{diag}(I_{p_0}, 0_{p_1})U$, $I - P_{A^T} = U^T \text{diag}(0_{p_0}, I_{p_1})U$ and $U^T U = I_p$, where $p_0 = \text{rank}(A)$ and $p_1 = p - p_0$.

$$\begin{aligned}\Omega_{00} &= U_0^T \nabla^2 f_{y_{\text{exact}}}(x^\star) U_0, \\ B_{11} &= U_1^T \nabla^2 g(x^\star) U_1.\end{aligned}$$

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

Lemma 3. *Assume that Ω_{00} and B_{11} are of full rank. Under the assumptions on f_y, g and assumption (i) on δ stated in Section 4.1,*

$$\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx = \tau^{p_0/2} \gamma^{p_1} \frac{(2\pi)^{p/2} e^{x_0^T H x_0/(2\tau)}}{[\det(\Omega_{00}) \det(B_{11})]^{1/2}} [1 + o_P(1)].$$

In particular, this implies that

$$\Delta_0(B(0,\delta)) = C_H \tau^{-p_0/2} \gamma^{-p_1} \int_{\mathcal{X} \setminus B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau} dx \, [1 + o_P(1)], \quad (21)$$

where $C_H = (2\pi)^{-p/2} [\det(\Omega_{00}) \det(B_{11})]^{1/2} \, e^{-x_0^T H x_0/(2\tau)}$.

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

5 Convergence rate when $x^\star$ is on the boundary

In this section we consider a special case where the assumption that $x^\star$ is an interior point of $\mathcal{X}$ does not hold. This is an example of so called nonregular models that have been considered mostly for a one-dimensional nonregular parameter (ref), and, as far as we are aware, have not been considered in the context of inverse problems. As we shall see, the rate of convergence is different in this case. We shall see that for some probability distributions, it makes it possible to observe exact data under the considered probabilistic model (Section 2).

In this section we assume that the parameter space is $\mathcal{X} = [0, \infty)^p$, and that each coordinate of $x^\star$ is on the boundary of $\mathcal{X} = [0, \infty)^p$, i.e. $x^\star = 0$.

This is an important benchmark case where there is no signal. Such setup arises, for example, in image analysis, where x is the vector of the unknown intensities, and we want to test whether there is any image present. 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.

5.1 Assumptions

We make the same assumptions on the prior distribution as in Section 4.1, however, we only need the smoothness and the convergence assumptions for up to the second derivative only, rather than up to the third. Assumptions on δ – the radius of approximation – are also changed.

Smoothness in x.

There exists $\delta > 0$ such that there exist bounded second order derivatives $\exists f_y''$, $\exists g''$ on $B(x^\star, \delta)$ for all $y \in \mathcal{Y}_{\text{loc}}$, i.e. $\exists C_{\tilde{f},2}$, $C_{g,2} < \infty$ such that for all $x \in B(x^\star, \delta)$, for all $y \in \mathcal{Y}_{\text{loc}}$,

$$\max_{1 \leqslant i,j \leqslant n} |\nabla_{ij} \tilde{f}_y(x)| \leqslant C_{\tilde{f},2}, \quad \max_{1 \leqslant i,j \leqslant p} |\nabla_{ij} g(x)| \leqslant C_{g,2}. \tag{22}$$

Convergence in Y.

$\exists M_{\tilde{f},1} < \infty$ such that for all $1 \leqslant j \leqslant p$ and for all $y \in \mathcal{Y}_{\text{loc}}$,

$$|\nabla_j \tilde{f}_y(Ax^\star) - \nabla_j \tilde{f}_{y_{\text{exact}}}(Ax^\star)| \leqslant M_{\tilde{f},1} ||y - y_{\text{exact}}||. \tag{23}$$

Assumptions on δ.

Assume that $\delta > 0$ satisfies the following conditions as $\tau \to 0$:

1.
$$\delta \to 0, \quad \frac{\delta}{\tau} \to 0,$$
$$\frac{\delta}{\gamma^2} \to \infty \quad \text{(not necessary if } A^T A \text{ is of full rank)}. \tag{24}$$

2. With high probability,
$$\Delta_0(B(0,\delta)) \to 0 \quad \text{as} \quad \tau \to 0, \tag{25}$$
where $\Delta_0(B(0,\delta))$ is defined by (15).

5.2 Rate of convergence in Ky Fan distance

Define

$$b(\omega) = \nabla h_{Y(\omega)}(x^\star).$$

Theorem 4. *Suppose we have the Bayesian model defined in Section 2.1, and let the assumptions on f_y, g and δ stated in Section 5.1 hold.*

Assume that $x^\star = 0$ and that $b_i(\omega) > 0$ for all i, and denote $b_{\min}(\omega) = \min_i b_i(\omega)$.

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

$$\rho_{\mathrm{P}}(\mu_{\mathrm{post}}(\omega), \delta_{x^\star}) \le \max\left\{\frac{\Delta_0}{1+\Delta_0}, \quad -\frac{\tau\sqrt{p}}{\bar{b}_{\min}(\omega)} \log\left(\frac{\tau}{\sqrt{p}\,\bar{b}_{\min}(\omega)}\right)(1+\Delta_4)\right\},$$

where $\Delta_0 = \Delta_0(B(0,\delta))$ is defined by (15) and $\Delta_4(\delta, Y(\omega))$ is defined by (37).

Recall that $b(\omega) = A^T \nabla \tilde{f}_{Y(\omega)}(x^\star) + \nu \nabla g(x^\star)$. Thus, if the image of A^T includes the whole set $\mathcal{X}$ (well-posed case), the leading term of $b(\omega)$ for each coordinate is a constant, then the rate of convergence is determined by $-\tau \log \tau$. However, if $\mathrm{rank}(A) < p$ (ill-posed case), then for some coordinates the leading term of $b(\omega)$ is $\nu\, const \to 0$, then the rate of convergence is determined by $-\gamma^2 \log \gamma$.

To have consistency in the ill-posed case, 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 $\tau \to 0$.

Now we apply Theorems 2 and 4 to obtain an upper bound on the Ky Fan distance. Define

$$b^\star = \nabla h_{y_{\mathrm{exact}}}(x^\star).$$

Theorem 5. *Consider the Bayesian model defined in Section 2.1, and suppose that the assumptions on f_y and g stated in Section 5.1 hold.*

Assume that $\mathcal{X} = [0,\infty)^p$, $x^\star = 0$, $\nabla_i \tilde{f}_{y_{\mathrm{exact}}}(G(y_{\mathrm{exact}})) > 0$ and $b_i^\star > 0$ for all i. Denote $b_{\min}^\star = \min_i b_i^\star$. If $\mathrm{rank}(A^T A) < p$, assume also that $\gamma \to 0$ and $\tau/\gamma^2 \to 0$ as $\tau \to 0$.

Then, for small enough τ, γ and ν,

$$\rho_{\mathrm{K}}(\mu_{\mathrm{post}}, \delta_{x^\star}) \;\le\; \max\left\{2\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}),\ \Delta_0^\star,\ -\frac{\tau\sqrt{p}}{b_{\min}^\star}\log\left(\frac{\tau}{\sqrt{p}\, b_{\min}^\star}\right)(1+\Delta_5^\star)\right\},$$

where $\Delta_0^\star$ *is defined by (21), and*

$$\begin{aligned}
\Delta_5^\star &= -1 + \frac{1+\Delta_4^\star}{1-\Delta_{11}}\left(1 - \frac{\log(1-\Delta_{11})}{\log\left(\frac{\tau}{\sqrt{p}\,b^\star_{\min}}\right)}\right),\\
\Delta_{11} &= \frac{M_{f1}}{b^\star_{\min}}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \delta\frac{p[C_{f2}+\nu C_{g2}/2]}{b^\star_{\min}},\\
\Delta_4^\star &= \frac{\log\left((1+\Delta_1^\star)/(1+\Delta_0^\star)\right)}{\log\left(\sqrt{p}\,b^\star_{\min}[1-\Delta_{11}]/\tau\right)},\\
\Delta_1^\star &= -1 + \left(\frac{1-\Delta_{11}}{1+\Delta_{11}}\right)^p \left[1 - e^{-\max_i b_i^\star(1+\Delta_{11})\delta/(\sqrt{p}\tau)}\right]^p.
\end{aligned}$$

Under the assumptions on τ*,* γ *and* δ *given in Section 5.1,* $\Delta_5^\star(\delta) = o(1)$ *as* $\tau \to 0$*.*

Hence, in the case that the solution is on the boundary, we have a different rate of convergence of the posterior distribution that is faster than the corresponding rate in the case the solution is an interior point. This fits with other studies of the rate of convergence of the posterior distribution for the error densities with jump (**?**; **?**).

Examples.

1. Rescaled Poisson distribution $Y_t/\tau \sim Pois(A_t x/\tau)$, independent. For $x^\star = 0$, we have $\mathbb{P}(Y_t = 0) = 1$ for all t. The Ky Fan distance between the data and its limit is zero, so we observe exact data. In this case, we can recover $P_{A^T}x$ exactly.

   If A^TA is of full rank, the Ky Fan distance 0 and we recover $x^\star$ exactly. If A^TA is not of full rank, the upper bound is of order $-\gamma^2\log(\gamma^2)$ and can be arbitrarily small. This rate is faster than the rate in the case $x^\star$ is an interior point.

2. Exponential error distribution: $Y_t - A_t x \sim \mathrm{Exp}(\lambda_t/\tau)$, independent. For $x^\star = 0$, we have $Y_t \sim \mathrm{Exp}(\lambda_t/\tau)$. In the well-posed case, the Ky Fan distance between the data and its limit is $-\Lambda_E\tau\log\tau$, i.e. is of the same order as the rate of contraction of the posterior distribution to its maximum, where Λ_E is a function of $\lambda_1, \ldots, \lambda_n$. In the ill-posed case, the dominating rate is of order $-\gamma^2\log(\gamma^2)$ which is faster than the corresponding rate when $x^\star \in int(\mathcal{X})$.

A comprehensive study of the rate of convergence of inverse problems under a more general setting (when $x^\star$ is an arbitrary point on the boundary) is beyond the scope of this paper and is current work in progress.

Acknowledgements

This work has arisen from ongoing joint research with Peter Green (University of Bristol, UK) on ill-posed inverse problems with a non-regular likelihood in a more general setting where some coordinates of $x^\star$ are on the boundary.

.1 Proofs of the results in Section 4

Lemma 4. *Denote* $H = A^T V_y(x^\star)A + \nu B(x^\star)$, $\kappa_A = \frac{p}{3}C_{f3}$, $\quad \kappa_B = \frac{p}{3}C_{g3}$.

Assume that H *is invertible, and that* $x^\star$ *is an interior point of* $\mathcal{X}$*.*

Let $x \in B_\delta = \{x \in \mathcal{X} : ||x - x^\star|| \le \delta\}$*, and denote* $v = (x - x^\star)/\sqrt{\tau}$*.*

1. Upper bound. *Then, for small enough* δ *and* ν*, we have the following upper bound:*

$$\begin{aligned}[h_y(x) - h_y(x^\star)]/\tau &\le ||\widetilde{H}^{1/2}(v - \widetilde{H}^{-1}Hx_0/\sqrt{\tau})||^2/2 \\ &+ \frac{[M_{f1}||y - y_{\text{exact}}|| + \nu||P_{A^T}\nabla g(x^\star)||]^2}{\tau[\lambda_{\min \text{pos}}(A^T V_y(x^\star)A) - \delta\kappa_A]},\end{aligned}$$

where $D = \kappa_A P_{A^T} + \nu\kappa_B I$ *and* $\widetilde{H} = H + \delta\sqrt{p}D$*.*

2. Lower bound. *For small enough* δ *and* ν*, we have the following lower bound:*

$$\begin{aligned}[h_y(x) - h_y(x^\star)]/\tau &\ge ||\bar{H}^{1/2}(v - \bar{H}^{-1}Hx_0/\sqrt{\tau})||^2/2 \\ &+ \frac{[M_{f1}||y - y_{\text{exact}}|| - \nu||P_{A^T}\nabla g(x^\star)||]^2}{\tau[\lambda_{\min \text{pos}}(A^T V_y(x^\star)A) + \delta\kappa_A + \nu\lambda_{\min, P_{A^T}}(B)]},\end{aligned}$$

where $\bar{H} = H - \delta\sqrt{p}D$*.*

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

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

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

$$\begin{aligned}\Delta_{00}(\delta) &= \frac{1}{6}\sum_{ijk}\nabla_{ijk}h_y(x_c)(x_i - x_i^\star)(x_j - x_j^\star)(x_k - x_k^\star) \\ &= \frac{1}{6}\sum_i (x_i - x_i^\star)\frac{\partial}{\partial z_i}\left[(x - x^\star)^T\nabla^2 h_y(z)(x - x^\star)\right]_{z=x_c}\end{aligned}$$

Note that

$$(x - x^\star)^T\nabla^2 h_y(z)(x - x^\star) = (x - x^\star)^T P_{A^T}\nabla^2 f_y(z)P_{A^T}(x - x^\star) + \nu(x - x^\star)^T\nabla^2 g(z)(x - x^\star).$$

Differentiating with respect to z and bounding the third derivatives of f_y and g using the Smoothness Assumption, we have that for every i, with high probability,

$$|(x-x^\star)^T P_{A^T}\nabla_i\nabla^2 f_y(z)P_{A^T}(x-x^\star)| \leq C_{f3}\,||P_{A^T}(x-x^\star)||_1^2 \leq pC_{f3}\,||P_{A^T}(x-x^\star)||_2^2,$$

and, similarly,

$$|(x-x^\star)^T\nabla_i\ \nabla^2 g(z)(x-x^\star)| \leq C_{g3}\,||x-x^\star||_1^2 \leq pC_{g3}\,||x-x^\star||_2^2.$$

Applying these inequalities together, we have

$$\begin{aligned}|\Delta_{00}(\delta)| &\leq \frac{1}{6}||x-x^\star||_1 \max_i |(x-x^\star)^T\nabla_i\nabla^2 h_y(z)(x-x^\star)| \\ &\leq \frac{\delta\sqrt{p}}{6}(x-x^\star)^T\left(pC_{f3}P_{A^T}+\nu pC_{g3}I\right)(x-x^\star).\end{aligned}$$

1. *The upper bound.* Making the change of variables $v=(x-x^\star)/\sqrt{\tau}$, we have

$$\begin{aligned}[h_y(x)-h_y(x^\star)]/\tau &\leq -v^T x_0/\sqrt{\tau}+\frac{1}{2}v^T(H+\delta\sqrt{p}D)v \\ &= \frac{1}{2}(v-\widetilde{H}^{-1}Hx_0/\sqrt{\tau})^T\widetilde{H}(v-\widetilde{H}^{-1}Hx_0/\sqrt{\tau})-\frac{1}{2\tau}||\widetilde{H}^{-1/2}Hx_0||^2.\end{aligned}$$

2. *The lower bound.* A similar argument leads to the following lower bound:

$$\begin{aligned}[h_y(x)-h_y(x^\star)]/\tau &\geq -v^T x_0/\sqrt{\tau}+\frac{1}{2}v^T(H-\delta\sqrt{p}D)v \\ &= \frac{1}{2}(v-\bar{H}^{-1}Hx_0/\sqrt{\tau})^T\bar{H}(v-\bar{H}^{-1}Hx_0/\sqrt{\tau})-\frac{1}{2\tau}||\bar{H}^{-1/2}Hx_0||^2.\end{aligned}$$

□

Proposition 2. *Let assumptions on f_y, g in Section 4.1 and assumptions (13) on δ hold. Assume that $H=A^TV_y(x^\star)A+\nu B(x^\star)$ is of full rank, and that $\gamma\to 0$ and $\nu\to 0$ as $\tau\to 0$.*

Then, for any $\varepsilon\in(c_1\rho_{\mathrm{K}}(Y,y_{\mathrm{exact}})+c_2\nu,\delta)$ such that $\varepsilon/\gamma\to\infty$,

$$\frac{\int_{\mathcal{X}\setminus B(x^\star,\varepsilon)}e^{-h_y(x)/\tau}dx}{\int_{\mathcal{X}}e^{-h_y(x)/\tau}dx}\leqslant\left[1-\Gamma\left(\frac{\lambda_{\min}(\bar{H})[\varepsilon-||\bar{H}^{-1}Hx_0||]^2}{2\tau}\mid\frac{p}{2}\right)\right]\frac{1+\Delta_2}{1+\Delta_0}+\frac{\Delta_0}{1+\Delta_0},$$

and, in particular,

$$\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau}dx \geq \frac{\tau^{p/2}}{[\det(\tilde{H})]^{1/2}}[2\pi]^{p/2}\exp\left\{\frac{x_0^T H\tilde{H}^{-1}Hx_0}{2\tau}\right\}[1+\widetilde{\Delta}_3],$$

where Δ_2 *is defined by*

$$\begin{aligned}\Delta_2(\delta,y) &= \left[\Gamma\left(\frac{\lambda_{\max}(\widetilde{H})(\delta-||\tilde{H}^{-1}Hx_0||)^2}{2\tau} \mid \frac{p}{2}\right)\right]^{-1}\left[\frac{1+\delta\sqrt{p}\lambda_{HD}}{1-\delta\sqrt{p}\lambda_{HD}}\right]^{p/2} \\ &\times \exp\left\{\delta\sqrt{p}\,(\kappa_A+\nu\kappa_B)\frac{[M_{f1}||y-y_{\text{exact}}||+\nu||P_A\nabla g(x^\star)]^2}{\tau[\lambda_{\min}^2(A^TV_y(x^\star)A)-\delta^2p\kappa_A^2]}\right\}, \quad (26)\end{aligned}$$

$$\widetilde{\Delta}_3(\delta,y) = -1+\left[\Gamma\left(\frac{\lambda_{\max}(\widetilde{H})(\delta-||\tilde{H}^{-1}Hx_0||)^2}{2\tau} \mid \frac{p}{2}\right)\right]. \quad (27)$$

Here $\lambda_{HD}=\lambda_{\min}(HD^{-1})=\min\left\{\frac{\lambda_{\min,P_{A,V}}(A^TV_y(x^\star)A+\nu B(x^\star))}{\kappa_A+\nu\kappa_B},\frac{\lambda_{\min,I-P_{A,V}}(B(x^\star))}{\kappa_B}\right\}$.

Proof of Proposition 2. Making the change of variables $v=(x-x^\star)/\sqrt{\tau}$ with Jacobian $J=\tau^{p/2}$ and applying Lemmas 4 and 8, we have

$$\begin{aligned}&\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_{y_{\text{exact}}}(x)]/\tau}dx \geq \tau^{p/2}\exp\left\{||\widetilde{H}^{-1/2}Hx_0||^2/(2\tau)\right\} \\ &\times\int_{B(0,\delta/\sqrt{\tau})}\exp\left\{-\left(v-\frac{\widetilde{H}^{-1}Hx_0}{\sqrt{\tau}}\right)^T\widetilde{H}\left(v-\frac{\widetilde{H}^{-1}Hx_0}{\sqrt{\tau}}\right)/2\right\}dv \\ \geq\;& \tau^{p/2}\exp\left\{||\widetilde{H}^{-1/2}Hx_0||^2/(2\tau)\right\}[2\pi]^{p/2}[\det(\widetilde{H})]^{-1/2}\Gamma\left(\frac{\lambda_{\min}(\widetilde{H})[\delta+||\widetilde{H}^{-1}Hx_0||]^2}{2\tau}\mid\frac{p}{2}\right).\end{aligned}$$

In particular, we have the statement of Lemma 3:

$$\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau}dx \;\geq\; \tau^{p/2}\exp\left\{x_0^T H\tilde{H}^{-1}Hx_0/(2\tau)\right\}[2\pi]^{p/2}[\det(\tilde{H})]^{-1/2}[1+\widetilde{\Delta}_3]$$

with $\widetilde{\Delta}_3$ defined by

$$\widetilde{\Delta}_3(\delta,y) = -1+\Gamma\left(\frac{\lambda_{\max}(\widetilde{H})[\delta+||\widetilde{H}^{-1}Hx_0||]^2}{2\tau}\mid\frac{p}{2}\right).$$

The error $\widetilde{\Delta}_3 \to 0$ as $\tau \to 0$, since $\lambda_{\min}(H)\delta^2/\tau \to \infty$ with $\mathbb{P}_{y_{\text{exact}}}$ probability $\to 1$.

Similarly, we can obtain an upper bound on this integral:

$$\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-[h_y(x)-h_y(x^\star)]/\tau}dx \le \tau^{p/2}\exp\left\{||\bar{H}^{1/2}Hx_0||^2/(2\tau)\right\}$$
$$\times \int_{\varepsilon/\sqrt{\tau}\le||v||\le\delta/\sqrt{\tau}} \exp\left\{-\frac{1}{2}||\bar{H}^{1/2}(v-\tau^{-1/2}\bar{H}^{-1}Hx_0)||^2\right\}dv.$$

Assume that δ is small enough so that $\bar{H}$ is positive definite.

Combining these results together, we have that for $\varepsilon > ||\bar{H}^{-1}Hx_0||$,

$$\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} \le \left[1-\Gamma\left(\frac{\lambda_{\min}(\bar{H})[\varepsilon-||\bar{H}^{-1}Hx_0||]^2}{2\tau} \mid \frac{p}{2}\right)\right]$$
$$\times\left[\Gamma\left(\frac{\lambda_{\max}(\widetilde{H})[\delta+||\widetilde{H}^{-1}Hx_0||]^2}{2\tau} \mid \frac{p}{2}\right)\right]^{-1}\left[\frac{\det(\widetilde{H})}{\det(\bar{H})}\right]^{1/2}$$
$$\times\exp\left\{\delta\sqrt{p}\,(\nabla h_y(x^\star))^T\bar{H}^{-1}D\widetilde{H}^{-1}\nabla h_y(x^\star)/\tau\right\},$$

since

$$\bar{H}^{-1}-\widetilde{H}^{-1} = \widetilde{H}^{-1}(\widetilde{H}-\bar{H})\bar{H}^{-1} = 2\delta\sqrt{p}\widetilde{H}^{-1}D\bar{H}^{-1}$$

The ratio of the determinants can be bounded by

$$\frac{\det(\widetilde{H})}{\det(\bar{H})} = \frac{\det(I+\delta\sqrt{p}H^{-1}D)}{\det(I-\delta\sqrt{p}H^{-1}D)} \le \left(\frac{1+\delta\sqrt{p}\lambda_{\max}(DH^{-1})}{1-\delta\sqrt{p}\lambda_{\max}(DH^{-1})}\right)^p,$$

By Lemma 8,

$$(\nabla h_y(x^\star))^T\bar{H}^{-1}D\widetilde{H}^{-1}\nabla h_y(x^\star) \le (\kappa_A+\nu\kappa_B)\frac{[M_{f1}||y-y_{\text{exact}}||+\nu||P_A\nabla g(x^\star)]^2}{\lambda^2_{\min}(A^TV_y(x^\star)A)-\delta^2p\kappa_A^2}.$$

Thus, we have

$$\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} \le \left[1-\Gamma\left(\frac{\lambda_{\min}(\bar{H})[\varepsilon-||\bar{H}^{-1}Hx_0||]^2}{2\tau} \mid \frac{p}{2}\right)\right]$$
$$\times\left[\Gamma\left(\frac{\lambda_{\max}(\widetilde{H})(\delta+||\tilde{H}^{-1}Hx_0||)^2}{2\tau} \mid \frac{p}{2}\right)\right]^{-1}\left[\frac{1+\delta\sqrt{p}\lambda_{\max}(DH^{-1})}{1-\delta\sqrt{p}\lambda_{\max}(DH^{-1})}\right]^{p/2}$$
$$\times\exp\left\{\delta\sqrt{p}\,(\kappa_A+\nu\kappa_B)\frac{[M_{f1}||y-y_{\text{exact}}||+\nu||P_A\nabla g(x^\star)]^2}{\tau[\lambda^2_{\min}(A^TV_y(x^\star)A)-\delta^2p\kappa_A^2]}\right\}.$$

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

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

Substituting the previous upper bound, we have the required statement.

□

Proof of Theorem 1. 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 ξ and $x^\star$.

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

$$\begin{aligned}\frac{\int_{B(x^\star,\delta)\backslash B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau}dx} &\leqslant \widetilde{\Delta}_0 + \left[1 - \Gamma\left(\frac{(\varepsilon - ||\bar{H}^{-1}Hx_0||)^2\lambda_{\min}(\bar{H})}{2\tau} \mid \frac{p}{2}\right)\right](1+\widetilde{\Delta}_2) \\ &\leq \varepsilon,\end{aligned}$$

where $\widetilde{\Delta}_0 = \Delta_0/(1+\Delta_0)$ and $\widetilde{\Delta}_2 = (1+\Delta_2)/(1+\Delta_0) - 1$. The last inequality implies that as $\tau/\lambda_{\min}(H) \to 0$, $\varepsilon \to 0$ and $\varepsilon^2\lambda_{\min}(H)/\tau \to \infty$. Hence, using Lemma 1, we have that

$$\varepsilon \;\leq\; ||\bar{H}^{-1}Hx_0|| + \sqrt{-\frac{\tau}{\lambda_{\min}(\bar{H})}\log\left(C_p\left(\frac{\tau}{\lambda_{\min}(\bar{H})(1+\widetilde{\Delta}_2)^2}\right)^{\kappa_p}\right)}.$$

By Lemma 8,

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

which we can substitute into the upper bound for ε, and

$$||\bar{H}^{-1}H|| = ||(I - \delta\sqrt{p}H^{-1}D)^{-1}|| \leq \left[1 - \delta\sqrt{p}\lambda_{\max}(DH^{-1})\right]^{-1}.$$

Hence an upper bound on the Ky Fan distance is the smallest $\varepsilon \geq \widetilde{\Delta}_0 > 0$ that satisfies the obtained upper bound. Therefore, the Ky Fan distance (and thus, the corresponding Prokhorov distance) is bounded from above by:

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

where $\lambda_{\min}(\omega) = \lambda_{\min}(H_{Y(\omega)}(x^\star))$, $\Delta_0 = \Delta_0(B(0,\delta))$ defined by (15) and $\Delta_\star$ is defined by

$$\Delta_\star(\delta, y) = \frac{1+\Delta_0}{(1+\Delta_2)} \left[1 + 2\frac{\log(1+\Delta_2) - \log(1+\Delta_0)}{\log\left(\lambda_{\min}(\bar{H})/\tau\right)} \right]^{1/2} - 1. \quad (28)$$

□

Proof of Theorem 2. First we note that

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

On the other hand,

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

Putting these together implies

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

hence, using Lemma 6, we have

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

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

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

□

Proof of Theorem 3. Now we prove Theorem 3 in the notation defined in the proof of Theorem 1.

We apply Theorem 2 with $\Omega_1 = \{\omega : ||Y(\omega) - y_{\mathrm{exact}}|| \leq \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})\}$ and $\Omega_2 = \Omega \setminus \Omega_1$ with $\mathbb{P}(\Omega_2) \leq \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})$ by the definition of Ky Fan distance, with the bound Φ_1 given in Theorem 3 which we modify to depend on y only via $||y - y_{\mathrm{exact}}||$. For small enough τ, the assumption of the theorems that $H = A^T V_y(x^\star)A + \nu B(x^\star)$ is of full rank holds on Ω_1, as we shall show below.

The upper bound depends on y via $||y - y_{\mathrm{exact}}||$, $\lambda_{\min}(H_y(x^\star))$, $\lambda_{\min\ \mathrm{pos}}(A^T V_y(x^\star)A)$, Δ_0 and Δ_2.

We start bounding the eigenvalues from below. Denote $H_\nu = H_{y_{\mathrm{exact}}}(x^\star)$. Since $|[H_y(x^\star) - H_{y_{\mathrm{exact}}}(x^\star)]_{ij}| = |[\nabla^2 f_y(x^\star) - \nabla^2 f_{y_{\mathrm{exact}}}(x^\star)]_{ij}| \leq M_{f2}||y - y_{\mathrm{exact}}||$, on Ω_1 we have, by Lemma 9,

$$\lambda_{\max}(\tilde{H}) \leq \lambda_{\max}(H_{y_{\mathrm{exact}}}(x^\star)) + M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \delta\sqrt{p}\lambda_{\min}(DH_\nu^{-1}).$$

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

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

hence $\lambda_{\min\ \mathrm{pos}}(A^T V_Y(x^\star)A) \geq \lambda_{\min\ \mathrm{pos}}(A^T V_{y_{\mathrm{exact}}}(x^\star)A) - M_{f2}||Y - y_{\mathrm{exact}}||$. The lower bound is positive for small enough τ.

We also need to bound $\lambda_{\max}(DH^{-1})$ from above, or equivalently, its inverse from below, on Ω_1,

$$\begin{aligned} \lambda_{\min}(HD^{-1}) &= \min\left\{\frac{\lambda_{\min, P_{A,V}}(A^T V_y(x^\star)A + \nu B(x^\star))}{\kappa_A + \nu\kappa_B}, \frac{\lambda_{\min, I-P_{A,V}}(B(x^\star))}{\kappa_B}\right\} \\ &\geq \min\left\{\frac{\lambda_{\min,\mathrm{pos}}(A^T V_{y_{\mathrm{exact}}}(x^\star)A) - M_{f2}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})}{\kappa_A + \nu\kappa_B}, \frac{\lambda_{\min, I-P^\star_{A,V}}(B(x^\star))}{\kappa_B}\right\} \\ &\stackrel{def}{=} \lambda_{DH}. \end{aligned} \tag{29}$$

Also, on Ω_1:

$$||\tilde{H}^{-1}Hx_0|| \quad \leq \quad \widetilde{c}_1||Y - y_{\text{exact}}|| + \widetilde{c}_2\nu,$$

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

Hence, on Ω_1,

$$\begin{aligned}
\widetilde{\Delta}_2(\delta, Y(\omega)) \quad &\leq \quad -1 + \exp\left\{\frac{\delta p^{3/2}\left(C_{f3} + \nu C_{g3}\right)[\widetilde{c}_1\rho_{\text{K}}(Y, y_{\text{exact}}) + \widetilde{c}_2\nu]^2}{3\tau[1 - \delta^2 p/\lambda_{DH}^2]}\right\} \\
&\times \quad \left[\Gamma\left(\frac{[\lambda_{\max}(H_\nu) + M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}}) + \delta\sqrt{p}\lambda_{\min}(DH_\nu^{-1})]\,[\delta + \widetilde{c}_1\rho_{\text{K}}(Y, y_{\text{exact}}) + \widetilde{c}_2\nu]^2}{\tau} \mid \frac{p}{2}\right)\right]^{-1} \\
&\times \quad \left[\frac{1 + \delta\sqrt{p}/\lambda_{DH}}{1 - \delta\sqrt{p}/\lambda_{DH})}\right]^{p/2}[1 + \Delta_0^\star]^{-1} \\
&\stackrel{def}{=} \quad \Delta_2^\star. \qquad (30)
\end{aligned}$$

By Lemma 9,

$$\det(H) \leq \det(H_\nu)\left[1 + \frac{M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})}{\lambda_{\text{min pos}}(A^T V_{y_{\text{exact}}}(x^\star)A)}\right]^{\text{rank}(A^TA)}.$$

Hence, Δ_0 is bounded on Ω_1 from above by

$$\begin{aligned}
\Delta_0^\star(B(0, \delta)) \quad &= \quad \frac{\tau^{p/2}\int_{\mathcal{X}\setminus B(x^\star,\delta)} \exp\left\{-\tau^{-1}[h_{Y(\omega)}(x) - h_{Y(\omega)}(x^\star)]\right\} dx}{\exp\left\{[2\tau]^{-1}[\bar{c}_1\nu - \bar{c}_2\rho_{\text{K}}(Y, y_{\text{exact}})]^2\right\}} \qquad (31) \\
&\times \quad \frac{[1 + \Delta_{22}]^{\text{rank}(A^TA)/2}}{[1 + \Delta_3^\star]}\frac{[\det(H_\nu)]^{1/2}}{[2\pi]^{p/2}}, \qquad (32)
\end{aligned}$$

where $\Delta_{22} = M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}})/\lambda_{\text{min pos}}(A^T V_{y_{\text{exact}}}(x^\star)A)$. $\Delta_3^\star$ is a lower bound on Δ_3 on Ω_1 derived in a similar way:

$$\Delta_3^\star = -1 + \Gamma\left(\frac{[\lambda_{\max}(H_\nu) + M_{f2}\rho_{\text{K}}(Y, y_{\text{exact}}) + \delta\sqrt{p}\lambda_{\min}(DH_\nu^{-1})]\,[\delta + \widetilde{c}_1\rho_{\text{K}}(Y, y_{\text{exact}}) + \widetilde{c}_2\nu]^2}{\tau} \mid \frac{p}{2}\right). (3$$

Therefore, we have that, on Ω_1,

$$\begin{aligned}
\varepsilon \quad &\leq \quad \frac{\widetilde{c}_1||Y - y_{\text{exact}}|| + \widetilde{c}_2\nu}{(1 - \delta\sqrt{p}/\lambda_{DH})} \\
&+ \quad \sqrt{-\frac{\tau}{\lambda_{\min}(1 - \Delta_{22})}\log\left(C_p\left(\frac{\tau}{\lambda_{\min}(1 - \Delta_{22})(1 + \ \Delta_2^\star)^2}\right)^{\kappa_p}\right)},
\end{aligned}$$

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

The bound on ε increases as a function of $||y - y_{\mathrm{exact}}||$. Using the lifting Theorem 2, we have that, for small enough τ, ν,

$$\begin{aligned}\rho_{\mathrm{K}}(\mu_{\mathrm{post}}, \delta_{x^\star}) &\leq \max\{2\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}),\ \Delta_0^\star,\ [\bar{c}_1\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}}) + \bar{c}_2\nu] \\ &+ \frac{\sqrt{\tau}}{\sqrt{\lambda_{\min}(1-\Delta_{22})}}\sqrt{-\log\left(C_p\left(\frac{\tau}{\lambda_{\min}(1-\Delta_{22})(1+\ \Delta_2^\star)^2}\right)^{\kappa_p}\right)}\right\}.\end{aligned}$$

Denoting

$$\Delta_{\star,K}(\delta) = \frac{1}{(1+\Delta_2^\star)(1-\Delta_{22})^{1/2}}\left[1 + \frac{2\log(1+\Delta_2^\star) + \log(1-\Delta_{22})}{\log(\lambda_{\min}/\tau)}\right]^{1/2} - 1 \qquad (34)$$

we have the statement of Theorem 3.

□

.2 Ky Fan distance inequalities

Lemma 5. *Assume that $A \to 0$ and $A \leqslant e^{-1}$. Then the solution of*

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

satisfies

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

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

Proof. Proof of Lemma 5. 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 6. *If* $\mathbb{P}(d(X,Y) > \varepsilon_1) \leqslant \varepsilon_2$ *for some* $\varepsilon_1, \varepsilon_2 \in (0,1)$*, then* $\rho_{\mathrm{K}}(X,Y) \leqslant \max(\varepsilon_1, \varepsilon_2)$*.*

.3 Proof of results in Section 3.

Proof of Lemma 2. 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 λ 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 5, 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 $\omega \leqslant 0$.

□

Proof of Proposition 1. 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. $z = [n\tau^{m(K)/2} L_K]^{1/(K+1)}$.

□

.4 Proofs of the results in Section 5

Lemma 7. *Denote* $\delta_b = \frac{\delta_p}{2}[C_{\tilde{f},2} A^T A + \nu C_{g2} I]\mathbf{1}$*, and assume that* $b_i(\omega) > 0$ *for all* i*.*

Let $x \in B_\delta = \{x \in \mathcal{X} : ||x - x^\star|| \leq \delta\}$, $x^\star = 0$. Then, for small enough δ and ν, we have the following bounds:

$$\begin{aligned} h_y(x) - h_y(x^\star) &\leq (b(\omega) + \delta_b)^T (x - x^\star), \\ h_y(x) - h_y(x^\star) &\geq (b(\omega) - \delta_b)^T (x - x^\star). \end{aligned}$$

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

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

Similarly to the proof of Lemma 4, bound Δ_{00} for $w = x - x^\star \in B(0, \delta) \cap (\mathcal{X} - x^\star)$ using Taylor decomposition of $h_y(x)$: $\exists x_c \in \langle x, x^\star \rangle$:

$$\begin{aligned} |\Delta_{00}(\delta)| &= \left| \frac{1}{2} \sum_{ij} \nabla_{ij} h_y(x_c)(x_i - x_i^\star)(x_j - x_j^\star) \right| \\ &= \left| \frac{1}{2} \sum_{ij} [\sum_{k\ell} A_{ki} A_{\ell j} \nabla_{k\ell} \tilde{f}_y(x_c) + \nabla_{ij} g(x_c)](x_i - x_i^\star)(x_j - x_j^\star) \right| \\ &\leq \frac{1}{2} \left[C_{\tilde{f},2} ||A(x - x^\star)||_1^2 + \nu C_{g2} ||x - x^\star||_1^2 \right] \\ &\leq \frac{p}{2} (x - x^\star)^T \left(C_{\tilde{f},2} A^T A + \nu C_{g2} I \right) (x - x^\star) \\ &\leq \frac{\delta p}{2} (x - x^\star)^T \left(C_{\tilde{f},2} A^T A + \nu C_{g2} I \right) \mathbf{1} = \delta_b^T (x - x^\star), \end{aligned}$$

since $x_i - x_i^\star \in [0, \delta]$.

Thus, we obtain an upper bound

$$h_y(x) - h_y(x^\star) \leq (b + \delta_b)^T (x - x^\star)$$

and the lower bound:

$$h_y(x) - h_y(x^\star) \geq (b - \delta_b)^T (x - x^\star).$$

□

Proposition 3. *Let assumptions on f_y, g and δ in Section 5.1 hold.*

Assume that $x^\star = 0$, $b_i = \nabla_i h_y(x^\star) > 0$ for all i, and that $\gamma \to 0$ and $\nu \to 0$ as $\tau \to 0$.

Then, for any $\varepsilon \in (0, \delta)$, such that $b_{\min}\varepsilon/\tau \to \infty$,

$$\frac{\int_{\mathcal{X}\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau}dx}{\int_{\mathcal{X}} e^{-h_y(x)/\tau}dx} \leqslant pe^{-\bar{b}_{\min}\varepsilon/(\sqrt{p}\tau)}\frac{1+\Delta_1}{1+\Delta_0} + \frac{\Delta_0}{1+\Delta_0},$$

and, in particular,

$$\int_{B(x^\star,\delta)} e^{-[h_y(x)-h_y(x^\star)]/\tau}dx \geq \tau^p \prod_i b_i^{-1}[1+\widetilde{\Delta}_3],$$

where Δ_1 and $\widetilde{\Delta}_3$ are defined by

$$\Delta_1(\delta, b) \quad = \quad -1 + \prod_i \frac{b_i - \delta_{b,i}}{b_i + \delta_{b,i}} \left[1 - e^{-\min_i \widetilde{b}_i\delta/(\sqrt{p}\tau)}\right]^{-p}, \tag{35}$$

$$\widetilde{\Delta}_3(\delta, y) \quad = \quad -1 + \left[1 + \max_i \delta_{b,\,i}/b_i\right]^{-p}\left[1 - e^{-\min_i \widetilde{b}_i\delta/(\sqrt{p}\tau)}\right]^{p}. \tag{36}$$

Proof of Proposition 3. Making the change of variables $v = (x - x^\star)/\tau$ with Jacobian $J = \tau^p$, we have

$$\int_{B(x^\star,\delta)\cap\mathcal{X}} e^{-[h_y(x)-h_{y_{\text{exact}}}(x)]/\tau}dx \geq \tau^p \int_{B(0,\delta/\tau)\cap(\mathcal{X}-x^\star)} \exp\left\{-(b+\delta_b)^T v\right\}dv$$

$$\geq \tau^p \int_{[0,\delta/(\sqrt{p}\tau)]^p} \exp\left\{-\tilde{b}^T v\right\}dv = \tau^p \prod_i \tilde{b}_i^{-1} \prod_i \left[1 - \exp\left\{-\tilde{b}_i\delta/(\sqrt{p}\tau)\right\}\right]$$

$$\geq \tau^p \prod_i b_i^{-1}[1+\widetilde{\Delta}_3]$$

with $\widetilde{\Delta}_3$ defined by (36). The error $\widetilde{\Delta}_3 \to 0$ as $\tau \to 0$, since $\delta \to 0$ and $b_{\min}\delta/\tau \to \infty$, with $\mathbb{P}_{y_{\text{exact}}}$ probability $\to 1$.

Similarly, we obtain an upper bound on the following integral:

$$\begin{aligned}
\int_{(\mathcal{X}\cap B(x^\star,\delta))\setminus B(x^\star,\varepsilon)} e^{-[h_y(x)-h_y(x^\star)]/\tau}dx &\leq \tau^p \int_{\varepsilon/\tau\leq||v||\leq\delta/\tau,\, v_i\geq 0} \exp\left\{-\bar{b}^T v\right\}dv \\
&\leq \tau^p \int_{||v||\geq\varepsilon/\tau} \exp\left\{-\bar{b}^T v\right\}dv \\
&\leq \sum_i \tau^p \int_{v_i\geq\varepsilon/(\sqrt{p}\tau),\, v_j\geq 0\,\forall j} \exp\left\{-\bar{b}^T v\right\}dv \\
&= \tau^p \prod_i \bar{b}_i^{-1} \sum_i \exp\{-\bar{b}_i\varepsilon/(\sqrt{p}\tau)\} \\
&\leq p\tau^p \prod_i \bar{b}_i^{-1} \exp\{-\min_i \bar{b}_i\varepsilon/(\sqrt{p}\tau)\},
\end{aligned}$$

where $\bar{b} = b - \delta_b$. Assume that δ is small enough so that $\bar{b}_i > 0$ for all i.

Therefore,

$$\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau} dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau} dx} \le p e^{-\bar{b}_{\min}\, \varepsilon/(\sqrt{p}\tau)}(1 + \Delta_1(\delta, b)),$$

where

$$\Delta_1(\delta, b) = -1 + \prod_i \frac{b_i - \delta_{b,i}}{b_i + \delta_{b,i}} \frac{\left[1 - e^{-\max_i \widetilde{b}_i \delta/(\sqrt{p}\tau)}\right]^p}{p \exp\{-\min_i \bar{b}_i \varepsilon/(\sqrt{p}\tau)\}}.$$

Hence, Δ_1 is small if $b_{\min}\delta/\tau \to \infty$ as $\tau \to 0$.

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

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

Thus, we have the required statement.

□

Proof of Theorem 4. We proceed similarly as in the proof of Theorem 1.

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 ξ and $x^\star$.

Using Proposition 3, we have an upper bound ε on the Ky Fan distance satisfies

$$\frac{\int_{B(x^\star,\delta)\setminus B(x^\star,\varepsilon)} e^{-h_y(x)/\tau} dx}{\int_{B(x^\star,\delta)} e^{-h_y(x)/\tau} dx} \leqslant \left[\widetilde{\Delta}_0 + p \exp\left\{-\frac{\varepsilon \bar{b}_{\min}}{\sqrt{p}\tau}\right\} \frac{1 + \Delta_1}{1 + \Delta_0}\right] \le \varepsilon,$$

where $\widetilde{\Delta}_0 = \Delta_0/(1 + \Delta_0)$.

An upper bound on the Ky Fan distance is the smallest $\varepsilon > 0$ such that

$$\begin{aligned}
\widetilde{\Delta}_0 &\le \varepsilon, \\
p \exp\left\{-\frac{\varepsilon \bar{b}_{\min}}{\sqrt{p}\tau}\right\} \frac{1 + \Delta_1}{1 + \Delta_0} &\le \varepsilon.
\end{aligned}$$

The last inequality implies that as $\tau/\bar{b}_{\min} \to 0$, $\varepsilon \to 0$. Hence, using Lemma 5, we have that

$$\varepsilon \leq -\frac{\tau\sqrt{p}}{\bar{b}_{\min}}\left[\log\left(\frac{\tau}{\sqrt{p}\,\bar{b}_{\min}}\right) - \log\left(\frac{1+\Delta_1}{1+\Delta_0}\right)\right].$$

Therefore, the Ky Fan distance is bounded from above by the maximum of the two expressions:

$$\begin{aligned}\rho_{\mathrm{P}}(\mu_{\mathrm{post}}(\omega), \delta_{x^\star}) &\leqslant \max\left\{\frac{\Delta_0}{1+\Delta_0}, \quad -\frac{\tau\sqrt{p}}{\bar{b}_{\min}}\left[\log\left(\frac{\tau}{\sqrt{p}\,\bar{b}_{\min}}\right) - \log\left(\frac{1+\Delta_1}{1+\Delta_0}\right)\right]\right\} \\ &= \max\left\{\frac{\Delta_0}{1+\Delta_0}, \quad -\frac{\tau\sqrt{p}}{\bar{b}_{\min}(\omega)}\log\left(\frac{\tau}{\sqrt{p}\,\bar{b}_{\min}(\omega)}\right)(1+\Delta_4(\delta, Y(\omega)))\right\},\end{aligned}$$

where $\Delta_0 = \Delta_0(B(0,\delta))$ is defined by (15) and Δ_4 is defined by

$$\Delta_4(\delta, y) = \frac{\log\left((1+\Delta_1)/(1+\Delta_0)\right)}{\log\left(\sqrt{p}\,\bar{b}_{\min}(\omega)/\tau\right)}. \tag{37}$$

□

Proof of Theorem 5. Now we prove Theorem 5 in the notation defined in the proof of Theorem 4.

We apply Theorem 2 with $\Omega_1 = \{\omega : ||Y(\omega) - y_{\mathrm{exact}}|| \leq \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})\}$ and $\Omega_2 = \Omega \setminus \Omega_1$ with $\mathbb{P}(\Omega_2) \leq \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})$ by the definition of Ky Fan distance, with the bounds given in Theorem 5 which we modify to depend on y only via $||y - y_{\mathrm{exact}}||$. For small enough τ, given that $b_i^\star > 0$, the assumption of the theorems that $b_i > 0$ holds on Ω_1 for small enough τ, as we shall show below.

The upper bound depends on y via $||y - y_{\mathrm{exact}}||$, $b(\omega)$, Δ_0 and Δ_1.

We have that, on Ω_1,

$$\begin{aligned} b_i &= \sum_j A_{ji}\nabla_j \tilde{f}_y(Ax^\star) + \nu\nabla_i g(x^\star) \\ &\geq \sum_j A_{ji}\nabla_j[\tilde{f}_{y_{\mathrm{exact}}}(Ax^\star) - M_{\tilde{f},1}\rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})] + \nu\nabla_i g(x^\star) \\ &= b_i^\star - \rho_{\mathrm{K}}(Y, y_{\mathrm{exact}})M_{\tilde{f},1}\sum_j A_{ji}, \end{aligned}$$

and also

$$b_i - \delta_{b,i} \geq b_i^\star - [\rho_{\rm K}(Y, y_{\rm exact}) M_{\tilde{f},1} + \delta p C_{\tilde{f},2} ||A||_{1,1}/2] \sum_j A_{j,i} - \nu \delta p C_{g2}/2, \quad (38)$$

Note that if $\sum_j A_{j,i} = 0$, then $b_i - \delta_{b,i} = \nu[\nabla_i g(x^\star) - \delta p C_{g2}/2]$, i.e. the leading term in the lower bound is of order ν. If $\sum_j A_{j,i} \neq 0$, then the leading term in the lower bound is a positive constant $\sum_j A_{ji} \nabla_j \tilde{f}_y(Ax^\star)$. Denote $i^\star = \arg\min_i b_i^\star$ and assume that τ and δ are small enough so that the minimum of the lower bound in (38) is also achieved at $i^\star$. Introduce Δ_{11} such that

$$\Delta_{11} = \left[\rho_{\rm K}(Y, y_{\rm exact}) M_{\tilde{f},1} + 0.5 \delta p C_{\tilde{f},2} ||A||_{1,1}\right] \frac{\sum_j A_{ji^\star}}{b^\star_{\rm min}} + \delta \frac{\nu C_{g,2}}{2 b^\star_{\rm min}}.$$

If $\sum_j A_{j,i^\star} = 0$.

$$\Delta_{11} = \delta \frac{\nu C_{g,2}}{2 b^\star_{\rm min}} = \delta \frac{C_{g,2}}{2 \nabla_{i^\star} g(x^\star)}.$$

Then, an upper bound on the Ky Fan distance is given by

$$\begin{aligned} \varepsilon &\leq -\frac{\tau \sqrt{p}}{\bar{b}_{\rm min}} \log\left(\frac{\tau}{\sqrt{p}\, \bar{b}_{\rm min}}\right)(1 + \Delta_4) \\ &\leq -\frac{\tau\sqrt{p}}{b^\star_{\rm min}[1 - \Delta_{11}]} \log\left(\frac{\tau}{\sqrt{p}\, b^\star_{\rm min}[1-\Delta_{11}]}\right)\left[1 + \Delta_4^\star\right], \end{aligned}$$

since the function $-x \log x$ increases for $x < 1/e$. This bound on ε on Ω_1 is independent of y. The error term $\Delta_4^\star$ is given by

$$\Delta_4^\star = \frac{\log\left((1 + \Delta_1^\star)/(1 + \Delta_0^\star)\right)}{\log\left(\sqrt{p}\, b^\star_{\rm min}[1 - \Delta_{11}]/\tau\right)}$$

Using the lifting Theorem 2, we have that, for small enough τ, ν,

$$\rho_{\rm K}(\mu_{\rm post}, \delta_{x^\star}) \leq \max\left\{2\rho_{\rm K}(Y, y_{\rm exact}),\ \Delta_0^\star,\ -\frac{\tau\sqrt{p}}{b^\star_{\rm min}} \log\left(\frac{\tau}{\sqrt{p} b^\star_{\rm min}}\right)(1 + \Delta_5^\star)\right\},$$

where

$$\Delta_5^\star = -1 + \frac{1 + \Delta_4^\star}{1 - \Delta_{11}} \left(1 - \frac{\log(1 - \Delta_{11})}{\log\left(\frac{\tau}{\sqrt{p}\, b^\star_{\rm min}}\right)}\right).$$

Thus, we have the statement of Theorem 5.

□

.5 Auxiliary results

Define the following projections

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

Lemma 8. *If $[A^T V_y(x) A : B(x)]$ is of full rank,*

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

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.

Proof of Lemma 8. The norm of H^{-1} is given by

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

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 $V_{y_{\mathrm{exact}}}(x^\star)$ is of full rank, the projection on the range of A^T coincides with the projection on the range of $A^T V_{y_{\mathrm{exact}}}(x^\star) A$.

Now we find an upper bound on $||H_{y_{\mathrm{exact}}}(x^\star)^{-1} \nabla h_y(x^\star)||$ using the first statement in Lemma 9:

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

□

Lemma 9. *1.* $||(C+\delta I)^{-1}x|| \leqslant (\delta+\lambda_k(C))^{-1}||P_Cx||+\delta^{-1}||(I-P_C)x||$ *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* $C_L = P_LCP_L$*. Then, for any* $j = 1,\dots,n-k$,

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

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

Proof of Lemma 9. 3. $\lambda_j(A^TDA) = \lambda_j(D^{1/2}AA^TD^{1/2})$, and since $j \geqslant \text{rank}(A^TDA) = \text{rank}(D^{1/2}AA^TD^{1/2})$,

$$\lambda_j(D^{1/2}AA^TD^{1/2}) \geqslant \min_{D_i>0} D_i\lambda_j(P_DAA^TP_D) \geqslant \min_{D_i>0} D_i\lambda_{j+m}(AA^T)$$

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

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

□