Supplementary Material

Rates of contraction of posterior distributions based on p-exponential priors

SERGIOS AGAPIOU[1,*], MASOUMEH DASHTI[2] and TAPIO HELIN[3]

[1] *Department of Mathematics and Statistics, University of Cyprus.*
E-mail: agapiou.sergios@ucy.ac.cy

[2] *Department of Mathematics, University of Sussex.*
E-mail: m.dashti@sussex.ac.uk

[3] *School of Engineering Science, LUT University.*
E-mail: tapio.helin@lut.fi

Note that in order to ease readability, in this supplementary material we use a different type of numbering for sections, results and displayed equations, compared to the main body of the article. In particular, we use letters for sections, and the letter of the section together with number, for results and displayed equations.

A. Proofs of Section 2

Proof of Proposition 2.10. By Proposition 2.9 we have $\overline{\mathcal{Z}}^{\|\cdot\|_X} \subset \overline{\mathcal{Q}}^{\|\cdot\|_X} \subset X$. For any arbitrary $x \in X$ and given $\epsilon > 0$, there exists N such that $x^N = \sum_{\ell=1}^{N} x_\ell \psi_\ell$ satisfies $\|x^N - x\|_X < \epsilon$, where (ψ_ℓ) is the Schauder basis in X and (x_ℓ) the corresponding coefficients of x. Since clearly $x^N \in \mathcal{Q} \cap \mathcal{Z}$, we conclude that $X = \overline{\mathcal{Z}}^{\|\cdot\|_X} = \overline{\mathcal{Q}}^{\|\cdot\|_X}$.

Since μ is a measure on X, we have $\mathrm{supp}(\mu) \subset X$. On the other hand, the topological support of any Radon measure in X is non-empty and by definition closed in X. By Proposition 2.4 we get that $0 \in \mathrm{supp}(\mu)$, thus $\mathcal{Q} \subset \mathrm{supp}(\mu)$. Taking closures in X, we get $X = \overline{\mathcal{Q}}^{\|\cdot\|_X} \subset \mathrm{supp}(\mu)$ and thus the claimed result. □

*Corresponding author

Proof of Proposition 2.11. By Proposition 2.7, letting $V = \frac{|\cdot|^p}{p}$, we have

$$\begin{aligned}\mu(\epsilon B_X + h) &= \int_{\epsilon B_X} \lim_{N\to\infty} e^{\left(\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)\right)} \mu(du) \\ &= e^{-\sum_{\ell=1}^\infty V\left(\frac{h_\ell}{\gamma_\ell}\right)} \int_{\epsilon B_X} \lim_{N\to\infty} e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)} \mu(du) \\ &= e^{-\sum_{\ell=1}^\infty V\left(\frac{h_\ell}{\gamma_\ell}\right)} \int_{\epsilon B_X} \lim_{N\to\infty} \frac{1}{2} \Bigg(e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell - h_\ell}{\gamma_\ell}\right)\right)} \\ &\qquad + e^{\sum_{\ell=1}^N \left(V\left(\frac{u_\ell}{\gamma_\ell}\right) + V\left(\frac{h_\ell}{\gamma_\ell}\right) - V\left(\frac{u_\ell + h_\ell}{\gamma_\ell}\right)\right)} \Bigg) \mu(du),\end{aligned}$$

where in the last equality we used symmetry. In the following, we show that the integrand in the last line above is bounded below by 1. Notice that our proof applies for any $V : \mathbb{R} \to \mathbb{R}_{\geq 0}$ convex and symmetric with $V(0) = 0$, differentiable in $\mathbb{R} \setminus \{0\}$ with concave derivative on the positive axis. The functions $\frac{|\cdot|^p}{p}$, $1 \leq p \leq 2$, clearly satisfy this assumption.

Since $\mathrm{e}^a + \mathrm{e}^{-a} \geq 2$ for any $a \in \mathbb{R}$, we observe that

$$\begin{aligned}&\mathrm{e}^{V(x)+V(y)-V(x-y)} + \mathrm{e}^{V(x)+V(y)-V(x+y)} \\ &= \mathrm{e}^{V(x)+V(y)-\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)} \left(\mathrm{e}^{-\frac{1}{2}V(x-y)+\frac{1}{2}V(x+y)} + \mathrm{e}^{\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)}\right) \\ &\geq 2\,\mathrm{e}^{V(x)+V(y)-\frac{1}{2}V(x-y)-\frac{1}{2}V(x+y)}.\end{aligned}$$

In consequence, we need to show that

$$G(x,y) := V(x) + V(y) - \frac{1}{2}V(x-y) - \frac{1}{2}V(x+y) \geq 0.$$

Notice that G has a number of symmetries. Namely, it satisfies

$$G(x,y) = G(-x,y) = G(x,-y) = G(y,x). \tag{A.1}$$

Therefore, without loss of generality, we can assume that $x, y \geq 0$. We note that if $x = 0$ or $y = 0$ then clearly $G(x,y) = V(0) = 0$. Consequently, due to (A.1) it will be sufficient to show that $\frac{\partial G}{\partial x}(x,y) \geq 0$ for any $x > 0$ and $y > 0$.

Let us briefly consider the derivative $R(x) = V'(x)$ for $x > 0$ and define $R(0) = \lim_{x\to 0^+} V'(x)$. By assumption on V, $V'(x)$ is concave hence continuous for all $x > 0$, implying that the limit exists although it may be $-\infty$. Combining with the convexity of V and since V has a minimum at the origin, we get that the limit is non-negative, $R(0) \geq 0$. The function R defined on $[0,\infty)$ is concave with $R(0) \geq 0$, hence it is subadditive.

We first observe that

$$G(x,x) = 2V(x) - \frac{1}{2}V(2x)$$

and due to the subadditivity of R, we must have $G(x,x) \geq 0$. For $x \neq y$, we have

$$\frac{\partial G}{\partial x}(x,y) = V'(x) - \frac{1}{2}V'(x-y) - \frac{1}{2}V'(x+y).$$

For $x > y$, by concavity of V' on the positive axis, we have

$$V'(x) = V'\Big(\frac{1}{2}(x-y) + \frac{1}{2}(x+y)\Big) \geq \frac{1}{2}V'(x-y) + \frac{1}{2}V'(x+y),$$

implying that $\frac{\partial G}{\partial x}(x, y) \geq 0$ for $x \geq y$. If $x < y$, since by symmetry of V it holds $V'(x-y) = -V'(y-x)$, we can write

$$\frac{\partial G}{\partial x}(x, y) = V'(x) + \frac{1}{2}V'(y-x) - \frac{1}{2}V'(x+y)$$

where the arguments of V' in the right-hand side are positive and we can use the concavity of V' on the positive axis. As above, using the auxiliary function R and since concave functions which are non-negative at zero are subadditive, we have

$$V'(x+y) = V'(2x+y-x) \leq 2V'(x) + V'(y-x).$$

Thus $\frac{dG}{dx}(x; y) \geq 0$ for any $y > x > 0$ as well, and the result follows. □

Proof of Theorem 2.13. Let $h \in \mathcal{Z}$ such that $\|h - w\|_X \leq \epsilon$. Then by the triangle inequality, for any $x \in X$ we have $\|x - w\|_X \leq \epsilon + \|x - h\|_X$, hence if $\|x - h\|_X \leq \epsilon$ then $\|x - w\|_X \leq 2\epsilon$. We thus have,

$$\mu(w + 2\epsilon B_X) \geq \mu(h + \epsilon B_X) \geq e^{-\frac{1}{p}\|h\|_{\mathcal{Z}}^p}\mu(\epsilon B_X),$$

where for the last inequality we used Proposition 2.11. To finish the proof we take the negative logarithm and optimize over $h \in \mathcal{Z}$. □

Proof of Proposition 2.15. Without loss of generality we work in $\mathbb{R}^\infty$. Recall $\gamma = (\gamma_\ell)$ and $\xi = (\xi_\ell)$ from the definition of the p-exponential measure μ, Definition 2.1. The inequality follows from [15, Theorem 2.4], see also [11, Theorem 4.19]. These theorems state that for the infinite (unscaled) independent product of standard p-exponential one-dimensional measures, μ_∞ in $\mathbb{R}^\infty$, there exists a universal constant $K > 0$ depending only on p, such that for all $\tilde{r} > 0$

$$\mu_\infty(A + \sqrt{\tilde{r}}B_2 + \tilde{r}^{\frac{1}{p}}B_p) \geq 1 - \frac{1}{\mu_\infty(A)}\exp\left(-\frac{\tilde{r}}{K}\right),$$

where B_p, B_2 are the closed unit balls in ℓ_p, ℓ_2 respectively. Letting $r = \tilde{r}^{\frac{1}{p}}$, we get

$$\mu_\infty(A + r^{\frac{p}{2}}B_2 + rB_p) \geq 1 - \frac{1}{\mu_\infty(A)}\exp\left(-\frac{r^p}{K}\right).$$

Defining $\Gamma : \mathbb{R}^\infty \to \mathbb{R}^\infty$, such that $x \in \mathbb{R}^\infty \mapsto (\gamma_\ell x_\ell)$, we get that for any μ-measurable set $A \subset \mathbb{R}^\infty$

$$\mu(A + r^{\frac{p}{2}}B_{\mathcal{Q}} + rB_{\mathcal{Z}}) = \mathbb{P}(\Gamma\xi \in A + r^{\frac{p}{2}}B_{\mathcal{Q}} + rB_{\mathcal{Z}}) = \mathbb{P}(\xi \in \Gamma^{-1}(A + r^{\frac{p}{2}}B_{\mathcal{Q}} + rB_{\mathcal{Z}}))$$
$$= \mu_\infty(\Gamma^{-1}A + r^{\frac{p}{2}}B_2 + rB_p) \geq 1 - \frac{1}{\mu_\infty(\Gamma^{-1}A)}\exp\left(-\frac{r^p}{K}\right) = 1 - \frac{1}{\mu(A)}\exp\left(-\frac{r^p}{K}\right).$$

□

B. Proof of general contraction theorem in Section 3

Proof of Theorem 3.1. Assume $\varphi_{w_0}(\epsilon_n) \le n\epsilon_n^2$. It follows by Theorem 2.13 that

$$\mathbb{P}(\|W - w_0\|_X < 2\epsilon_n) = \exp\left(\log \mu(w_0 + 2\epsilon_n B_X)\right) \ge \exp(-\varphi_{w_0}(\epsilon_n)) \ge e^{-n\epsilon_n^2},$$

and, consequently, the claim (9) follows.

We now consider the existence of sets X_n such that (7) and (8) hold. We set

$$X_n = \epsilon_n B_X + M_n^{\frac{p}{2}} B_{\mathcal{Q}} + M_n B_{\mathcal{Z}}, \tag{B.2}$$

where $M_n > 0$ will be chosen below. By Proposition 2.15, we have

$$\mathbb{P}(W \notin X_n) \le \frac{1}{\mu(\epsilon_n B_X)} \exp\left(-\frac{M_n^p}{K}\right) = \exp\left(\varphi_0(\epsilon_n) - \frac{M_n^p}{K}\right). \tag{B.3}$$

Next, for any $C > 1$, we denote $M_n = (K(C+1)n\epsilon_n^2)^{\frac{1}{p}}$ which is bounded away from zero for all n by assumption. Since

$$\varphi_0(\epsilon_n) \le \varphi_{w_0}(\epsilon_n) \le n\epsilon_n^2 \tag{B.4}$$

we obtain the claim (8) by combining (B.3) with (B.4).

For the final claim (7), we cannot use directly Proposition 2.11 to bound the complexity of X_n, since Proposition 2.11 refers to shifts in $\mathcal{Z}$ while X_n involves a ball in $\mathcal{Q}$. We can however, find a large enough ball $\overline{M}_n\mathcal{Z}$ which is such that a $2\epsilon_n$-cushion in X around it contains X_n. We can then use Proposition 2.11 to bound the complexity of $2\epsilon_n B_X + \overline{M}_n B_{\mathcal{Z}}$, which in turn implies a bound on the complexity of X_n.

Define

$$\overline{M}_n = 2\left(M_n \vee (1 + \frac{1}{n}) f(M_n^{\frac{p}{2}})^{\frac{1}{p}} g(\epsilon_n)^{\frac{1}{p} - \frac{1}{2}}\right).$$

Then using (6) we can show that

$$X_n \subset 2\epsilon_n B_X + \overline{M}_n B_{\mathcal{Z}}. \tag{B.5}$$

Indeed, for every $x \in {M_n}^{\frac{p}{2}} B_{\mathcal{Q}}$, we have by (6) that

$$\inf_{z \in \mathcal{Z}: \|z - x\|_X \le \epsilon} \|z\|_{\mathcal{Z}}^p \le f(M_n^{\frac{p}{2}}) g(\epsilon)^{1 - \frac{p}{2}}$$

and, in consequence, there exists $y \in \left(1 + \frac{1}{n}\right) f(M_n^{\frac{p}{2}})^{\frac{1}{p}} g(\epsilon_n)^{\frac{1}{p} - \frac{1}{2}} B_{\mathcal{Z}}$, with $\|x - y\|_X \le \epsilon_n$. The term $1 + 1/n$ does not play any significant role, and can be replaced by any constant over 1. Thus any $x \in {M_n}^{\frac{p}{2}} B_{\mathcal{Q}} + M_n B_{\mathcal{Z}}$ is within ϵ_n $\|\cdot\|_X$-distance from some point in $\overline{M}_n B_{\mathcal{Z}}$ and (B.5) follows.

Let $h_1, \ldots, h_N \in \overline{M}_n B_{\mathcal{Z}}$ be $2\epsilon_n$-apart in $\|\cdot\|_X$. Clearly, the balls $h_j + \epsilon_n B_X$ are disjoint and hence by Proposition 2.11 we obtain

$$1 \geq \sum_{j=1}^{N} \mathbb{P}(W \in h_j + \epsilon_n B_X) \geq \sum_{j=1}^{N} e^{-\frac{\|h_j\|_{\mathcal{Z}}^p}{p}} \mathbb{P}(W \in \epsilon_n B_X) \geq N e^{-\frac{\overline{M}_n^p}{p} - \varphi_0(\epsilon_n)}. \quad \text{(B.6)}$$

If the set of points $h_1, \ldots, h_N$ is maximal in $\overline{M}_n B_{\mathcal{Z}}$ (that is, it achieves the maximum number of points $2\epsilon_n$-apart in $\|\cdot\|_X$ that can fit in $\overline{M}_n B_{\mathcal{Z}}$), then the balls $h_j + 2\epsilon_n B_X$ cover $\overline{M}_n B_{\mathcal{Z}}$ and combining with (B.5) we get that

$$X_n \subset \bigcup_{j=1}^{N} (h_j + 4\epsilon_n B_X). \quad \text{(B.7)}$$

Combining (B.6) together with (B.7) we obtain

$$N(4\epsilon_n, X_n, \|\cdot\|_X) \leq N \leq \exp\left(\frac{\overline{M}_n^p}{p} + \varphi_0(\epsilon_n)\right).$$

Using the definitions of $M_n, \overline{M}_n$ we get

$$\begin{aligned}
&\log N(4\epsilon_n, X_n, \|\cdot\|_X) \\
&\leq \frac{2^p}{p}\left(K(C+1)n\epsilon_n^2 \vee (1+\frac{1}{n})^p f(\sqrt{K(C+1)}n^{\frac{1}{2}}\epsilon_n) g(\epsilon_n)^{1-\frac{p}{2}}\right) + n\epsilon_n^2.
\end{aligned}$$

Finally, using that $n\epsilon_n^2 \gtrsim 1$ and f is non-decreasing with $f(a) \to \infty$ at most polynomially as $a \to \infty$, we get (7) . This completes the proof. □

C. Proofs of Section 5

Proof of Lemma 5.2. The proof is very similar to the proof of [4, Theorem 5], taking into account Proposition 2.5. We include it for the reader's convenience.

We first show that for $s < \alpha$, it holds $\mu(B_q^s) = 1$. Indeed, we have

$$\mathbb{E}_\mu \|u\|_{B_q^s}^q = \mathbb{E} \sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d}+\frac{1}{2})-1} |\gamma_\ell|^q |\xi_\ell|^q = \mathbb{E}|\xi_1|^q \sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha)}{d}-1}.$$

If $s < \alpha$, then the expectation is finite, hence $\mu(B_q^s) = 1$.

We next show that if $\mu(B_q^s) = 1$ then $s < \alpha$. If $\mu(B_q^s) = 1$ then $\|u\|_{B_q^s} < \infty$ almost surely with respect to μ, hence

$$\sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha)}{d}-1} |\xi_\ell|^q < \infty, \quad \text{almost surely.} \quad \text{(C.8)}$$

By contradiction to the Law of Large Numbers, we then get that

$$s < \alpha + \frac{d}{q}. \tag{C.9}$$

Define $\zeta_\ell = \ell^{\frac{q(s-\alpha)}{d}-1}|\xi_\ell|^q$, which are independent non-negative random variables. By [10, Proposition 4.14], (C.8) also implies that

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell \wedge 1] < \infty. \tag{C.10}$$

We have

$$\mathbb{E}\zeta_\ell = \mathbb{E}[\zeta_\ell \mathbb{1}_{\zeta_\ell \le 1}] + \mathbb{E}[\zeta_\ell \mathbb{1}_{\zeta_\ell > 1}] \le \mathbb{E}[\zeta_\ell \wedge 1] + I_\ell,$$

where

$$I_\ell = \mathbb{E}\left[\ell^{\frac{q(s-\alpha)}{d}-1}|\xi_\ell|^q \mathbb{1}_{\{|\xi_\ell| > \ell^{\frac{\alpha-s}{d}+\frac{1}{q}}\}}\right] = c_p \ell^{\frac{q(s-\alpha)}{d}-1} \int_{\ell^{\frac{\alpha-s}{d}+\frac{1}{q}}}^{\infty} x^q e^{-\frac{x^p}{p}} dx,$$

where c_p depends only on p. Since $p, q \ge 1$, it holds that $x^q e^{-\frac{x^p}{p}} \le C_1 e^{-C_2 x}$, for constants $C_1, C_2 > 0$ sufficiently large and small, respectively. This results in the bound

$$I_\ell \le c_p \frac{C_1}{C_2} \ell^{\frac{q(s-\alpha)}{d}-1} \exp(-C_2 \ell^{\frac{\alpha-s}{d}+\frac{1}{q}}) := \iota_\ell,$$

where by (C.9) ι_ℓ are summable. Combining, we get that

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell] \le \sum_{\ell=1}^{\infty} \mathbb{E}[\zeta_\ell \wedge 1] + \sum_{\ell=1}^{\infty} \iota_\ell < \infty.$$

We thus have

$$\sum_{\ell=1} \mathbb{E}[\zeta_\ell] = \mathbb{E}[|\xi_1|^q] \sum_{\ell=1}^{\infty} \ell^{\frac{q(s-\alpha}{d}-1} < \infty,$$

and therefore $s < \alpha$.

Finally, Proposition 2.5 implies that $\mu(B_q^s) = 0$, for $s \ge \alpha$. □

Proof of Proposition 5.4. We examine each case separately.

i) For $q \ge 2$, by Lemmas 5.12 and 5.13, we have that as $\epsilon \to 0$ the concentration function satisfies

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{\beta p - \alpha p - d}{\beta}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta < \alpha + \frac{d}{p}, \\ (-\log \epsilon)^{\frac{q-p}{q}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta = \alpha + \frac{d}{p}, \\ 1 + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta > \alpha + \frac{d}{p}. \end{cases} \tag{C.11}$$

Determining which term dominates in each case, we find

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{\beta p - \alpha p - d}{\beta}}, & \text{for } \alpha \geq \beta, \\ \epsilon^{-\frac{d}{\alpha}}, & \text{for } \alpha < \beta. \end{cases} \tag{C.12}$$

Computing the minimal solution ϵ_n such that $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, we arrive at $\epsilon_n \asymp r_n^{\alpha,\beta,p,q}$.

ii) For $q < 2$ and $p \leq q$, by Lemmas 5.12 and 5.13, we have that as $\epsilon \to 0$ the concentration function satisfies

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{2\beta pq - 2\alpha pq - 2qd}{2\beta q + qd - 2d}} + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta < \alpha + \frac{d}{p}, \\ (-\log \epsilon)^{\frac{q-p}{q}} + \epsilon^{-\frac{d}{\alpha}} & \text{for } \beta = \alpha + \frac{d}{p}, \\ 1 + \epsilon^{-\frac{d}{\alpha}}, & \text{for } \beta > \alpha + \frac{d}{p}. \end{cases} \tag{C.13}$$

In this case, determining which of the two terms dominates for $\beta < \alpha + \frac{d}{p}$ is a bit more complicated and leads to a quadratic equation for the value of α balancing the two terms. This quadratic equation has two solutions, a negative one which is rejected since $\alpha > 0$ and $\alpha = \frac{\beta p - d + a}{2p}$, where a as in the statement of the lemma. Notice that the expression under the square root in a is positive by the assumption on β. We thus have the following bound:

$$\varphi_{w_0}(\epsilon) \lesssim \begin{cases} \epsilon^{\frac{2\beta pq - 2\alpha pq - 2qd}{2\beta q + qd - 2d}}, & \text{for } \alpha \geq \frac{\beta p - d + a}{2p}, \\ \epsilon^{-\frac{d}{\alpha}}, & \text{for } \alpha < \frac{\beta p - d + a}{2p}. \end{cases} \tag{C.14}$$

Computing the minimal solution ϵ_n such that $\varphi_{w_0}(\epsilon_n) \leq n\epsilon_n^2$, we again arrive at $\epsilon_n \asymp r_n^{\alpha,\beta,p,q}$.

iii) For $q < 2$, thus $p > q$, the proof is similar to (ii) above, using the expressions corresponding to this case from Lemma 5.13.

Finally, notice that the constants in all the used upper bounds on the concentration function from Lemmas 5.12 and 5.13, depend on w_0 only through its B_q^β norm, hence so do the constants in the rates derived above. □

Proof of Proposition 5.8. In all of the studied cases except when $\alpha = \beta - \frac{d}{p}$ for $q > p$, Lemmas 5.12 and 5.13 result in a bound of the form

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim \lambda^{-p}\epsilon^{-s} + (\epsilon/\lambda)^{-\frac{d}{\alpha}},$$

for some $s = s(\alpha, \beta, p, q) \geq 0$. We optimize the choice of λ by balancing the two terms, obtaining the choice

$$\lambda \asymp \epsilon^{\frac{d - \alpha s}{d + \alpha p}}. \tag{C.15}$$

We then equate the resulting bound on $\bar{\varphi}_{w_0}(\epsilon)$ with $n\epsilon^2$, thus getting

$$\epsilon_n \asymp n^{-(2 + \frac{dp + ds}{d + \alpha p})^{-1}}. \tag{C.16}$$

The last rate coincides with the rate $m_n = n^{-\frac{\beta}{d+2\beta}}$, if and only if

$$s = s_0(\alpha, \beta, p) := \frac{d + \alpha p - \beta p}{\beta}.$$

The corresponding optimal λ is

$$\lambda_n \asymp n^{\frac{\alpha-\beta}{d+2\beta}}. \tag{C.17}$$

For $s > s_0$, the rate ϵ_n is polynomially slower than m_n, and this holds for the optimal and hence any choice of λ_n. The case $s < s_0$ does not arise, as seen below (it cannot arise as it would lead to a faster rate than the minimax rate m_n). The proof thus proceeds by comparing the values of s obtained in Lemma 5.13 to s_0.

Recall that $q < 2$. If $q \le p$ and $\alpha \ge \beta - \frac{d}{q}$, we have

$$s = \frac{2\alpha pq + 2pd - 2\beta pq}{2\beta q + dq - 2d}. \tag{C.18}$$

It holds

$$s - s_0 = \frac{dq(\beta - \alpha - d/q)(p-2) + (d^2 + 2\alpha d)(p-q)}{\beta(2\beta q + dq - 2d)} \ge 0,$$

since both numerator and denominator are positive by our assumptions. The difference $s - s_0$ vanishes if and only if $\alpha = \beta - \frac{d}{q}$ and $q = p$, in which case $\epsilon_n \asymp m_n$ for $\lambda_n \asymp n^{-\frac{d}{p(d+2\beta)}}$. In all other cases, for any λ_n, ϵ_n is polynomially slower than m_n.

If $\alpha < \beta - \frac{d}{p}$ for any relationship between q and p, we have $s = 0 > s_0$, hence for any λ_n, ϵ_n is polynomially slower than m_n.

If $q > p$ and $\alpha > \beta - \frac{d}{p}$, then

$$s = \frac{2\alpha pq + 2qd - 2\beta pq}{2\beta q + qd - 2d}.$$

It holds

$$s - s_0 = \frac{(\beta - \alpha - \frac{d}{p})(q-2)dp}{\beta(2\beta q + qd - 2d)} > 0,$$

hence for any λ_n, ϵ_n is polynomially slower than m_n.

We next turn to the case $q > p$ for $\alpha = \beta - \frac{d}{p}$ (recall $q < 2$). By Lemmas 5.12 and 5.13 we have the bound

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim \lambda^{-p} (\log 1/\epsilon)^{\frac{q-p}{q}} + (\epsilon/\lambda)^{-\frac{d}{\alpha}}.$$

We again optimize the choice of λ by balancing the two terms, to find

$$\lambda \asymp \epsilon^{\frac{d}{\beta p}} (\log 1/\epsilon)^{\frac{q-p}{q}\frac{\beta p - d}{\beta p^2}}. \tag{C.19}$$

The resulting bound on $\bar{\varphi}_{w_0}(\epsilon)$ is

$$\bar{\varphi}_{w_0}(\epsilon) \lesssim (\log 1/\epsilon)^{\frac{d(q-p)}{\beta q p}} \epsilon^{-\frac{d}{\beta}}.$$

We equate this bound with $n\epsilon^2$, obtaining the equation

$$\epsilon_n^{\frac{d+2\beta}{\beta}} \log^{-\frac{(q-p)d}{\beta qp}}(1/\epsilon_n) = n^{-1}.$$

We then solve for ϵ using [12, Lemma 3], included below as Lemma G.2 for the reader's convenience. This gives the claimed value of $\bar{r}_n$ and we can in turn compute the value of λ_n, by plugging $\epsilon_n \asymp \bar{r}_n$ in (C.19). It holds that $\omega > 0$ (see the expression for λ_n in the statement), since $\alpha = \beta - \frac{d}{p} > 0$.

Finally, we return to the case $q < p$ and determine the best achievable rate. If $\alpha \leq \beta - \frac{d}{q}$, then $s = 0$ and by (C.16), the rate is

$$\epsilon_n \asymp n^{-(2+\frac{dp}{d+\alpha p})^{-1}},$$

which is (uniquely) optimized when we choose α as large as possible, $\alpha = \beta - \frac{d}{q}$. Plugging this choice of α into the last expression and (C.15), we obtain the claimed values for $\bar{r}_n$ and λ_n, respectively. It remains to verify that for $\alpha > \beta - \frac{d}{q}$ the resulting rate ϵ_n in (C.16) is not better than the aforementioned $\bar{r}_n$. Indeed, comparing ϵ_n and $\bar{r}_n$, we conclude that this is the case if and only if the following inequality holds

$$s \geq p^2 \frac{\alpha q + d - \beta q}{dq - dp + p\beta q},$$

with equality if and only if the two rates coincide. Using the expression for s from (C.18), we get that the last inequality is equivalent to

$$\frac{2}{2\beta q + dq - 2d} \geq \frac{p}{\beta pq + dq - dp},$$

where notice that the two denominators are positive since $\beta \geq \frac{d}{q} - \frac{d}{2}$ and $p \leq 2$. It is then straightforward to see that the last inequality is equivalent to $p \leq 2$ and so indeed the rate for $\alpha > \beta - \frac{d}{q}$ is strictly worse than $\bar{r}_n$ when $p < 2$, while for $p = 2$ the rates are identical for any $\alpha \geq \beta - \frac{d}{q}$.

As always, notice that the constants in all the used upper bounds on the concentration function from Lemmas 5.12 and 5.13, depend on w_0 only through its B^β_q norm, hence so do the constants in the rates derived above. □

Proof of Lemma 5.13. Let $w_0 := (w_{0,\ell})_{\ell \in \mathbb{N}} \in B^\beta_q$, so that under the assumption on β it holds that $w_0 \in \ell_2$, see Lemma G.1 below. Since $w_0 \in B^\beta_q$ we find that

$$|w_{0,\ell}| \leq \|w_0\|_{B^\beta_q} \ell^{-\frac{\beta}{d} - \frac{1}{2} + \frac{1}{q}}. \tag{C.20}$$

Consider now approximations $h_{1:L} = (w_1, ..., w_L, 0, ...) \in \mathbb{R}^\infty$ of w_0, where $L \in \mathbb{N}$. Obviously, we have $h_{1:L} \in \mathcal{Z}_\alpha$ for any L. We first study how large L needs to be, in order to have

$$\|h_{1:L} - w_0\|_{\ell_2} \leq \epsilon. \tag{C.21}$$

In case $q > 2$, we have

$$\begin{aligned}\|h_{1:L} - w_0\|_{\ell_2}^2 &= \sum_{\ell > L} \ell^{-\frac{2\beta}{d}-1+\frac{2}{q}} \ell^{\frac{2\beta}{d}+1-\frac{2}{q}} w_{0,\ell}^2 \\ &\leq \left(\sum_{\ell > L} \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} |w_{0,\ell}|^q \right)^{\frac{2}{q}} \left(\sum_{\ell > L} \ell^{-\frac{2\beta q}{d(q-2)}-1} \right)^{\frac{q-2}{q}} \leq \|w_0\|_{B_q^\beta}^2 L^{-\frac{2\beta}{d}},\end{aligned}$$

where we have used the Hölder inequality $(\frac{q}{2}, \frac{q}{q-2})$ and comparison of the sum to an integral. If $q = 2$, we obtain

$$\|h_{1:L} - w_0\|_{\ell_2}^2 = \sum_{\ell > L} \ell^{\frac{2\beta}{d}} w_{0,\ell}^2 \ell^{-\frac{2\beta}{d}} \leq L^{-\frac{2\beta}{d}} \|w_0\|_{B_q^\beta}^2 .$$

If $q < 2$, applying (C.20) yields

$$\begin{aligned}\|h_{1:L} - w_0\|_{\ell_2}^2 &= \sum_{\ell > L} |w_{0,\ell}|^q |w_{0,\ell}|^{2-q} \leq \|w_0\|_{B_q^\beta}^{2-q} \sum_{\ell > L} |w_{0,\ell}|^q \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} \ell^{-\frac{2\beta}{d}-1+\frac{2}{q}} \\ &\leq L^{\frac{2}{q}-1-\frac{2\beta}{d}} \|w_0\|_{B_q^\beta}^2 ,\end{aligned}$$

where we used the assumption on β and where all the constants depend on w_0 only through its B_q^β-norm.

For (C.21) to hold with minimal $L \in \mathbb{N}$, we choose L as

$$L = c \begin{cases} \epsilon^{-\frac{d}{\beta}}, & \text{if } q \geq 2, \\ \epsilon^{\frac{1}{\frac{1}{q}-\frac{1}{2}-\frac{\beta}{d}}}, & \text{if } q < 2, \end{cases} \tag{C.22}$$

where the constant c depends on w_0 only through its B_q^β-norm.

Testing $h_{1:L} \in \mathcal{Z}_\alpha$ in the infimum we aim to bound, yields an upper bound

$$I(\epsilon) := \inf_{h \in \mathcal{Z}_\alpha : \|h - w_0\|_{\ell_2} \leq \epsilon} \|h\|_{\mathcal{Z}_\alpha}^p \leq \|h_{1:L}\|_{\mathcal{Z}_\alpha}^p = \sum_{\ell \leq L} \ell^{\frac{p}{2}+\frac{\alpha p}{d}} |w_{0,\ell}|^p.$$

i) For $q \leq p$, we use (C.20) to get

$$\begin{aligned}\|h_{1:L}\|_{\mathcal{Z}_\alpha}^p &= \sum_{\ell \leq L} \ell^{\frac{p}{2}+\frac{\alpha p}{d}} |w_{0,\ell}|^p = \sum_{\ell \leq L} \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} |w_{0,\ell}|^q |w_{0,\ell}|^{p-q} \ell^{\frac{p}{2}+\frac{\alpha p}{d}-\frac{\beta q}{d}-\frac{q}{2}+1} \\ &\leq \|w_0\|_{B_q^\beta}^{p-q} \sum_{\ell \leq L} \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} |w_{0,\ell}|^q \ell^{\frac{(\alpha-\beta)p}{d}+\frac{p}{q}} \\ &\leq \begin{cases} \|w_0\|_{B_q^\beta}^p , & \text{if } \beta \geq \alpha + \frac{d}{q}, \\ \|w_0\|_{B_q^\beta}^p L^{\frac{(\alpha-\beta)p}{d}+\frac{p}{q}}, & \text{if } \beta < \alpha + \frac{d}{q}. \end{cases}\end{aligned} \tag{C.23}$$

ii) For $q > p$, we use Hölder inequality with $(\frac{q}{p}, \frac{q}{q-p})$ to bound

$$\begin{aligned}
\|h_{1:L}\|_{\mathcal{Z}_\alpha}^p &= \sum_{\ell \le L} \ell^{\frac{p}{2}+\frac{\alpha p}{d}} |w_{0,\ell}|^p = \sum_{\ell \le L} \ell^{\frac{\beta p}{d}+\frac{p}{2}-\frac{p}{q}} |w_{0,\ell}|^p \ell^{\frac{\alpha p}{d}-\frac{\beta p}{d}+\frac{p}{q}} \\
&\le \left(\sum_{\ell \le L} \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} |w_{0,\ell}|^q \right)^{\frac{p}{q}} \left(\sum_{\ell \le L} \ell^{(\frac{\alpha p}{d}-\frac{\beta p}{d}+\frac{p}{q})\frac{q}{q-p}} \right)^{\frac{q-p}{q}} \\
&\le \begin{cases} \|w_0\|_{B_q^\beta}^p, & \text{if } \beta \ge \alpha + \frac{d}{p}, \\ \|w_0\|_{B_q^\beta}^p (\log L)^{\frac{q-p}{q}}, & \text{if } \beta = \alpha + \frac{d}{p}, \\ \|w_0\|_{B_q^\beta}^p L^{\frac{\alpha p - \beta p + d}{d}}, & \text{if } \beta < \alpha + \frac{d}{p}. \end{cases} \qquad \text{(C.24)}
\end{aligned}$$

Combining the bounds (C.23) and (C.24) with the choice of L from (C.22) completes the proof. □

Proof of Lemma 5.14. We first show that f and g defined by (18) satisfy (6). Let $h = (h_\ell) \in aB_{\mathcal{Q}_\alpha}$ and define $x_{1:L} = (h_1, ..., h_L, 0, ...)$ for the smallest $L = L(\epsilon; a) \in \mathbb{N}$ such that $\gamma_L \le \frac{\epsilon}{a}$. Such L satisfies

$$L = \lceil a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} \rceil \le a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} + 1.$$

Clearly, $x_{1:L} \in \mathcal{Z}_\alpha$ and we obtain

$$\|h - x_{1:L}\|_{\ell_2}^2 \le \gamma_L^2 \sum_{\ell > L} \gamma_\ell^{-2} h_\ell^2 \le \gamma_L^2 \|h\|_{\mathcal{Q}_\alpha}^2 \le \epsilon^2.$$

Consequently, by applying the Hölder inequality for $\left(\frac{2}{p}, \frac{2}{2-p}\right)$ we obtain

$$\begin{aligned}
\inf_{x \in \mathcal{Z}_\alpha : \|h-x\|_{\ell^2} \le \epsilon} \|x\|_{\mathcal{Z}_\alpha}^p &\le \sum_{\ell \le L} \gamma_\ell^{-p} |h_\ell|^p \\
&\le L^{1-\frac{p}{2}} \left(\sum_{\ell \le L} \gamma_\ell^{-2} h_\ell^2 \right)^{\frac{p}{2}} \\
&\le (a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} + 1)^{1-\frac{p}{2}} a^p \\
&\le 2^{1-\frac{p}{2}} (a^{\frac{2d}{d+2\alpha}} \epsilon^{-\frac{2d}{d+2\alpha}} \vee 1)^{1-\frac{p}{2}} a^p \\
&\le 2^{1-\frac{p}{2}} (a^{\frac{2d-pd}{d+2\alpha}} \vee 1)(\epsilon^{-\frac{2d-pd}{d+2\alpha}} \vee 1) a^p \\
&= f(a) g(\epsilon)^{1-\frac{p}{2}}.
\end{aligned}$$

For the second part of the claim, let $\epsilon_n = r_n^{\alpha,\beta,p,q}$ as in Proposition 5.4 and observe that since $n\epsilon_n^2 > 1$ and $p \le 2$ we have

$$f(n^{\frac{1}{2}} \epsilon_n) = n^{\frac{d+\alpha p}{d+2\alpha}} \epsilon_n^{\frac{2d+2\alpha p}{d+2\alpha}}$$

and since $\epsilon_n < 1$ we have $g(\epsilon_n) = 2\epsilon_n^{-\frac{2d}{d+2\alpha}}$. For $p = 2$ the claim holds trivially since $f(n^{\frac{1}{2}}\epsilon_n) = n\epsilon_n^2$ and $g(\epsilon_n)^{1-\frac{p}{2}} = 1$. For $p \in [1, 2)$, first notice that

$$f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} = 2^{1-\frac{p}{2}}\epsilon_n^p n^{\frac{d+p\alpha}{d+2\alpha}},$$

so that $f(n^{\frac{1}{2}}\epsilon_n)g(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2$ is equivalent to

$$\epsilon_n \gtrsim n^{-\frac{\alpha}{d+2\alpha}}. \tag{C.25}$$

This is always true, since in the proof of Proposition 5.4 we computed ϵ_n by finding an upper bound, say $B(\epsilon_n)$, on the concentration function (see for example (C.11)), and then solving $B(\epsilon_n) = n\epsilon_n^2$. Since $B(\epsilon_n) \geq \epsilon^{-\frac{d}{\alpha}}$, we have that (C.25) holds. □

Proof of Lemma 5.15. Fix $a, \epsilon > 0$ and let $h \in aB_{\bar{\mathcal{Q}}_\alpha}$. We need to show that $\bar{f}, \bar{g}$ satisfy the bound

$$\inf_{x\in\bar{\mathcal{Z}}_\alpha:\|h-x\|_{\ell_2}\leq\epsilon} \|x\|^p_{\bar{\mathcal{Z}}_\alpha} \leq \bar{f}(a)\bar{g}(\epsilon)^{1-\frac{p}{2}}. \tag{C.26}$$

Noticing that $aB_{\bar{\mathcal{Q}}_\alpha} = \lambda aB_{\mathcal{Q}_\alpha}$ and using Lemma 5.14, we get that

$$\inf_{x\in\bar{\mathcal{Z}}_\alpha:\|h-x\|_{\ell_2}\leq\epsilon} \|x\|^p_{\bar{\mathcal{Z}}_\alpha} = \inf_{x\in\mathcal{Z}:\|h-x\|_{\ell_2}\leq\epsilon} \lambda^{-p}\|x\|^p_{\mathcal{Z}_\alpha} \leq \lambda^{-p}f(\lambda a)g(\epsilon)^{1-\frac{p}{2}} = \bar{f}(a)\bar{g}(\epsilon)^{1-\frac{p}{2}},$$

hence (C.26) is indeed satisfied.

For the second part of the claim, let $\epsilon_n = \bar{r}_n$, for $\bar{r}_n$ as in Proposition 5.8 and observe that since $n\epsilon_n^2 > 1$ and $p \leq 2$ we have $\bar{f}(n^{\frac{1}{2}}\epsilon_n) = \lambda_n^{\frac{(2-p)d}{d+2\alpha}} n^{\frac{d+\alpha p}{d+2\alpha}} \epsilon_n^{\frac{2d+2\alpha p}{d+2\alpha}}$ and since $\epsilon_n < 1$ we have $g(\epsilon_n) = 2\epsilon_n^{-\frac{2d}{d+2\alpha}}$.

For $p = 2$ the claim holds trivially since $\bar{f}(n^{\frac{1}{2}}\epsilon_n) = n\epsilon_n^2$ and $\bar{g}(\epsilon_n)^{1-\frac{p}{2}} = 1$. For $p \in [1, 2)$, first notice that

$$\bar{f}(n^{\frac{1}{2}}\epsilon_n)\bar{g}(\epsilon_n)^{1-\frac{p}{2}} = 2^{1-\frac{p}{2}}\lambda_n^{\frac{(2-p)d}{d+2\alpha}}\epsilon_n^p n^{\frac{d+p\alpha}{d+2\alpha}},$$

so that $\bar{f}(n^{\frac{1}{2}}\epsilon_n)\bar{g}(\epsilon_n)^{1-\frac{p}{2}} \lesssim n\epsilon_n^2$ is equivalent to

$$\epsilon_n \gtrsim n^{-\frac{\alpha}{d+2\alpha}}\lambda_n^{\frac{d}{d+2\alpha}}. \tag{C.27}$$

This is always true, since in the proof of Proposition 5.8 we computed ϵ_n by finding an upper bound, say $B(\epsilon_n, \lambda_n)$, on the concentration function, and then solving $B(\epsilon_n, \lambda_n) = n\epsilon_n^2$. Since $B(\epsilon_n, \lambda_n) \geq \epsilon_n^{-\frac{d}{\alpha}}\lambda_n^{\frac{d}{\alpha}}$, we have that (C.27) holds. □

D. Proofs of Section 6

Proof of Proposition 6.1. The proof follows the techniques of the proof of [4, Corollary 5] taking into account the form of the Schauder basis functions ψ_{kl}. In particular recall that ψ_{kl} are ϑ-Hölder continuous with $\vartheta \le S\wedge 1$, $S>\alpha$. Denote by $\kappa_n(W)$ the nth cumulant of a random variable W. Let $u\sim\mu$. Since the odd cumulants of centered random variables are zero and the cumulants are additive for independent random variables, we have for any integer $q\ge 1$ and any $x,y\in[0,1], x\neq y$

$$
\begin{aligned}
&|\kappa_{2q}(u(x)-u(y))| = \left|\sum_{k=1}^{\infty}\sum_{l=1}^{2^k}\kappa_{2q}(\xi_{kl})2^{-(1+2\alpha)qk}(\psi_{kl}(x)-\psi_{kl}(y))^{2q}\right|\\
&\le C_q\sum_{k=1}^{\infty}\sum_{l=1}^{2^k}2^{-(1+2\alpha)qk}|\psi_{kl}(x)-\psi_{kl}(y)|^{2q}\\
&\le C_q\sum_{k=1}^{\infty}\min\Big\{\sum_{l=1}^{2^k}2^{2(\vartheta-\alpha)qk}|x-y|^{2q\vartheta}, 2^{-(1+2\alpha)qk}\sum_{l=1}^{2^k}(|\psi_{kl}(x)|^{2q}+|\psi_{kl}(y)|^{2q})\Big\}\\
&\le C_q\sum_{k=1}^{\infty}\min\Big\{2^k 2^{2(\vartheta-\alpha)qk}|x-y|^{2q\vartheta}, 2^{-2\alpha qk}\Big\}\\
&\le C_q\sum_{\ell=1}^{\infty}\min\Big\{\ell^{2(\vartheta-\alpha)q}|x-y|^{2q\vartheta}, \ell^{-2\alpha q-1}\Big\}\\
&\le C_q\left(|x-y|^{2q\vartheta}\int_1^{|x-y|^{\frac{-2q\vartheta}{1+2q\vartheta}}}\ell^{2(\vartheta-\alpha)q}d\ell+\int_{|x-y|^{\frac{-2q\vartheta}{1+2q\vartheta}}}^{\infty}\ell^{-2\alpha q-1}d\ell\right)\\
&\le C_q|x-y|^{\frac{4\alpha\vartheta q^2}{1+2q\vartheta}},
\end{aligned}
$$

where C_q is a constant depending only on (q,p), changing from line to line. For the second inequality we used the ϑ-Hölder continuity of ψ_{kl} as described in (21) and the bound $(a+b)^{2q}\le 2^{2q-1}(a^{2q}+b^{2q})$, for any $a,b\in\mathbb{R}$. For the third inequality we used (22), which in turn implies that

$$\left\|\sum_{l=1}^{2^k}|\psi_{kl}|\right\|_{L_\infty}\le C_2 2^{\frac{k}{2}},$$

by letting $x^*\in\arg\max_x\sum_{l=1}^{2^k}|\psi_{kl}(x)|$ and taking $u_{kl}=\mathrm{sign}\{\psi_{kl}(x^*)\}$.

Since the random variables $u(x)$ are centered, all moments of even order $2q, q\ge 1$, can be written as homogeneous polynomials of the even cumulants of order up to $2q$, hence

$$\mathbb{E}|u(x)-u(y)|^{2q}\le C_q|x-y|^{\frac{4\alpha\vartheta q^2}{1+2q\vartheta}},$$

uniformly for $x,y\in[0,1]$. The result follows from Kolmogorov's continuity theorem, since we can choose q arbitrarily large, see [4, Corollary 4]. □

Proof of Proposition 6.3. As a first step we generalize [14, Lemma 2.1] which holds for standard jointly normal variables, to the case of independent p-exponential variables with $p \in [1,2]$. Due to independence, we can use the product rule instead of Sidak's inequality. The lemma immediately generalizes due to the estimates in Lemma F.1 below.

To get the result we then follow the proof of [14, Theorem 1.3]. For u drawn from an α-regular p-exponential measure, it holds by (22)

$$\|u\|_{L_\infty} \le c \sum_{k=0}^{\infty} 2^{\frac{k}{2}} \sup_{1\le l\le 2^k} |u_{kl}| \le c \sum_{k=0}^{\infty} 2^{-\alpha k} \sup_{1\le l \le 2^k} |\xi_{kl}|.$$

For $\epsilon > 0$, let n be an integer such that

$$\frac{2}{1-2^{-\alpha/2}} 2^{-\alpha n} \le \epsilon < \frac{2}{1-2^{-\alpha/2}} 2^{-\alpha(n-1)}.$$

Define

$$b_k := \begin{cases} 2^{\frac{3}{2}(k-n)\alpha}, & \text{if } k < n \\ 2^{\frac{1}{2}(k-n)\alpha}, & \text{if } k \ge n \end{cases}$$

and notice that

$$\sum_{k=0}^{\infty} 2^{-\alpha k} b_k \le \frac{2}{1-2^{-\alpha/2}} 2^{-\alpha n}.$$

Then, if $|\xi_{kl}| \le b_k$ for all $k \ge 0$, $1 \le l \le 2^k$, we have that $\|u\|_{L_\infty} \le \epsilon$. Therefore, by [14, Lemma 2.1] we have

$$\mathbb{P}(\|u\|_{L_\infty} \le \epsilon) \ge \exp(-C 2^n)$$

and the proof is complete since 2^n is of order $\epsilon^{-\frac{1}{\alpha}}$. □

Proof of Lemma 6.4. Let w_{kl} be the coefficients of w_0 in the wavelet basis ψ_{kl}. Consider $h_{1:K} \in C[0,1]$ with coefficients $h_{kl} = w_{kl}$ for $k \le K$ and $1 \le l \le 2^k$ and $h_{kl} = 0$ for $k > K$. Then $h_{1:K} \in \mathcal{Z}_\alpha$ for any $K \in \mathbb{N}$ and

$$\|h_{1:K} - w_0\|_{L_\infty} \le c \sum_{k>K} 2^{\frac{k}{2}} \sup_{1\le l\le 2^k} |w_{kl}| \le c \|w_0\|_{B^\beta_{\infty\infty}} \sum_{k>K} 2^{-\beta k} \le c \|w_0\|_{B^\beta_{\infty\infty}} 2^{-\beta K},$$

for $c > 0$ a changing constant independent of w_0 and where for the first inequality we used (22). Choosing $K \in \mathbb{N}$ minimal so that $\epsilon \ge c\|w_0\|_{B^\beta_{\infty\infty}} 2^{-\beta K}$, we get $\|h_{1:K} - w_0\|_{L_\infty} < \epsilon$. Then

$$\begin{aligned} \inf_{h\in\mathcal{Z}_\alpha : \|h-w_0\|_{L_\infty}\le\epsilon} \|h\|^p_{\mathcal{Z}_\alpha} &\le \|h_{1:K}\|^p_{\mathcal{Z}_\alpha} = \sum_{k\le K} \sum_{1\le l\le 2^k} |w_{kl}|^p 2^{p(\frac{1}{2}+\alpha)k} 2^{-\beta k} 2^{\beta k} \\ &\le \|w_0\|^p_{B^\beta_{\infty\infty}} \sum_{k\le K} \sum_{1\le l\le 2^k} 2^{p(\alpha-\beta)k} = \|w_0\|^p_{B^\beta_{\infty\infty}} \sum_{k\le K} 2^{(p(\alpha-\beta)+1)k}. \end{aligned}$$

The sum on the right hand side converges as $K \to \infty$ if $\beta > \alpha + \frac{1}{p}$, for $\beta = \alpha + \frac{1}{p}$ blows-up as $K \asymp \log(1/\epsilon)$ and for $\beta < \alpha + \frac{1}{p}$ blows-up as $2^{K((\alpha-\beta)p+1)} \asymp \epsilon^{\frac{\beta p - \alpha p - 1}{\beta}}$. □

Proof of Lemma 6.6. Let $h = (h_{kl}) \in aB_{\mathcal{Q}_\alpha}$ and define the approximation $x_{1:K} = \sum_{k\leq K}\sum_{l=1}^{2^k} h_{kl}\psi_{kl}$ for K to be determined below. We have that $x_{1:K} \in \mathcal{Z}_\alpha, \forall K \in \mathbb{N}$ and by (22) and Cauchy-Schwarz inequality, we get the bound

$$\|h - x_{1:K}\|_{L_\infty} \leq c\left(\sum_{k=K+1}^{\infty} 2^{-2\alpha k}\right)^{\frac{1}{2}}\left(\sum_{k=1}^{\infty} 2^{(1+2\alpha)k}\sup_{1\leq l\leq 2^k}|h_{kl}|^2\right)^{\frac{1}{2}} \leq c2^{-\alpha K}\|h\|_{\mathcal{Q}_\alpha},$$

where $c > 0$ is a constant with a value that changes below, dependent only on the Schauder basis, α and later on p. For $K = K(\epsilon; a) \in \mathbb{N}$ minimal such that $\epsilon \geq ca2^{-\alpha K}$, we have that $\|h - x_{1:K}\|_{L_\infty} \leq \epsilon$. This K satisfies $K = \lceil \log_2(ca^{\frac{1}{\alpha}}\epsilon^{-\frac{1}{\alpha}})\rceil \leq \log_2(ca^{\frac{1}{\alpha}}\epsilon^{-\frac{1}{\alpha}}) + 1$, hence $2^K \leq 2ca^{\frac{1}{\alpha}}\epsilon^{-\frac{1}{\alpha}}$ and by Hölder inequality we have the bound

$$\inf_{x\in\mathcal{Z}_\alpha:\|h-x\|_\infty\leq\epsilon}\|x\|_{\mathcal{Z}_\alpha}^p \leq \|x_{1:K}\|_{\mathcal{Z}_\alpha}^p = \sum_{k\leq K}\sum_{l=1}^{2^k}\frac{|h_{kl}|^p}{\gamma_{kl}^p} \leq (2^K-1)^{1-\frac{p}{2}}\|h\|_{\mathcal{Q}_\alpha}^p \leq c\epsilon^{-\frac{2-p}{2\alpha}}a^{\frac{2-p+2\alpha p}{2\alpha}}.$$

Therefore, $f(a) = ca^{\frac{2-p+2\alpha p}{2\alpha}}$ and $g(\epsilon) = \epsilon^{-\frac{1}{\alpha}}$ satisfy (6) and the proof is complete. □

E. Shift spaces of scaled independent product measures

Proposition E.1. *Let ν be the law of the scaled sequence $(\gamma_\ell\xi_\ell)_{\ell\in\mathbb{N}}$, where ξ_ℓ independent and identically distributed univariate random variables and $\gamma = (\gamma_\ell)$ deterministic decaying sequence of positive numbers. Assume that the common distribution of ξ_ℓ, has finite Fisher information and variance and has a density ρ_ℓ with respect to the Lebesgue measure which is everywhere positive and continuous. Then for any $h \in \mathbb{R}^\infty$ it holds that the translated measure ν_h and ν are either singular or equivalent. The shift space of the measure ν is*

$$\mathcal{Q}(\nu) = \{h \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} h_\ell^2\gamma_\ell^{-2} < \infty\}.$$

Furthermore, letting $\rho_{\ell,h_\ell} = \rho_\ell(\cdot - h_\ell)$, we have for $h \in \mathcal{Q}(\nu)$,

$$\frac{d\nu_h}{d\nu}(u) = \lim_{N\to\infty}\prod_{\ell=1}^{N}\frac{d\rho_{\ell,h_\ell}}{d\rho_\ell}(u_\ell) \qquad in\ \ L^1(\mathbb{R}^\infty, \nu).$$

Proof. The positivity and continuity assumption on the density of ξ_ℓ, secures that for each ℓ we have that ρ_ℓ and the translate $\rho_{\ell,h_\ell} = \rho_\ell(\cdot - h_\ell)$ are equivalent. Hence by the Kakutani Theorem [1, Theorem 2.12.7] ν and ν_h are either singular or equivalent.

The rest of the proof relies on [9, Section 1] which builds on [13]. In these papers it is shown that if $Z = (Z_1, Z_2, \dots)$ is a sequence of independent of random variables with variance $0 < \sigma_j^2 < \infty$, then a sufficient condition for Z and $Z+\alpha$, where $\alpha = (\alpha_1, \alpha_2, \dots)$, to be singular is that $\sum_{j=1}^{\infty}\alpha_j^2\sigma_j^{-2}=\infty$. If in addition the Fisher information I_j of Z_j is finite for all j, then a necessary condition for Z and $Z+\alpha$, to be mutually singular is that $\sum_{j=1}^{\infty}\alpha_j^2 I_j=\infty$.

In our assumed setting, since the Fisher information of $\gamma_\ell\xi_\ell$ is $I_\ell=\gamma_\ell^{-2}I$ where I is, the assumed to be finite, Fisher information of ξ_ℓ, and since $\mathrm{Var}(\gamma_\ell\xi_\ell)=\gamma_\ell^2\mathrm{Var}(\xi_\ell)$, we have that the necessary and sufficient condition for the singularity of ν with ν_h is $\sum_{\ell=1}^{\infty}h_\ell^2\gamma_\ell^{-2}=\infty$. Since ν and ν_h are either singular or equivalent, the shift space is as claimed.

The Radon-Nikodym derivative follows again from Kakutani theorem in the form presented in [3, Theorem 2.7], noting that in [9, Section 1] it is shown that the Hellinger integral $H(\nu,\nu_h)$ is positive when $\sum_{\ell=1}^{\infty}h_\ell^2 I_\ell<\infty$. □

F. Estimates for the univariate *p*-exponential distribution

Lemma F.1. *Let $\xi\sim f_p(x)$, where $f_p(x)\propto\exp(-\frac{|x|^p}{p}), x\in\mathbb{R}$, $p\in[1,2]$. Then there exist constants $0<r_1<1$ and $r_2>0$ depending only on p, such that*

$$\mathbb{P}(|\xi|\le x)\ge\begin{cases} r_1x, & \text{if } 0\le x\le 1\\ \exp(-r_2\exp(-\frac{1}{p}x^p)), & \text{if } x>1.\end{cases}$$

Proof. For $x\le 1$ we have

$$\mathbb{P}(|\xi|\le x)=2\int_0^x c_pe^{-\frac{t^p}{p}}dt\ge 2\int_0^x c_pe^{-1/p}dt\ge r_1x,$$

where $r_1=2c_pe^{-1/p}=\frac{e^{-\frac{1}{p}}p^{-\frac{1}{p}}}{\Gamma(1+\frac{1}{p})}<1$ for $p\in[1,2]$.

For $x>1$, let

$$g_p(x)=\mathbb{P}(|\xi|\le x)-\exp(-r_2\exp(-\frac{1}{p}x^p))=\frac{\int_0^x e^{-\frac{t^p}{p}}dt}{\int_0^\infty e^{-\frac{t^p}{p}}dt}-\exp(-r_2\exp(-\frac{1}{p}x^p)),$$

for some $r_2>0$ large enough, so that $g_p(1)>0$. Such an r_2 exists since the first term in $g_p(1)$ is fixed and positive and the second term is decreasing to zero as r_2 grows. The derivative of $g_p(x)$ is

$$\frac{d}{dx}g_p(x)=e^{-\frac{x^p}{p}}\left(\frac{1}{\int_0^\infty e^{-\frac{t^p}{p}}dt}-r_2x^{p-1}\exp(-r_2\exp(-\frac{1}{p}x^p))\right).$$

The term inside the parenthesis, as $x\ge 1$ grows, starts from a possibly positive value and is monotonically decreasing, eventually becoming negative. This means that the derivative $\frac{d}{dx}g_p(x)$, as $x\ge 1$ grows starts from a possibly positive value and eventually becomes negative too, and thus has at most one root which corresponds to at most a unique critical point of $g_p(x), x\ge 1$, which if exists is a maximum. Noting that $\lim_{x\to+\infty}g_p(x)=0$, and since $g_p(1)>0$, we get that $g_p(x)\ge 0,\forall x>1$ and the proof is complete. □

G. Other technical results

Lemma G.1. *Let $q \geq 1, d \in \mathbb{N}$ and $\beta > \frac{d}{q} - \frac{d}{2}$. Then $B_q^\beta \subset \ell_2$.*

Proof. Let $w := (w_\ell)_{\ell \in \mathbb{N}} \in B_q^\beta$. For $q = 2$ the claim is trivially true. If $q > 2$, then by Hölder inequality for $(\frac{q}{2}, \frac{q}{q-2})$, we have

$$\|w\|_{\ell_2}^2 = \sum_{\ell=1}^{\infty} w_\ell^2 \ell^{2(\frac{\beta}{d}+\frac{1}{2})-\frac{2}{q}} \ell^{\frac{2}{q}-2(\frac{\beta}{d}+\frac{1}{2})} \leq \|w\|_{B_q^\beta}^2 \left(\sum_{\ell=1}^{\infty} \ell^{\frac{2q}{q-2}(\frac{1}{q}-\frac{\beta}{d}-\frac{1}{2})} \right)^{\frac{q-2}{q}},$$

where the last sum is finite if and only if $\beta > \frac{d}{q} - \frac{d}{2}$. If $q < 2$,

$$\|w\|_{\ell_2}^2 = \sum_{\ell=1}^{\infty} w_\ell^2 = \sum_{\ell=1}^{\infty} |w_\ell|^q \ell^{\frac{\beta q}{d}+\frac{q}{2}-1} \ell^{-\frac{\beta q}{d}-\frac{q}{2}+1} |w_\ell|^{2-q} \leq c \, \|w\|_{B_q^\beta}^q ,$$

provided $\ell^{-\frac{\beta q}{d}-\frac{q}{2}+1} |w_\ell|^{2-q} \leq c$. Using that $w \in B_q^\beta$, we have

$$|w_\ell| \leq \|w\|_{B_q^\beta} \, \ell^{-\frac{\beta}{d}-\frac{1}{2}+\frac{1}{q}}.$$

The last estimate gives

$$\ell^{-\frac{\beta q}{d}-\frac{q}{2}+1} |w_\ell|^{2-q} \leq \|w\|_{B_q^\beta}^{2-q} \, \ell^{-\frac{2\beta}{d}-1+\frac{2}{q}},$$

which is bounded for $\beta \geq \frac{d}{q} - \frac{d}{2}$. □

Lemma G.2 (Lemma 3 [12]). *Given $a, b > 0$ consider the functions*

$$r_{a,b}(s) := s^a \log^{-b}(1/s), \ 0 < s < 1,$$

and

$$v_{a,b}(s) := s^{1/a} \log^{b/a}(1/s^{1/a}), \ 0 < s < 1.$$

Then we have that

$$\lim_{s \to 0} \frac{r_{a,b}^{-1}(s)}{v_{a,b}(s)} = 1.$$

H. Minimax and linear-minimax rates and their relationship to posterior contraction rates, in the white noise model

Consider the estimation of a function $w \in L_2[0,1]$ observed under scaled Gaussian white noise, as in Section 4.2. Let d be a metric on $L_2[0,1]$, and consider the minimax risk over a class $\mathcal{F} \subset L_2[0,1]$

$$\inf_{\tilde{w}_n} \sup_{w \in \mathcal{F}} \mathbb{E}_{P_w^{(n)}} [d(\tilde{w}_n, w)], \tag{H.28}$$

where the infimum is taken over all estimators $\tilde{w}_n$ constructed using the sample path $X^{(n)} \sim P_w^{(n)}$. *The minimax rate of estimation in d-risk over* $\mathcal{F}$, is the fastest rate of decay r_n of the minimax risk in (H.28), as $n \to \infty$. See [8, Definition 6.3.1] for details.

We can also consider a more general minimax framework, and in particular can embed convergence in d-loss in probability in the minimax framework; see [16, Remark (2) in Section 2.1]. In this case, we study the minimax risk

$$\inf_{\tilde{w}_n} \sup_{w \in \mathcal{F}} P_w^{(n)}(d(\tilde{w}_n, w) \geq r_n). \tag{H.29}$$

The *minimax rate of estimation in d-loss in probability over* $\mathcal{F}$, is the fastest rate r_n for which the minimax risk in (H.29) vanishes as $n \to \infty$.

In both (H.28) and (H.29), we can restrict the infimum to *linear* estimators, in which case we have the corresponding notions of *linear-minimax rates.* The minimax and linear-minimax rates in L_2-risk (as in (H.28)) under Besov smoothness for function estimation in the white noise model can be found in [5, Theorem 1]; see also (14) and (15) in Section 5.1 of the present article.

The minimax rate in L_2-loss *in probability* (as in (H.29)) is a benchmark for rates of contraction in L_2-loss, because if the posterior contracts at a rate ϵ_n at a $w_0 \in L_2[0,1]$, then the center of the smallest ball containing at least half the posterior mass, is an estimator converging at the same rate ϵ_n in L_2-distance, in $P_{w_0}^{(n)}$-probability; see for example [7, Theorem 8.7]. Furthermore, the typical approach for establishing minimax rates in L_2-risk (more generally d-risk, as in (H.28)) and in particular lower bounds, is by establishing a lower bound in probability and using Markov's inequality to obtain a lower bound in expectation, see for example [8, Section 6.3.1]. Hence, it is implicit in the typical derivation of the minimax rate in L_2-risk, that the same rate is also the minimax rate in L_2-loss in probability. This is indeed the approach used when establishing the minimax rates in L_2-risk for the white noise model under Besov-type smoothness, see for example [8, Section 6.3.3], and hence the minimax rate m_n defined in (14) *is* a benchmark for rates of contraction in L_2-loss under Besov-type smoothness.

We next turn to the question of whether Gaussian priors are fundamentally limited by the minimax rate in L_2-risk over linear estimators. In the white noise model, Gaussian priors are conjugate to the Gaussian likelihood, hence the posterior is also Gaussian. By Anderson's inequality, see for example Proposition 2.4, the posterior mean which is a linear estimator, coincides with the center of the smallest ball containing at least half the posterior mass. Following the train of thought of the previous paragraph, one would thus expect that the contraction rates in L_2-loss of Gaussian priors under Besov-type smoothness cannot be faster than the linear-minimax rate l_n in L_2-risk defined in (15). However, linear-minimax rates in L_2-risk over Besov-bodies are established by directly working with L_2-risk and not by establishing lower bounds in probability, see [5, Section 6] and [6]. In other words, the linear-minimax rates in L_2-risk (as in (H.28)) do not necessarily coincide with the linear-minimax rates in L_2-loss in probability (as in (H.29)), over Besov-bodies. In order to establish that Gaussian priors are fundamentally limited by the linear-minimax rate l_n in L_2-risk, one needs either to establish that l_n is also the linear-minimax rate in L_2-loss in probability (that is, to establish the corresponding lower bound in probability), or to show that, in this Gaussian-conjugate setting, posterior contraction in L_2-loss at a rate ϵ_n implies convergence in expected L_2-distance of the posterior mean at the same rate. Both of these tasks appear to be non-trivial.

An alternative, more modest, approach for establishing that posterior contraction rates in L_2-loss for Gaussian priors under Besov-type smoothness cannot be faster than the linear-minimax rate in L_2-risk, is to study lower bounds on the contraction rate and establish that the upper bounds obtained in this article are sharp. Lower bounds on contraction rates in L_2-loss for Gaussian priors under *Sobolev* smoothness, have been studied in [2, Theorem 2]. Once more, it is not immediately obvious how this result can be generalized to obtain good lower bounds under Besov-type smoothness, since its proof relies on the weighted ℓ_2-type structure of Sobolev spaces.

Although beyond the scope of the present paper, we believe that the question of whether Gaussian priors are limited by linear-minimax rates in L_2-risk is extremely interesting: an affirmative answer would rigorously show that when interested in reconstructing spatially inhomogeneous unknown functions (that is functions in B^{β}_{q} for $q < 2$), from the point of view of contraction rates in L_2-loss, it is better to use a Laplace prior rather than a Gaussian prior.

References

[1] V. I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[2] I. Castillo. Lower bounds for posterior rates with gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299, 2008.

[3] G. Da Prato. *An introduction to infinite-dimensional analysis*. Springer Science & Business Media, 2006.

[4] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification. Vol. 1, 2, 3*, pages 311–428. Springer, Cham, 2017.

[5] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26(3):879–921, 1998.

[6] D. L. Donoho, R. C. Liu, and B. MacGibbon. Minimax risk over hyperrectangles, and implications. *Ann. Statist.*, 18(3):1416–1437, 1990.

[7] S. Ghosal and A. van der Vaart. *Fundamentals of nonparametric Bayesian inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[8] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.

[9] A. Kagan. Statistical approach to some mathematical problems. *Austrian Journal of Statistics*, 32:71–83, 2003.

[10] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.

[11] M. Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2005.

[12] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[13] L. Shepp. Distinguishing a sequence of random variables from a translate of itself. *The Annals of Mathematical Statistics*, 36(4):1107–1112, 1965.

[14] W. Stolz. Some small ball probabilities for Gaussian processes under nonuniform norms. *Journal of Theoretical Probability*, 9(3):613–630, 1996.

[15] M. Talagrand. The supremum of some canonical processes. *American Journal of Mathematics*, 116(2):283–325, 1994.

[16] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.