Appendix

The Appendix is organized as follows. In Section A we prove Propositions 1 and 2. Section B is devoted to the analysis of the bias. We study spectral properties of the diffusion operator L to give sufficient and general conditions for the compactness assumption from Theorem 2 and Proposition 3 to hold. Section C provides concentration inequalities for the operators involved in Proposition 2. We conclude by Section D that gives explicit rates of convergence for the bias when μ is a 1-D Gaussian (this result could be easily extended to higher dimensional Gaussians).

A Proofs of Proposition 1 and 2

Recall that $L_0^2(\mu)$ is the subspace of $L^2(\mu)$ of zero mean functions: $L_0^2(\mu) := \{f \in L^2(\mu),\ \int f(x)d\mu(x) = 0\}$ and that we similarly defined $\mathcal{H}_0 := \mathcal{H} \cap L_0^2(\mu)$. Let us also denote by $\mathbb{R}\mathbb{1}$, the set of constant functions.

Proof of Proposition 1. The proof is simply the following reformulation of Equation (1). Under assumption **(Ass. 1)**:

$$\begin{aligned}
\mathcal{P}_\mu &= \sup_{f \in H^1(\mu) \setminus \mathbb{R}\mathbb{1}} \frac{\int_{\mathbb{R}^d} f(x)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2}{\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x)} \\
&= \sup_{f \in \mathcal{H} \setminus \mathbb{R}\mathbb{1}} \frac{\int_{\mathbb{R}^d} f(x)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2}{\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x)} \\
&= \sup_{f \in \mathcal{H}_0 \setminus \{0\}} \frac{\int_{\mathbb{R}^d} f(x)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2}{\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x)}.
\end{aligned}$$

We then simply note that

$$\left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2 = \left(\left\langle f, \int_{\mathbb{R}^d} K_x d\mu(x) \right\rangle_{\mathcal{H}}\right)^2 = \langle f, m \rangle_{\mathcal{H}}^2 = \langle f, (m \otimes m) f \rangle_{\mathcal{H}}.$$

Similarly,

$$\int_{\mathbb{R}^d} f(x)^2 d\mu(x) = \langle f, \Sigma f \rangle_{\mathcal{H}} \quad \text{and} \quad \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x) = \langle f, \Delta f \rangle_{\mathcal{H}}.$$

Note here that $\mathrm{Ker}(\Delta) \subset \mathrm{Ker}(C)$. Indeed, if $f \in \mathrm{Ker}(\Delta)$, then $\langle f, \Delta f \rangle_{\mathcal{H}} = 0$. Hence, μ-almost everywhere, $\nabla f = 0$ so that f is constant and $Cf = 0$. Note also the previous reasoning shows that $\mathrm{Ker}(\Delta)$ is the subset of $\mathcal{H}$ made of constant functions, and $(\mathrm{Ker}(\Delta))^\perp = \mathcal{H} \cap L_0^2(\mu) = \mathcal{H}_0$.

Thus we can write,

$$\mathcal{P}_\mu = \sup_{f \in \mathcal{H} \setminus \mathrm{Ker}(\Delta)} \frac{\langle f, (\Sigma - m \otimes m) f \rangle_{\mathcal{H}}}{\langle f, \Delta f \rangle_{\mathcal{H}}} = \left\| \Delta^{-1/2} C \Delta^{-1/2} \right\|,$$

where we consider Δ^{-1} as the inverse of Δ restricted to $(\mathrm{Ker}(\Delta))^\perp$ and thus get Proposition 1. □

Proof of Proposition 2. We refer to Lemmas 5 and 6 in Section C for the explicit bounds. We have the following inequalities:

$$\begin{aligned}
\left| \widehat{\mathcal{P}}_\mu - \mathcal{P}_\mu^\lambda \right| &= \left| \left\| \widehat{\Delta}_\lambda^{-1/2} \widehat{C} \widehat{\Delta}_\lambda^{-1/2} \right\| - \left\| \Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2} \right\| \right| \\
&\leqslant \left| \left\| \widehat{\Delta}_\lambda^{-1/2} \widehat{C} \widehat{\Delta}_\lambda^{-1/2} \right\| - \left\| \widehat{\Delta}_\lambda^{-1/2} C \widehat{\Delta}_\lambda^{-1/2} \right\| \right| + \left| \left\| \widehat{\Delta}_\lambda^{-1/2} C \widehat{\Delta}_\lambda^{-1/2} \right\| - \left\| \Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2} \right\| \right| \\
&\leqslant \left\| \widehat{\Delta}_\lambda^{-1/2} (\widehat{C} - C) \widehat{\Delta}_\lambda^{-1/2} \right\| + \left| \left\| C^{1/2} \widehat{\Delta}_\lambda^{-1} C^{1/2} \right\| - \left\| C^{1/2} \Delta_\lambda^{-1} C^{1/2} \right\| \right| \\
&\leqslant \left\| \widehat{\Delta}_\lambda^{-1/2} (\widehat{C} - C) \widehat{\Delta}_\lambda^{-1/2} \right\| + \left\| C^{1/2} (\widehat{\Delta}_\lambda^{-1} - \Delta_\lambda^{-1}) C^{1/2} \right\|.
\end{aligned}$$

Consider an event where the estimates of Lemmas 5, 6 and 7 hold for a given value of $\delta > 0$. A simple computation shows that this event has a probability $1 - 3\delta$ at least. We study the two terms above separately. First, provided that $n \geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}$ and $\lambda \in (0, \|\Delta\|]$ in order to use Lemmas 6 and 7,

$$\begin{aligned}
\left\|\widehat{\Delta}_\lambda^{-1/2}(\widehat{C}-C)\widehat{\Delta}_\lambda^{-1/2}\right\| &= \left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\Delta_\lambda^{-1/2}(\widehat{C}-C)\Delta_\lambda^{-1/2}\Delta_\lambda^{1/2}\widehat{\Delta}_\lambda^{-1/2}\right\| \\
&\leqslant \underbrace{\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2}_{\text{Lemma 7}}\underbrace{\left\|\Delta_\lambda^{-1/2}(\widehat{C}-C)\Delta_\lambda^{-1/2}\right\|}_{\text{Lemma 5}} \\
&\leqslant 2\,(\text{Lemma 5}).
\end{aligned}$$

For the second term,

$$\begin{aligned}
\left\|C^{1/2}(\widehat{\Delta}_\lambda^{-1}-\Delta_\lambda^{-1})C^{1/2}\right\| &= \left\|C^{1/2}\widehat{\Delta}_\lambda^{-1}(\Delta-\widehat{\Delta})\Delta_\lambda^{-1}C^{1/2}\right\| \\
&= \left\|C^{1/2}\Delta_\lambda^{-1/2}\Delta_\lambda^{1/2}\widehat{\Delta}_\lambda^{-1}\Delta_\lambda^{1/2}\Delta_\lambda^{-1/2}(\Delta-\widehat{\Delta})\Delta_\lambda^{-1/2}\Delta_\lambda^{-1/2}C^{1/2}\right\| \\
&\leqslant \underbrace{\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2}_{\text{Lemma 7}}\underbrace{\left\|C^{1/2}\Delta_\lambda^{-1/2}\right\|^2}_{\mathcal{P}_\mu^\lambda}\underbrace{\left\|\Delta_\lambda^{-1/2}(\Delta-\widehat{\Delta})\Delta_\lambda^{-1/2}\right\|}_{\text{Lemma 6}} \\
&\leqslant 2\cdot\mathcal{P}_\mu^\lambda\cdot(\text{Lemma 6}).
\end{aligned}$$

The leading order term in the estimate of Lemma 6 is of order $\left(\frac{2\mathcal{K}_d\log(4\mathrm{Tr}\Delta/\lambda\delta)}{\lambda n}\right)^{1/2}$ whereas the leading one in Lemma 5 is of order $\frac{8\mathcal{K}\log(2/\delta)}{\lambda\sqrt{n}}$. Hence, the latter is the dominant term in the final estimation. □

B Analysis of the bias: convergence of the regularized Poincaré constant to the true one

We begin this section by proving Proposition 3. We then investigate the compactness condition required in the assumptions of Proposition 3 by studying the spectral properties of the diffusion operator L. In Proposition 6, we derive, under some general assumption on the RKHS and usual growth conditions on V, some convergence rate for the bias term.

B.1 General condition for consistency: proof of Proposition 3

To prove Proposition 3, we first need a general result on operator norm convergence.

Lemma 1. *Let $\mathcal{H}$ be a Hilbert space and suppose that $(A_n)_{n\geqslant 0}$ is a family of bounded operators such that $\forall n \in \mathbb{N}$, $\|A_n\| \leqslant 1$ and $\forall f \in \mathcal{H}$, $A_n f \xrightarrow{n\to\infty} Af$. Suppose also that B is a compact operator. Then, in operator norm,*

$$A_n B A_n^* \xrightarrow{n\to\infty} ABA^*.$$

Proof. Let $\varepsilon > 0$. As B is compact, it can be approximated by a finite rank operator $B_{n_\varepsilon} = \sum_{i=1}^{n_\varepsilon} b_i\langle f_i, \cdot\rangle g_i$, where $(f_i)_i$ and $(g_i)_i$ are orthonormal bases, and $(b_i)_i$ is a sequence of nonnegative numbers with limit zero (singular values of the operator). More precisely, n_ε is chosen so that

$$\|B - B_{n_\varepsilon}\| \leqslant \frac{\varepsilon}{2}.$$

Moreover, ε being fixed, $A_n B_{n_\varepsilon} A_n^* = \sum_{i=1}^{n_\varepsilon} b_i\langle A_n f_i, \cdot\rangle A_n g_i \underset{n\infty}{\longrightarrow} \sum_{i=1}^{n_\varepsilon} b_i\langle Af_i, \cdot\rangle Ag_i = AB_{n_\varepsilon}A^*$ in operator norm, so that, for $n \geqslant N_\varepsilon$, with $N_\varepsilon \geqslant n_\varepsilon$ sufficiently large, $\|A_n B_{n_\varepsilon} A_n^* - AB_{n_\varepsilon}A^*\| \leqslant \frac{\varepsilon}{2}$. Finally, as $\|A\| \leqslant 1$, it holds, for $n \geqslant N_\varepsilon$

$$\begin{aligned}
\|A_n B_{n_\varepsilon} A_n^* - ABA^*\| &\leqslant \|A_n B_{n_\varepsilon} A_n^* - AB_{n_\varepsilon}A^*\| + \|A(B_{n_\varepsilon} - B)A^*\| \\
&\leqslant \|A_n B_{n_\varepsilon} A_n^* - AB_{n_\varepsilon}A^*\| + \|B_{n_\varepsilon} - B\| \leqslant \varepsilon.
\end{aligned}$$

This proves the convergence in operator norm of $A_n B A_n^*$ to ABA^* when n goes to infinity. □

We can now prove Proposition 3.

Proof of Proposition 3. Let $\lambda > 0$, we want to show that

$$\mathcal{P}_\mu^\lambda = \|\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2}\| \underset{\lambda \to 0}{\longrightarrow} \|\Delta^{-1/2} C \Delta^{-1/2}\| = \mathcal{P}_\mu.$$

Actually, with Lemma 1, we will show a stronger result which is the norm convergence of the operator $\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2}$ to $\Delta^{-1/2} C \Delta^{-1/2}$. Indeed, denoting by $B = \Delta^{-1/2} C \Delta^{-1/2}$ and by $A_\lambda = \Delta_\lambda^{-1/2} \Delta^{1/2}$ both defined on $\mathcal{H}_0$, we have $\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2} = A_\lambda B A_\lambda^*$ with B compact and $\|A_\lambda\| \leqslant 1$. Furthermore, let $(\phi_i)_{i \in \mathbb{N}}$ be an orthonormal family of eigenvectors of the compact operator Δ associated to eigenvalues $(\nu_i)_{i \in \mathbb{N}}$. Then we can write, for any $f \in \mathcal{H}_0$,

$$A_\lambda f = \Delta_\lambda^{-1/2} \Delta^{1/2} f = \sum_{i=0}^{\infty} \sqrt{\frac{\nu_i}{\lambda + \nu_i}} \langle f, \phi_i \rangle_{\mathcal{H}} \, \phi_i \underset{\lambda \to 0}{\longrightarrow} f.$$

Hence by applying Lemma 1, we have the convergence in operator norm of $\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2}$ to $\Delta^{-1/2} C \Delta^{-1/2}$, hence in particular the convergence of the norms of the operators.

□

B.2 Introduction of the operator L

In all this section we focus on a distribution $d\mu$ of the form $d\mu(x) = \mathrm{e}^{-V(x)} dx$.

Let us give first a characterization of the function that allows to recover the Poincaré constant, i.e., the function in $H^1(\mu)$ that minimizes $\frac{\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x)}{\int_{\mathbb{R}^d} f(x)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2}$. We call f_* this function. We recall that we denote by Δ^L the standard Laplacian in $\mathbb{R}^d$: $\forall f \in H^1(\mu)$, $\Delta^L f = \sum_{i=1}^d \frac{\partial^2 f_i}{\partial^2 x i}$. Let us define the operator $\forall f \in H^1(\mu)$, $Lf = -\Delta^L f + \langle \nabla V, \nabla f \rangle$, which is the opposite of the infinitesimal generator of the dynamics (3). We can verify that it is symmetric in $L^2(\mu)$. Indeed by integrations by parts for any $\forall f, g \in C_c^\infty$,

$$\begin{aligned}
\langle Lf, g \rangle_{L^2(\mu)} &= \int (Lf)(x) g(x) d\mu(x) \\
&= -\int \Delta^L f(x) g(x) \mathrm{e}^{-V(x)} dx + \int \langle \nabla V(x), \nabla f(x) \rangle g(x) \mathrm{e}^{-V(x)} dx \\
&= \int \left\langle \nabla f(x), \nabla \left(g(x) \mathrm{e}^{-V(x)} \right) \right\rangle dx + \int \langle \nabla V(x), \nabla f(x) \rangle g(x) \mathrm{e}^{-V(x)} dx \\
&= \int \langle \nabla f(x), \nabla g(x) \rangle \mathrm{e}^{-V(x)} dx - \int \langle \nabla f(x), \nabla V(x) \rangle g(x) \mathrm{e}^{-V(x)} dx \\
&\qquad + \int \langle \nabla V(x), \nabla f(x) \rangle g(x) \mathrm{e}^{-V(x)} dx \\
&= \int \langle \nabla f(x), \nabla g(x) \rangle d\mu(x).
\end{aligned}$$

The last equality being totally symmetric in f and g, we have the symmetry of the operator L: $\langle Lf, g \rangle_{L^2(\mu)} = \int \langle \nabla f, \nabla g \rangle d\mu = \langle f, Lg \rangle_{L^2(\mu)}$ (for the self-adjointness we refer to [Bakry et al., 2014]). Remark that the same calculation shows that $\nabla^* = -\mathrm{div} + \nabla V \cdot$, hence $L = \nabla^* \cdot \nabla = -\Delta^L + \langle \nabla V, \nabla \cdot \rangle$, where ∇^* is the adjoint of ∇ in $L^2(\mu)$.

Let us call π the orthogonal projector of $L^2(\mu)$ on constant functions: $\pi f : x \in \mathbb{R}^d \mapsto \int f d\mu$. The problem (4) then rewrites:

$$\mathcal{P}^{-1} = \inf_{f \in (H^1(\mu) \cap L_0^2(\mu)) \setminus \{0\}} \frac{\langle Lf, f \rangle_{L^2(\mu)}}{\|(I_{L^2(\mu)} - \pi) f\|^2}, \tag{8}$$

Until the end of this part, to alleviate the notation we omit to mention that the scalar product is the canonical one on $L^2(\mu)$. In the same way, we also denote $\mathbb{1} = I_{L^2(\mu)}$.

B.2.1 Case where $d\mu$ has infinite support

Proposition 4 (Properties of the minimizer)**.** *If* $\lim_{|x|\to\infty} \frac{1}{4}|\nabla V|^2 - \frac{1}{2}\Delta^L V = +\infty$, *the problem* (8) *admits a minimizer in* $H^1(\mu)$ *and every minimizer* f *is an eigenvector of* L *associated with the eigenvalue* $\mathcal{P}^{-1}$*:*

$$Lf = \mathcal{P}^{-1} f. \tag{9}$$

To prove the existence of a minimizer in $H^1(\mu)$, we need the following lemmas.

Lemma 2 (Criterion for compact embedding of $H^1(\mu)$ in $L^2(\mu)$)**.** *The injection $H^1(\mu) \hookrightarrow L^2(\mu)$ is compact if and only if the Schrödinger operator $-\Delta^L + \frac{1}{4}|\nabla V|^2 - \frac{1}{2}\Delta^L V$ has compact resolvent.*

Proof. See [Gansberger, 2010, Proposition 1.3] or [Reed and Simon, 2012, Lemma XIII.65]. □

Lemma 3 (A sufficient condition)**.** *If $\Phi \in C^\infty$ and $\Phi(x) \longrightarrow +\infty$ when $|x| \to \infty$, the Schrödinger operator $-\Delta^L + \Phi$ on $\mathbb{R}^d$ has compact resolvent.*

Proof. See [Helffer and Nier, 2005, Section 3] or [Reed and Simon, 2012, Lemma XIII.67]. □

Now we can prove Proposition 4.

Proof of Proposition 4. We first prove that (8) admits a minimizer in $H^1(\mu)$. Indeed, we have,

$$\mathcal{P}^{-1} = \inf_{f\in(H^1\cap L_0^2)\setminus\{0\}} \frac{\langle Lf, f\rangle_{L^2(\mu)}}{\|(\mathbb{1}-\pi)f\|^2} = \inf_{f\in(H^1\cap L_0^2)\setminus\{0\}} J(f), \quad \text{where } J(f) := \frac{\|\nabla f\|^2}{\|f\|^2}.$$

Let $(f_n)_{n\geqslant 0}$ be a sequence of functions in $H_0^1(\mu)$ equipped with the natural H^1-norm such that $(J(f_n))_{n\geqslant 0}$ converges to $\mathcal{P}^{-1}$. As the problem in invariant by rescaling of f, we can assume that $\forall n \geqslant 0$, $\|f_n\|^2_{L^2(\mu)} = 1$. Hence $J(f_n) = \|\nabla f_n\|^2_{L^2(\mu)}$ converges (to $\mathcal{P}^{-1}$). In particular $\|\nabla f_n\|^2_{L^2(\mu)}$ is bounded in $L^2(\mu)$, hence $(f_n)_{n\geqslant 0}$ is bounded in $H^1(\mu)$. Since by Lemma 2 and 3 we have a compact injection of $H^1(\mu)$ in $L^2(\mu)$, it holds, upon extracting a subsequence, that there exists $f \in H^1(\mu)$ such that

$$\begin{cases} f_n \to f & \text{strongly in } L^2(\mu) \\ f_n \rightharpoonup f & \text{weakly in } H^1(\mu). \end{cases}$$

Thanks to the strong $L^2(\mu)$ convergence, $\|f\|^2 = \lim_{n\infty} \|f_n\|^2 = 1$. By the Cauchy-Schwarz inequality and then taking the limit $n \to +\infty$,

$$\|\nabla f\|^2 = \lim_{n\infty} \langle \nabla f_n, \nabla f\rangle \leqslant \lim_{n\infty} \|\nabla f\|\|\nabla f_n\| = \|\nabla f\|\mathcal{P}^{-1}.$$

Therefore, $\|\nabla f\| \leqslant \mathcal{P}^{-1/2}$ which implies that $J(f) \leqslant \mathcal{P}^{-1}$, and so $J(f) = \mathcal{P}^{-1}$. This shows that f is a minimizer of J.

Let us next prove the PDE characterization of minimizers. A necessary condition on a minimizer f_* of the problem $\inf_{f\in H^1(\mu)}\{\|\nabla f\|_{L^2(\mu)},\ \|f\|^2 = 1\}$ is to satisfy the following Euler-Lagrange equation: there exists $\beta \in \mathbb{R}$ such that:

$$Lf_* + \beta f_* = 0.$$

Plugging this into (8), we have: $\mathcal{P}^{-1} = \langle Lf_*, f_*\rangle = -\beta\langle f_*, f_*\rangle = -\beta\|f_*\|_2^2 = -\beta$. Finally, the equation satisfied by f_* is:

$$Lf = -\Delta^L f_* + \langle \nabla V, \nabla f_*\rangle = \mathcal{P}^{-1} f_*,$$

which concludes the proof. □

B.2.2 Case where $d\mu$ has compact support

We suppose in this section that $d\mu$ has a compact support included in Ω. Without loss of generality we can take a set Ω with a C^∞ smooth boundary $\partial\Omega$. In this case, without changing the result of the variational problem, we can restrict ourselves to functions that vanish at the boundary, namely the Sobolev space $H^1_D(\mathbb{R}^d, d\mu) = \{f \in H^1(\mu) \text{ s.t. } f_{|\partial\Omega} = 0\}$. Note that, as V is smooth, $H^1(\mu) \supset H^1(\mathbb{R}^d, d\lambda)$ the usual "flat" space equipped with $d\lambda$, the Lebesgue measure. Note also that only in this section the domain of the operator L is $H^2 \cap H^1_D$.

Proposition 5 (Properties of the minimizer in the compact support case)**.** *The problem (8) admits a minimizer in H^1_D and every minimizer f satisfies the partial differential equation:*

$$Lf = \mathcal{P}^{-1} f. \tag{10}$$

Proof. The proof is exactly the same than the one of Proposition 4 since H^1_D can be compactly injected in L^2 without any additional assumption on V. □

Let us take in this section $\mathcal{H} = H^d(\mathbb{R}^d, d\lambda)$, which is the RKHS associated to the kernel $k(x, x') = \mathrm{e}^{-\|x-x'\|}$. As f_* satisfies (10), from regularity properties of elliptic PDEs, we infer that f_* is $C^\infty(\overline{\Omega})$. By the Whitney extension theorem [Whitney, 1934], we can extend f_* defined on $\overline{\Omega}$ to a smooth and compactly supported function in $\Omega' \supset \Omega$ of $\mathbb{R}^d$. Hence $f_* \in C_c^\infty(\mathbb{R}^d) \subset \mathcal{H}$.

Proposition 6. *Consider a minimizer f_* of (8). Then*

$$\mathcal{P}^{-1} \leqslant \mathcal{P}_\lambda^{-1} \leqslant \mathcal{P}^{-1} + \lambda \frac{\|f_*\|^2_{\mathcal{H}}}{\|f_*\|^2_{L^2(\mu)}}. \tag{11}$$

Proof. First note that f_* has mean zero with respect to $d\mu$. Indeed, $\int f d\mu = \mathcal{P}^{-1} \int Lf d\mu = 0$, by the fact that $d\mu$ is the stationary distribution of the dynamics.

For $\lambda > 0$,

$$\begin{aligned}
\mathcal{P}^{-1} \leqslant \mathcal{P}_\lambda^{-1} &= \inf_{f \in \mathcal{H} \setminus \mathbb{R}\mathbb{1}} \frac{\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x) + \lambda \|f\|^2_{\mathcal{H}}}{\int_{\mathbb{R}^d} f(x)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x) d\mu(x)\right)^2} \\
&\leqslant \frac{\int_{\mathbb{R}^d} \|\nabla f_*(x)\|^2 d\mu(x) + \lambda \|f_*\|^2_{\mathcal{H}}}{\int_{\mathbb{R}^d} f_*(x)^2 d\mu(x)} = \mathcal{P}^{-1} + \lambda \frac{\|f_*\|^2_{\mathcal{H}}}{\|f_*\|^2_{L^2(\mu)}},
\end{aligned}$$

which provides the result. □

C Technical inequalities

C.1 Concentration inequalities

We first begin by recalling some concentration inequalities for sums of random vectors and operators.

Proposition 7 (Bernstein's inequality for sums of random vectors)**.** *Let $z_1, \ldots, z_n$ be a sequence of independent identically and distributed random elements of a separable Hilbert space $\mathcal{H}$. Assume that $\mathbb{E}\|z_1\| < +\infty$ and note $\mu = \mathbb{E} z_1$. Let $\sigma, L \geqslant 0$ such that,*

$$\forall p \geqslant 2, \qquad \mathbb{E}\, \|z_1 - \mu\|^p_{\mathcal{H}} \leqslant \frac{1}{2} p! \sigma^2 L^{p-2}.$$

Then, for any $\delta \in (0, 1]$,

$$\left\| \frac{1}{n} \sum_{i=1}^n z_i - \mu \right\|_{\mathcal{H}} \leqslant \frac{2L \log(2/\delta)}{n} + \sqrt{\frac{2\sigma^2 \log(2/\delta)}{n}}, \tag{12}$$

with probability at least $1 - \delta$.

Proof. This is a restatement of Theorem 3.3.4 of [Yurinsky, 1995]. □

Proposition 8 (Bernstein's inequality for sums of random operators)**.** *Let $\mathcal{H}$ be a separable Hilbert space and let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed self-adjoint random operators on $\mathcal{H}$. Assume that $\mathbb{E}(X_i) = 0$ and that there exist $T > 0$ and S a positive trace-class operator such that $\|X_i\| \leqslant T$ almost surely and $\mathbb{E}X_i^2 \preccurlyeq S$ for any $i \in \{1, \ldots, n\}$. Then, for any $\delta \in (0, 1]$, the following inequality holds:*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i \right\| \leqslant \frac{2T\beta}{3n} + \sqrt{\frac{2\|S\|\beta}{n}}, \tag{13}$$

with probability at least $1 - \delta$ and where $\beta = \log \frac{2 \operatorname{Tr} S}{\|S\|\delta}$.

Proof. The theorem is a restatement of Theorem 7.3.1 of [Tropp, 2012] generalized to the separable Hilbert space case by means of the technique in Section 4 of [Stanislav, 2017]. □

C.2 Operator bounds

Lemma 4. *Under assumptions **(Ass. 2)** and **(Ass. 3)**, Σ, C and Δ are trace-class operators.*

Proof. We only prove the result for Δ, the proof for Σ and C being similar. Consider an orthonormal basis $(\phi_i)_{i \in \mathbb{N}}$ of $\mathcal{H}$. Then, as Δ is a positive self adjoint operator,

$$\begin{aligned} \operatorname{Tr} \Delta = \sum_{i=1}^{\infty} \langle \Delta \phi_i, \phi_i \rangle = \sum_{i=1}^{\infty} \mathbb{E}_\mu \left[\sum_{j=1}^{d} \langle \partial_j K_x, \phi_i \rangle^2 \right] = \mathbb{E}_\mu \left[\sum_{i=1}^{\infty} \sum_{j=1}^{d} \langle \partial_j K_x, \phi_i \rangle^2 \right] \\ = \mathbb{E}_\mu \left[\sum_{j=1}^{d} \|\partial_j K_x\|^2 \right] \leqslant \mathcal{K}_d. \end{aligned}$$

Hence, Δ is a trace-class operator. □

The following quantities are useful for the estimates in this section:

$$\mathcal{N}_\infty(\lambda) = \sup_{x \in \operatorname{supp}(\mu)} \left\| \Delta_\lambda^{-1/2} K_x \right\|_{\mathcal{H}}^2, \text{ and } \quad \mathcal{F}_\infty(\lambda) = \sup_{x \in \operatorname{supp}(\mu)} \left\| \Delta_\lambda^{-1/2} \nabla K_x \right\|_{\mathcal{H}}^2.$$

Note that under assumption **(Ass. 3)**, $\mathcal{N}_\infty(\lambda) \leqslant \frac{\mathcal{K}}{\lambda}$ and $\mathcal{F}_\infty(\lambda) \leqslant \frac{\mathcal{K}_d}{\lambda}$. Note also that under refined assumptions on the spectrum of Δ, we could have a better dependence of the latter bounds with respect to λ. Let us now state three useful lemmas to bound the norms of the operators that appear during the proof of Proposition 2.

Lemma 5. *For any $\lambda > 0$ and any $\delta \in (0, 1]$,*

$$\begin{aligned} \left\| \Delta_\lambda^{-1/2} (\widehat{C} - C) \Delta_\lambda^{-1/2} \right\| \leqslant \frac{4\mathcal{N}_\infty(\lambda) \log \frac{2 \operatorname{Tr} \Sigma}{\mathcal{P}_\mu^\lambda \lambda \delta}}{3n} + \left[\frac{2\, \mathcal{P}_\mu^\lambda\, \mathcal{N}_\infty(\lambda) \log \frac{2 \operatorname{Tr} \Sigma}{\mathcal{P}_\mu^\lambda \lambda \delta}}{n} \right]^{1/2} \\ + 8\mathcal{N}_\infty(\lambda) \left(\frac{\log(\frac{2}{\delta})}{n} + \sqrt{\frac{\log(\frac{2}{\delta})}{n}} \right) \\ + 16\mathcal{N}_\infty(\lambda) \left(\frac{\log(\frac{2}{\delta})}{n} + \sqrt{\frac{\log(\frac{2}{\delta})}{n}} \right)^2, \end{aligned}$$

with probability at least $1 - \delta$.

Proof of Lemma 5. We apply some concentration inequality to the operator $\Delta_\lambda^{-1/2}\widehat{C}\Delta_\lambda^{-1/2}$ whose mean is exactly $\Delta_\lambda^{-1/2}C\Delta_\lambda^{-1/2}$. The calculation is the following:

$$\begin{aligned}
\left\|\Delta_\lambda^{-1/2}(\widehat{C}-C)\Delta_\lambda^{-1/2}\right\| &= \left\|\Delta_\lambda^{-1/2}\widehat{C}\Delta_\lambda^{-1/2} - \Delta_\lambda^{-1/2}C\Delta_\lambda^{-1/2}\right\| \\
&\leqslant \left\|\Delta_\lambda^{-1/2}\widehat{\Sigma}\Delta_\lambda^{-1/2} - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right\| \\
&\quad + \left\|\Delta_\lambda^{-1/2}(\widehat{m}\otimes\widehat{m})\Delta_\lambda^{-1/2} - \Delta_\lambda^{-1/2}(m\otimes m)\Delta_\lambda^{-1/2}\right\| \\
&= \left\|\frac{1}{n}\sum_{i=1}^n \left[(\Delta_\lambda^{-1/2}K_{x_i})\otimes(\Delta_\lambda^{-1/2}K_{x_i}) - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right]\right\| \\
&\quad + \left\|(\Delta_\lambda^{-1/2}\widehat{m})\otimes(\Delta_\lambda^{-1/2}\widehat{m}) - (\Delta_\lambda^{-1/2}m)\otimes(\Delta_\lambda^{-1/2}m)\right\|.
\end{aligned}$$

We estimate the two terms separately.

Bound on the first term: we use Proposition 8. To do this, we bound for $i \in [\![1,n]\!]$:

$$\begin{aligned}
\left\|(\Delta_\lambda^{-1/2}K_{x_i})\otimes(\Delta_\lambda^{-1/2}K_{x_i}) - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right\| &\leqslant \left\|\Delta_\lambda^{-1/2}K_{x_i}\right\|_{\mathcal{H}}^2 + \left\|\Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right\| \\
&\leqslant 2\mathcal{N}_\infty(\lambda),
\end{aligned}$$

and, for the second order moment,

$$\begin{aligned}
&\mathbb{E}\left((\Delta_\lambda^{-1/2}K_{x_i})\otimes(\Delta_\lambda^{-1/2}K_{x_i}) - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right)^2 \\
&\quad = \mathbb{E}\left[\left\|\Delta_\lambda^{-1/2}K_{x_i}\right\|_{\mathcal{H}}^2 (\Delta_\lambda^{-1/2}K_{x_i})\otimes(\Delta_\lambda^{-1/2}K_{x_i})\right] - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1}\Sigma\Delta_\lambda^{-1/2} \\
&\quad \preccurlyeq \mathcal{N}_\infty(\lambda)\Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}.
\end{aligned}$$

We conclude this first part of the proof by some estimation of the constant $\beta = \log\frac{2\,\mathrm{Tr}(\Sigma\Delta_\lambda^{-1})}{\left\|\Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right\|\delta}$. Using $\mathrm{Tr}\Sigma\Delta_\lambda^{-1} \leqslant \lambda^{-1}\mathrm{Tr}\Sigma$, it holds $\beta \leqslant \log\frac{2\,\mathrm{Tr}\Sigma}{\mathcal{P}_\mu^\lambda\lambda\delta}$. Therefore,

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n \left[(\Delta_\lambda^{-1/2}K_{x_i})\otimes(\Delta_\lambda^{-1/2}K_{x_i}) - \Delta_\lambda^{-1/2}\Sigma\Delta_\lambda^{-1/2}\right]\right\| \\
&\qquad \leqslant \frac{4\mathcal{N}_\infty(\lambda)\log\frac{2\,\mathrm{Tr}\Sigma}{\mathcal{P}_\mu^\lambda\lambda\delta}}{3n} + \left[\frac{2\,\mathcal{P}_\mu^\lambda\,\mathcal{N}_\infty(\lambda)\log\frac{2\,\mathrm{Tr}\Sigma}{\mathcal{P}_\mu^\lambda\lambda\delta}}{n}\right]^{1/2}.
\end{aligned}$$

Bound on the second term. Denote by $v = \Delta_\lambda^{-1/2}m$ and $\widehat{v} = \Delta_\lambda^{-1/2}\widehat{m}$. A simple calculation leads to

$$\begin{aligned}
\|\widehat{v}\otimes\widehat{v} - v\otimes v\| &\leqslant \|v\otimes(\widehat{v}-v)\| + \|(\widehat{v}-v)\otimes v\| + \|(\widehat{v}-v)\otimes(\widehat{v}-v)\| \\
&\leqslant 2\|v\|\|\widehat{v}-v\| + \|\widehat{v}-v\|^2.
\end{aligned}$$

We bound $\|\widehat{v}-v\|$ with Proposition 7. It holds: $\widehat{v}-v = \Delta_\lambda^{-1/2}(\widehat{m}-m) = \frac{1}{n}\sum_{i=1}^n \Delta_\lambda^{-1/2}(K_{x_i}-m) = \frac{1}{n}\sum_{i=1}^n Z_i$, with $Z_i = \Delta_\lambda^{-1/2}(K_{x_i}-m)$. Obviously for any $i\in[\![1,n]\!]$, $\mathbb{E}(Z_i)=0$, and $\|Z_i\| \leqslant \|\Delta_\lambda^{-1/2}K_{x_i}\| + \|\Delta_\lambda^{-1/2}m\| \leqslant 2\sqrt{\mathcal{N}_\infty(\lambda)}$. Furthermore,

$$\begin{aligned}
\mathbb{E}\|Z_i\|^2 = \mathbb{E}\left\langle\Delta_\lambda^{-1/2}(K_{x_i}-m), \Delta_\lambda^{-1/2}(K_{x_i}-m)\right\rangle &= \mathbb{E}\left\|\Delta_\lambda^{-1/2}K_{x_i}\right\|^2 - \left\|\Delta_\lambda^{-1/2}m\right\|^2 \\
&\leqslant \mathcal{N}_\infty(\lambda).
\end{aligned}$$

Thus, for $p \geqslant 2$,

$$\mathbb{E}\|Z_i\|^p \leqslant \mathbb{E}\left(\|Z_i\|^{p-2}\|Z_i\|^2\right) \leqslant \frac{1}{2}p!\left(\sqrt{\mathcal{N}_\infty(\lambda)}\right)^2\left(2\sqrt{\mathcal{N}_\infty(\lambda)}\right)^{p-2},$$

hence, by applying Proposition 7 with $L = 2\sqrt{\mathcal{N}_\infty(\lambda)}$ and $\sigma = \sqrt{\mathcal{N}_\infty(\lambda)}$,

$$\begin{aligned}\|\widehat{v} - v\| &\leqslant \frac{4\sqrt{\mathcal{N}_\infty(\lambda)}\log(2/\delta)}{n} + \sqrt{\frac{2\mathcal{N}_\infty(\lambda)\log(2/\delta)}{n}} \\ &\leqslant 4\sqrt{\mathcal{N}_\infty(\lambda)}\left(\frac{\log(2/\delta)}{n} + \sqrt{\frac{\log(2/\delta)}{n}}\right).\end{aligned}$$

Finally, as $\|v\| \leqslant \sqrt{\mathcal{N}_\infty(\lambda)}$,

$$\begin{aligned}\|\widehat{v}\otimes\widehat{v} - v\otimes v\| \leqslant 8\mathcal{N}_\infty(\lambda)&\left(\frac{\log(2/\delta)}{n} + \sqrt{\frac{\log(2/\delta)}{n}}\right) \\ &+ 16\mathcal{N}_\infty(\lambda)\left(\frac{\log(2/\delta)}{n} + \sqrt{\frac{\log(2/\delta)}{n}}\right)^2.\end{aligned}$$

This concludes the proof of Lemma 5. □

Lemma 6. *For any $\lambda \in (0, \|\Delta\|\,]$ and any $\delta \in (0,1]$,*

$$\left\|\Delta_\lambda^{-1/2}(\widehat{\Delta} - \Delta)\Delta_\lambda^{-1/2}\right\| \leqslant \frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}}{3n} + \sqrt{\frac{2\;\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}}{n}},$$

with probability at least $1-\delta$.

Proof of Lemma 6. As in the proof of Lemma 5, we want to apply some concentration inequality to the operator $\Delta_\lambda^{-1/2}\widehat{\Delta}\Delta_\lambda^{-1/2}$, whose mean is exactly $\Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}$. The proof is almost the same as Lemma 5. We start by writing

$$\begin{aligned}\left\|\Delta_\lambda^{-1/2}(\widehat{\Delta} - \Delta)\Delta_\lambda^{-1/2}\right\| &= \left\|\Delta_\lambda^{-1/2}\widehat{\Delta}\Delta_\lambda^{-1/2} - \Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}\right\| \\ &= \left\|\frac{1}{n}\sum_{i=1}^n\left[(\Delta_\lambda^{-1/2}\nabla K_{x_i})\otimes(\Delta_\lambda^{-1/2}\nabla K_{x_i}) - \Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}\right]\right\|.\end{aligned}$$

In order to use Proposition 8, we bound for $i \in [\![1, n]\!]$,

$$\begin{aligned}\left\|(\Delta_\lambda^{-1/2}\nabla K_{x_i})\otimes(\Delta_\lambda^{-1/2}\nabla K_{x_i}) - \Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}\right\| &\leqslant \left\|\Delta_\lambda^{-1/2}\nabla K_{x_i}\right\|_{\mathcal{H}}^2 + \left\|\Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}\right\| \\ &\leqslant 2\mathcal{F}_\infty(\lambda),\end{aligned}$$

and, for the second order moment,

$$\begin{aligned}&\mathbb{E}\left[\left((\Delta_\lambda^{-1/2}\nabla K_{x_i})\otimes(\Delta_\lambda^{-1/2}\nabla K_{x_i}) - \Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}\right)^2\right] \\ &= \mathbb{E}\left[\left\|\Delta_\lambda^{-1/2}\nabla K_{x_i}\right\|_{\mathcal{H}}^2(\Delta_\lambda^{-1/2}\nabla K_{x_i})\otimes(\Delta_\lambda^{-1/2}\nabla K_{x_i})\right] - \Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1}\Delta\Delta_\lambda^{-1/2} \\ &\preccurlyeq \mathcal{F}_\infty(\lambda)\Delta_\lambda^{-1/2}\Delta\Delta_\lambda^{-1/2}.\end{aligned}$$

We conclude by some estimation of $\beta = \log\frac{2\,\mathrm{Tr}(\Delta\Delta_\lambda^{-1})}{\|\Delta_\lambda^{-1}\Delta\|\delta}$. Since $\mathrm{Tr}(\Delta\Delta_\lambda^{-1}) \leqslant \lambda^{-1}\mathrm{Tr}\Delta$ and for $\lambda \leqslant \|\Delta\|$, $\left\|\Delta_\lambda^{-1}\Delta\right\| \geqslant 1/2$, it follow that $\beta \leqslant \log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}$. The conclusion then follows from (13). □

Lemma 7 (Bounding operators)**.** *For any $\lambda > 0$, $\delta \in (0,1)$, and $n \geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}$,*

$$\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2 \leqslant 2,$$

with probability at least $1-\delta$.

The proof of this result relies on the following lemma (see proof in [Rudi and Rosasco, 2017, Proposition 8]).

Lemma 8. *Let $\mathcal{H}$ be a separable Hilbert space, A and B two bounded self-adjoint positive linear operators on $\mathcal{H}$ and $\lambda > 0$. Then*

$$\left\|(A+\lambda I)^{-1/2}(B+\lambda I)^{1/2}\right\| \leqslant (1-\beta)^{-1/2},$$

with $\beta = \lambda_{\max}\left((B+\lambda I)^{-1/2}(B-A)(B+\lambda I)^{-1/2}\right) < 1$, where $\lambda_{\max}(O)$ is the largest eigenvalue of the self-adjoint operator O.

We can now write the proof of Lemma 7.

Proof of Lemma 7. Thanks to Lemma 8, we see that

$$\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2 \leqslant \left(1-\lambda_{\max}\left(\Delta_\lambda^{-1/2}(\widehat{\Delta}-\Delta)\Delta_\lambda^{-1/2}\right)\right)^{-1},$$

and as $\left\|\Delta_\lambda^{-1/2}(\widehat{\Delta}-\Delta)\Delta_\lambda^{-1/2}\right\| < 1$, we have:

$$\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2 \leqslant \left(1-\left\|\Delta_\lambda^{-1/2}(\widehat{\Delta}-\Delta)\Delta_\lambda^{-1/2}\right\|\right)^{-1}.$$

We can then apply the bound of Lemma 6 to obtain that, if λ is such that $\frac{4\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}}{3n} + \sqrt{\frac{2\ \mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}}{n}} \leqslant \frac{1}{2}$, then $\left\|\widehat{\Delta}_\lambda^{-1/2}\Delta_\lambda^{1/2}\right\|^2 \leqslant 2$ with probability $1-\delta$. The condition on λ is satisfied when $n \geqslant 15\mathcal{F}_\infty(\lambda)\log\frac{4\,\mathrm{Tr}\Delta}{\lambda\delta}$.

□

D Calculation of the bias in the Gaussian case

We can derive a rate of convergence when μ is a one-dimensional Gaussian. Hence, we consider the one-dimensional distribution $d\mu$ as the normal distribution with mean zero and variance $1/(4a)$. Let $b > 0$, we consider also the following approximation $\mathcal{P}_\kappa^{-1} = \inf_{f\in\mathcal{H}} \frac{\mathbb{E}_\mu(f'^2)+\kappa\|f\|_{\mathcal{H}}^2}{\mathrm{var}_\mu(f)}$ where $\mathcal{H}$ is the RKHS associated with the Gaussian kernel $\exp(-b(x-y)^2)$. Our goal is to study how $\mathcal{P}_\kappa$ tends to $\mathcal{P}$ when κ tends to zero.

Proposition 9 (Rate of convergence for the bias in the one-dimensional Gaussian case)**.** *If $d\mu$ is a one-dimensional Gaussian of mean zero and variance $1/(4a)$ there exists $A > 0$ such that, if $\lambda \leqslant A$, it holds*

$$\mathcal{P}^{-1} \leqslant \mathcal{P}_\lambda^{-1} \leqslant \mathcal{P}^{-1}(1+B\lambda\ln^2(1/\lambda)), \tag{14}$$

where A and B depend only on the constant a.

We will show it by considering a specific orthonormal basis of $L^2(\mu)$, where all operators may be expressed simply in closed form.

D.1 An orthonormal basis of $L^2(\mu)$ and $\mathcal{H}$

We begin by giving an explicit a basis of $L^2(\mu)$ which is also a basis of $\mathcal{H}$.

Proposition 10 (Explicit basis)**.** *We consider*

$$f_i(x) = \left(\frac{c}{a}\right)^{1/4}\left(2^i i!\right)^{-1/2}\mathrm{e}^{-(c-a)x^2}H_i\left(\sqrt{2c}x\right),$$

where H_i is the i-th Hermite polynomial, and $c = \sqrt{a^2+2ab}$. Then,

- *$(f_i)_{i\geqslant 0}$ is an orthonormal basis of $L^2(\mu)$;*
- *$\tilde{f}_i = \lambda_i^{1/2} f_i$ forms an orthonormal basis of $\mathcal{H}$, with $\lambda_i = \sqrt{\frac{2a}{a+b+c}}\left(\frac{b}{a+b+c}\right)^i$.*

Proof. We can check that this is indeed an orthonormal basis of $L^2(\mu)$:

$$\begin{aligned}\langle f_k, f_m \rangle_{L^2(\mu)} &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi/4a}} \mathrm{e}^{-2ax^2} \left(\frac{c}{a}\right)^{1/2} \mathrm{e}^{-2(c-a)x^2} \left(2^k k!\right)^{-1/2} \left(2^m m!\right)^{-1/2} H_k(\sqrt{2c}x) H_m(\sqrt{2c}x) dx \\ &= \sqrt{2c/\pi} \left(2^k k!\right)^{-1/2} \left(2^m m!\right)^{-1/2} \int_{\mathbb{R}} \mathrm{e}^{-2cx^2} H_k(\sqrt{2c}x) H_m(\sqrt{2c}x) dx \\ &= \delta_{mk},\end{aligned}$$

using properties of Hermite polynomials. Considering the integral operator $T : L^2(\mu) \to L^2(\mu)$, defined as $Tf(y) = \int_{\mathbb{R}} \mathrm{e}^{-b(x-y)^2} f(x) d\mu(x)$, we have:

$$\begin{aligned}Tf_k(y) &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \int_{\mathbb{R}} \mathrm{e}^{-(c-a)x^2} H_k(\sqrt{2c}x) \frac{1}{\sqrt{2\pi/4a}} \mathrm{e}^{-2ax^2} \mathrm{e}^{-b(x-y)^2} dx \\ &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \mathrm{e}^{-by^2} \frac{1}{\sqrt{2\pi/4a}} \frac{1}{\sqrt{2c}} \int_{\mathbb{R}} \mathrm{e}^{-(a+b+c)x^2} H_k(\sqrt{2c}x) \mathrm{e}^{2bxy} \sqrt{2c} dx \\ &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \mathrm{e}^{-by^2} \frac{1}{\sqrt{2\pi/4a}} \frac{1}{\sqrt{2c}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{a+b+c}{2c}x^2} H_k(x) \mathrm{e}^{\frac{2b}{\sqrt{2c}}xy} dx.\end{aligned}$$

We consider u such that $\frac{1}{1-u^2} = \frac{a+b+c}{2c}$, that is, $1 - \frac{2c}{a+b+c} = \frac{a+b-c}{a+b+c} = \frac{b^2}{(a+b+c)^2} = u^2$, which implies that $u = \frac{b}{a+b+c}$; and then $\frac{2u}{1-u^2} = \frac{b}{c}$.

Thus, using properties of Hermite polynomials (see Section D.4), we get:

$$\begin{aligned}Tf_k(y) &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \mathrm{e}^{-by^2} \frac{1}{\sqrt{2\pi/4a}} \frac{1}{\sqrt{2c}} \sqrt{\pi} \sqrt{1-u^2} H_k(\sqrt{2c}y) \exp\left(\frac{u^2}{1-u^2} 2cy^2\right) u^k \\ &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \frac{1}{\sqrt{2\pi/4a}} \frac{1}{\sqrt{2c}} \sqrt{\pi} \frac{\sqrt{2c}}{\sqrt{a+b+c}} H_k(\sqrt{2c}y) \exp(buy^2 - by^2) u^k \\ &= \left(\frac{c}{a}\right)^{1/4} \left(2^k k!\right)^{-1/2} \frac{\sqrt{2a}}{\sqrt{a+b+c}} H_k(\sqrt{2c}y) \exp\left(-by^2 + 2cy^2\left(-1 + \frac{1}{1-u^2}\right)\right) u^k \\ &= \frac{\sqrt{2a}}{\sqrt{a+b+c}} \left(\frac{b}{a+b+c}\right)^k f_k(y) \\ &= \lambda_k f_k(y).\end{aligned}$$

This implies that $(\tilde{f}_i)$ is an orthonormal basis of $\mathcal{H}$. □

We can now rewrite our problem in this basis, which is the purpose of the following lemma:

Lemma 9 (Reformulation of the problem in the basis)**.** *Let $(\alpha_i)_i \in \ell^2(\mathbb{N})$. For $f = \sum_{i=0}^{\infty} \alpha_i f_i$, we have:*

- $\|f\|_{\mathcal{H}}^2 = \sum_{i=0}^{\infty} \alpha_i^2 \lambda_i^{-1} = \alpha^\top \operatorname{Diag}(\lambda)^{-1} \alpha$*;*
- $\operatorname{var}_\mu(f(x)) = \sum_{i=0}^{\infty} \alpha_i^2 - \left(\sum_{i=0}^{\infty} \eta_i \alpha_i\right)^2 = \alpha^\top (I - \eta\eta^\top) \alpha$*;*
- $\mathbb{E}_\mu f'(x)^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \alpha_i \alpha_j (M^\top M)_{ij} = \alpha^\top M^\top M \alpha$*,*

where η is the vector of coefficients of $\mathbf{1}_{L^2(\mu)}$ and M the matrix of coordinates of the derivative operator in the (f_i) basis. The problem can be rewritten under the following form:

$$\mathcal{P}_\kappa^{-1} = \inf_\alpha \frac{\alpha^\top (M^\top M + \kappa \operatorname{Diag}(\lambda)^{-1}) \alpha}{\alpha^\top (I - \eta\eta^\top) \alpha}, \tag{15}$$

where

- $\forall k \geqslant 0, \eta_{2k} = \left(\frac{c}{a}\right)^{1/4} \sqrt{\frac{2a}{a+c}} \left(\frac{b}{a+b+c}\right)^k \frac{\sqrt{(2k)!}}{2^k k!}$ *and* $\eta_{2k+1} = 0$
- $\forall i \in \mathbb{N}, \left(M^\top M\right)_{ii} = \frac{1}{c}\left(2i(a^2+c^2) + (a-c)^2\right)$ *and* $\left(M^\top M\right)_{i,i+2} = \frac{1}{c}\left((a^2-c^2)\sqrt{(i+1)(i+2)}\right)$.

Proof. **Covariance operator.** Since (f_i) is orthonormal for $L^2(\mu)$, we only need to compute for each i, $\eta_i = \mathbb{E}_\mu f_i(x)$, as follows (and using properties of Hermite polynomials):

$$\begin{aligned}
\eta_i = \langle 1, f_i \rangle_{L^2(\mu)} &= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \int_{\mathbb{R}} \mathrm{e}^{-(c-a)x^2} H_i(\sqrt{2c}x) \mathrm{e}^{-2ax^2} \sqrt{2a/\pi} dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{a/(\pi c)} \int_{\mathbb{R}} \mathrm{e}^{-\frac{a+c}{2c}x^2} H_i(x) dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{\frac{2a}{a+c}} \left(\frac{c-a}{c+a}\right)^{i/2} H_i(0) \mathrm{i}^i.
\end{aligned}$$

This is only non-zero for i even, and

$$\begin{aligned}
\eta_{2k} &= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k}(2k)!\right)^{-1/2} \sqrt{\frac{2a}{a+c}} \left(\frac{c-a}{c+a}\right)^k H_{2k}(0)(-1)^k \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k}(2k)!\right)^{-1/2} \sqrt{\frac{2a}{a+c}} \left(\frac{c-a}{c+a}\right)^k \frac{(2k)!}{k!} \\
&= \left(\frac{c}{a}\right)^{1/4} \sqrt{\frac{2a}{a+c}} \left(\frac{c-a}{c+a}\right)^k \frac{\sqrt{(2k)!}}{2^k k!} \\
&= \left(\frac{c}{a}\right)^{1/4} \sqrt{\frac{2a}{a+c}} \left(\frac{b}{a+b+c}\right)^k \frac{\sqrt{(2k)!}}{2^k k!}.
\end{aligned}$$

Note that we must have $\sum_{i=0}^\infty \eta_i^2 = \|1\|^2_{L^2(\mu)} = 1$, which can indeed be checked —the shrewd reader will recognize the entire series development of $(1-z^2)^{-1/2}$.

Derivatives. We have, using the recurrence properties of Hermite polynomials:

$$f_i' = \frac{a-c}{\sqrt{c}} \sqrt{i+1} f_{i+1} + \frac{a+c}{\sqrt{c}} \sqrt{i} f_{i-1},$$

for $i > 0$, while for $i = 0$, $f_0' = \frac{a-c}{\sqrt{c}} f_1$. Thus, if M is the matrix of coordinates of the derivative operator in the basis (f_i), we have $M_{i+1,i} = \frac{a-c}{\sqrt{c}}\sqrt{i+1}$ and $M_{i-1,i} = \frac{a+c}{\sqrt{c}}\sqrt{i}$. This leads to

$$\langle f_i', f_j' \rangle_{L^2(\mu)} = (M^\top M)_{ij}.$$

We have

$$\begin{aligned}
(M^\top M)_{ii} &= \langle f_i', f_i' \rangle_{L^2(\mu)} \\
&= \frac{1}{c}\Big((i+1)(a-c)^2 + i(a+c)^2\Big) \\
&= \frac{1}{c}\Big(2i(a^2+c^2) + (a-c)^2\Big) \text{ for } i \geqslant 0, \\
(M^\top M)_{i,i+2} &= \langle f_i', f_{i+2}' \rangle_{L^2(\mu)} \\
&= \frac{1}{c}\Big((a^2-c^2)\sqrt{(i+1)(i+2)}\Big) \text{ for } i \geqslant 0.
\end{aligned}$$

Note that we have $M\eta = 0$ as these are the coordinates of the derivative of the constant function (this can be checked directly by computing $(M\eta)_{2k+1} = M_{2k+1,2k}\eta_{2k} + M_{2k+1,2k+2}\eta_{2k+2}$).

□

D.2 Unregularized solution

Recall that we want to solve $\mathcal{P}^{-1} = \inf_f \frac{\mathbb{E}_\mu f'(x)^2}{\mathrm{var}_\mu(f(x))}$,. The following lemma characterizes the optimal solution completely.

Lemma 10 (Optimal solution for one dimensional Gaussian)**.** *We know that the solution of the Poincaré problem is $\mathcal{P}^{-1} = 4a$ which is attained for $f_*(x) = x$. The decomposition of f_* is the basis $(f_i)_i$ is given by $f_* = \sum_{i \geqslant 0} \nu_i f_i$, where $\forall k \geqslant 0$, $\nu_{2k} = 0$ and $\nu_{2k+1} = \left(\frac{c}{a}\right)^{1/4} \frac{\sqrt{a}}{2c} \left(\frac{2c}{a+c}\right)^{3/2} \left(\frac{b}{a+b+c}\right)^k \frac{\sqrt{(2k+1)!}}{2^k k!}$.*

Proof. We thus need to compute:

$$
\begin{aligned}
\nu_i &= \langle f_*, f_i \rangle_{L^2(\mu)} \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \int_{\mathbb{R}} \mathrm{e}^{-(c-a)x^2} H_i(\sqrt{2c}x) \mathrm{e}^{-2ax^2} \sqrt{2a/\pi} x dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{2a/\pi} \int_{\mathbb{R}} \mathrm{e}^{-(c+a)x^2} H_i(\sqrt{2c}x) x dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{2a/\pi} \frac{1}{2c} \int_{\mathbb{R}} \mathrm{e}^{-\frac{c+a}{2c}x^2} H_i(x) x dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{2a/\pi} \frac{1}{4c} \int_{\mathbb{R}} \mathrm{e}^{-\frac{c+a}{2c}x^2} [H_{i+1}(x) + 2i H_{i-1}(x)] dx \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^i i!\right)^{-1/2} \sqrt{2a/\pi} \frac{\sqrt{\pi}}{4c} \sqrt{\frac{2c}{a+c}} \Big(\left(\frac{c-a}{c+a}\right)^{(i+1)/2} H_{i+1}(0) \mathrm{i}^{i+1} \\
&\qquad + 2i \left(\frac{c-a}{c+a}\right)^{(i-1)/2} H_{i-1}(0) \mathrm{i}^{i-1} \Big),
\end{aligned}
$$

which is only non-zero for i odd. We have:

$$
\begin{aligned}
\nu_{2k+1} &= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{2a/\pi} \frac{\sqrt{\pi}}{4c} \sqrt{\frac{2c}{a+c}} \Big(\left(\frac{c-a}{c+a}\right)^{k+1} H_{2k+2}(0)(-1)^{k+1} \\
&\qquad + 2(2k+1) \left(\frac{c-a}{c+a}\right)^k H_{2k}(0)(-1)^k \Big) \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{2a/\pi} \frac{\sqrt{\pi}}{4c} \sqrt{\frac{2c}{a+c}} \Big(\left(\frac{c-a}{c+a}\right)^{k+1} H_{2k+2}(0)(-1)^{k+1} \\
&\qquad + 2(2k+1) \left(\frac{c-a}{c+a}\right)^k H_{2k}(0)(-1)^k \Big) \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{2a/\pi} \frac{\sqrt{\pi}}{4c} \sqrt{\frac{2c}{a+c}} \left(\frac{c-a}{c+a}\right)^k (-1)^k \\
&\qquad \Big(\left(\frac{c-a}{c+a}\right) 2(2k+1) H_{2k}(0) + 2(2k+1) H_{2k}(0) \Big) \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{2a/\pi} \frac{\sqrt{\pi}}{4c} \sqrt{\frac{2c}{a+c}} \left(\frac{c-a}{c+a}\right)^k (-1)^k 2(2k+1) H_{2k}(0) \frac{2c}{c+a} \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{a} \frac{1}{c\sqrt{2}} \left(\frac{2c}{a+c}\right)^{3/2} \left(\frac{c-a}{c+a}\right)^k (-1)^k (2k+1) H_{2k}(0) \\
&= \left(\frac{c}{a}\right)^{1/4} \left(2^{2k+1}(2k+1)!\right)^{-1/2} \sqrt{a} \frac{1}{c\sqrt{2}} \left(\frac{2c}{a+c}\right)^{3/2} \left(\frac{c-a}{c+a}\right)^k (2k+1) \frac{(2k)!}{k!} \\
&= \left(\frac{c}{a}\right)^{1/4} \frac{\sqrt{a}}{2c} \left(\frac{2c}{a+c}\right)^{3/2} \left(\frac{c-a}{c+a}\right)^k \frac{\sqrt{(2k+1)!}}{2^k k!} \\
&= \left(\frac{c}{a}\right)^{1/4} \frac{\sqrt{a}}{2c} \left(\frac{2c}{a+c}\right)^{3/2} \left(\frac{b}{a+b+c}\right)^k \frac{\sqrt{(2k+1)!}}{2^k k!}.
\end{aligned}
$$

□

Note that we have:

$$\mu^\top \nu = \langle 1, f_* \rangle_{L^2(\mu)} = 0$$
$$\|\nu\|^2 = \|f_*\|^2_{L^2(\mu)} = \frac{1}{4a}$$
$$M^\top M \nu = 4a\nu.$$

The first equality if obvious from the odd/even sparsity patterns. The third one can be checked directly. The second one can probably be checked by another shrewd entire series development.

If we had $\nu^\top \mathrm{Diag}(\lambda)^{-1}\nu$ finite, then we would have

$$\mathcal{P}^{-1} \leqslant \mathcal{P}_\kappa^{-1} \leqslant \mathcal{P}^{-1}\left(1 + \kappa \cdot \nu^\top \mathrm{Diag}(\lambda)^{-1}\nu\right),$$

which would very nice and simple. Unfortunately, this is not true (see below).

D.2.1 Some further properties for ν

We have: $\frac{c-a}{c+a} = \frac{b}{a+b+c}$, and the following equivalent $\frac{\sqrt{\sqrt{k}(2k/e)^{2k+1}}}{2^k\sqrt{k}(k/e)^k} \sim \frac{k^{1/4+k+1/2}}{k^{k+1/2}} \sim k^{1/4}$ (up to constants). Thus

$$|\nu_{2k+1}^2 \lambda_{2k+1}^{-1}| \leqslant \left(\frac{c}{a}\right)^{1/2} \frac{a}{c^2} \left(\frac{2c}{a+c}\right)^3 \left(\frac{b}{a+b+c}\right)^{2k-2k-1} \sqrt{\frac{a+b+c}{2a}} \sqrt{k} = \Theta(\sqrt{k})$$

hence,

$$\sum_{k=0}^{2m+1} \nu_k^2 \lambda_k^{-1} \sim \Theta(m^{3/2}).$$

Consequently, $\nu^\top \mathrm{Diag}(\lambda)^{-1}\nu = +\infty$.

Note that we have the extra recursion

$$\nu_k = \frac{1}{\sqrt{4c}}\left[\sqrt{k+1}\eta_{k+1} + \sqrt{k}\eta_{k-1}\right].$$

D.3 Truncation

We are going to consider a truncated version α, of ν, with only the first $2m+1$ elements. That is $\alpha_k = \nu_k$ for $k \leqslant 2m+1$ and 0 otherwise.

Lemma 11 (Convergence of the truncation)**.** *Consider $g^m = \sum_{k=0}^\infty \alpha_k f_k = \sum_{k=0}^{2m+1} \nu_k f_k$, recall that $u = \frac{b}{a+b+c}$. For $m \geqslant max\{-\frac{3}{4\ln u}, \frac{1}{6c}\}$, we have the following:*

(i) $\left|\|\alpha\|^2 - \frac{1}{4a}\right| \leqslant Lmu^{2m}$

(ii) $\alpha^\top \eta = 0$

(iii) $\left|\alpha^\top M^\top M\alpha - 1\right| \leqslant Lm^2u^{2m}$

(iv) $\alpha^\top \mathrm{Diag}(\lambda)^{-1}\alpha \leqslant Lm^{3/2}$,

where L depends only on a, b, c.

Proof. We show successively the four estimations.

(i) Let us calculate $\|\alpha\|^2$. We have: $\|\alpha\|^2 - \frac{1}{4a} = \|\alpha\|^2 - \|\nu\|^2 = \sum_{k=m+1}^\infty \nu_{2k+1}^2$. Recall that $u = \frac{b}{a+b+c} \leqslant 1$, by noting $A = \left(\frac{c}{a}\right)^{1/4} \frac{\sqrt{a}}{2c}\left(\frac{2c}{a+c}\right)^{3/2}$, we have

$$\|\alpha\|^2 - \frac{1}{4a} = A^2 \sum_{k=m+1}^\infty \frac{(2k+1)!}{(2^k k!)^2} u^{2k}.$$

Now by Stirling inequality:

$$\begin{aligned}
\frac{(2k+1)!}{(2^k k!)^2} u^{2k} &\leqslant \frac{e\,(2k+1)^{2k+1+1/2}\,\mathrm{e}^{-(2k+1)}}{(\sqrt{2\pi}2^k k^{k+1/2}\mathrm{e}^{-k})^2} u^{2k} \\
&= \frac{\sqrt{2}}{\pi}\left(1+\frac{1}{2k}\right)^{2k+1}\left(k+\frac{1}{2}\right)^{1/2} u^{2k}. \\
&\leqslant \frac{4e}{\pi}\sqrt{k}u^{2k}.
\end{aligned}$$

And for $m \geqslant -\frac{1}{4\ln u}$,

$$\begin{aligned}
\sum_{m+1}^{\infty}\sqrt{k}u^{2k} &\leqslant \int_m^{\infty}\sqrt{x}u^{2x}dx \\
&\leqslant \int_m^{\infty} x u^{2x}dx \\
&= u^{2m}\frac{(1-2m\ln u)}{(2\ln u)^2} \\
&\leqslant \frac{mu^{2m}}{\ln(1/u)}.
\end{aligned}$$

Hence finally:

$$\left|\|\alpha\|^2 - \frac{1}{4a}\right| \leqslant \frac{4A^2 e}{\pi\ln(1/u)} m u^{2m}.$$

(ii) is straightforward because of the odd/even sparsity of ν and η.

(iii) Let us calculate $\|M\alpha\|^2$. We have:

$$\begin{aligned}
\|M\alpha\|^2 - 1 &= \|M\alpha\|^2 - \|M\nu\|^2 \\
&= \sum_{k,j\geqslant m+1}\nu_{2k+1}\nu_{2j+1}\left(M^\top M\right)_{2k+1,2j+1} \\
&= \sum_{k=m+1}^{\infty}\nu_{2k+1}^2\left(M^\top M\right)_{2k+1,2k+1} + 2\sum_{k=m+1}^{\infty}\nu_{2k+1}\nu_{2k+3}\left(M^\top M\right)_{2k+1,2k+3} \\
&= \frac{A^2}{c}\sum_{k=m+1}^{\infty}\frac{(2k+1)!}{(2^k k!)^2}\left(2(2k+1)(a^2+c^2)+(a-c)^2\right)u^{2k} \\
&\qquad - \frac{2A^2ab}{c}\sum_{k=m+1}^{\infty}\frac{\sqrt{(2k+1)!}}{(2^k k!)}\frac{\sqrt{(2k+3)!}}{(2^{k+1}(k+1)!)}\sqrt{(2k+2)(2k+3)}u^{2k+1}.
\end{aligned}$$

Let us call the two terms u_m and v_m respectively. For the first term, when $m \geqslant \max\{-\frac{3}{4\ln u}, \frac{1}{6c}\}$ a calculation as in **(i)** leads to:

$$\begin{aligned}
|u_m| &\leqslant \frac{24A^2e(u^2+c^2)}{\pi c}\int_m^{\infty} x\sqrt{x}u^{2x}dx + \frac{(a-c)^2}{c}\left(\|\alpha\|^2 - \|\nu\|^2\right)\\
&\leqslant \frac{24A^2e(u^2+c^2)}{\pi c}\int_m^{\infty} x^2u^{2x}dx - \frac{4A^2e}{\pi\ln u}mu^{2m}\\
&= -\frac{24A^2e(u^2+c^2)}{\pi c}\frac{u^{2m}(2m\ln u(2m\ln(u)-2)+2)}{8\ln^3(u)} - \frac{4A^2e}{\pi\ln u}mu^{2m}\\
&\leqslant -\frac{12A^2e(a^2+c^2)}{\pi c\ln(u)}m^2u^{2m} - \frac{4A^2e}{\pi\ln u}mu^{2m}\\
&\leqslant -\frac{4A^2e}{\pi\ln u}\left(\frac{3(a^2+c^2)}{c}m+1\right)mu^{2m}\\
&\leqslant \frac{24A^2ce}{\pi\ln(1/u)}m^2u^{2m}.
\end{aligned}$$

and for the second term, applying another time Stirling inequality, we get:

$$\begin{aligned}
\frac{\sqrt{(2k+1)!}}{2^kk!}\frac{\sqrt{(2k+3)!}}{2^{k+1}(k+1)!}u^{2k+1} &\leqslant \frac{\mathrm{e}^{1/2}\,(2k+1)^{k+3/4}\,\mathrm{e}^{-(k+1/2)}}{\sqrt{2\pi}2^kk^{k+1/2}\mathrm{e}^{-k}}\frac{\mathrm{e}^{1/2}\,(2k+3)^{k+7/4}\,\mathrm{e}^{-(k+3/2)}}{\sqrt{2\pi}2^{k+1}(k+1)^{k+3/2}\mathrm{e}^{-(k+1)}}u^{2k+1}\\
&\leqslant \frac{(2k+1)^{k+3/4}}{\sqrt{2\pi}2^kk^{k+1/2}}\frac{(2k+3)^{k+7/4}}{\sqrt{2\pi}2^{k+1}(k+1)^{k+3/2}}u^{2k+1}\\
&= \frac{\sqrt{2}}{\pi}\frac{\left(1+\frac{1}{2k}\right)^{k+3/4}\left(1+\frac{3}{2k}\right)^{k+7/4}}{\left(1+\frac{1}{k}\right)^{k+3/2}}\sqrt{k}u^{2k+1}\\
&\leqslant \frac{\sqrt{2}}{\pi}\frac{\left(1+\frac{3}{2k}\right)^{2k}\left(1+\frac{3}{2k}\right)^{5/2}}{\left(1+\frac{1}{k}\right)^{k}\left(1+\frac{1}{k}\right)^{3/2}}\sqrt{k}u^{2k+1}\\
&\leqslant \frac{\sqrt{2}}{\pi}\left(1+\frac{3}{2k}\right)^{2k}\left(1+\frac{3}{2k}\right)^{5/2}\sqrt{k}u^{2k+1}\\
&\leqslant \frac{15e^3}{\pi}\sqrt{k}u^{2k+1}.
\end{aligned}$$

Hence, as $\sum_{k\geqslant m+1}\sqrt{k}u^{2k+1} \leqslant -\frac{mu^{2m+1}}{\ln u}$, we have $|v_m| \leqslant \frac{30A^2abe^3}{\pi c\ln(1/u)}mu^{2m}$.

(iv) Let us calculate $\alpha^\top\,\mathrm{Diag}(\lambda)^{-1}\alpha$. We have:

$$\begin{aligned}
\alpha^\top \operatorname{Diag}(\lambda)^{-1}\alpha &= \sum_{k=0}^{m} \nu_{2k+1}^2 \lambda_{2k+1}^{-1} \\
&= A^2 \sqrt{\frac{bu}{2a}} \sum_{k=0}^{m} \frac{(2k+1)!}{(2^k k!)^2} u^{2k} u^{-(2k+1)} \\
&= A^2 \sqrt{\frac{b}{2au}} \sum_{k=0}^{m} \frac{(2k+1)!}{(2^k k!)^2} \\
&\leqslant \frac{4A^2 e\sqrt{b}}{\pi\sqrt{2au}} \sum_{k=0}^{m} \sqrt{k} \\
&\leqslant \frac{8A^2 e\sqrt{b}}{\pi\sqrt{2au}} m^{3/2}.
\end{aligned}$$

(Final constant.) By taking $L = \max\left\{ \frac{4A^2 e}{\pi \ln(1/u)}, \frac{48A^2 ce}{\pi \ln(1/u)}, \frac{60A^2 abe^3}{\pi c \ln(1/u)}, \frac{8A^2 e\sqrt{b}}{\pi\sqrt{2au}} \right\}$, we have proven the lemma. □

We can now state the principal result of this section:

Proposition 11 (Rate of convergence for the bias)**.** *If* $\kappa \leqslant \min\{a^2, 1/5, u^{1/(3c)}\}$ *and such that* $\ln(1/\kappa)\kappa \leqslant \frac{\ln(1/u)}{2aL}$, *then*

$$\mathcal{P}^{-1} \leqslant \mathcal{P}_\kappa^{-1} \leqslant \mathcal{P}^{-1}\left(1 + \frac{L}{2\ln^2(1/u)}\kappa \ln^2(1/\kappa)\right). \tag{16}$$

Proof. The first inequality $\mathcal{P}^{-1} \leqslant \mathcal{P}_\kappa^{-1}$ is obvious. On the other side,

$$\mathcal{P}_\kappa^{-1} = \inf_\beta \frac{\beta^\top (M^\top M + \kappa \operatorname{Diag}(\lambda)^{-1})\beta}{\beta^\top (I - \eta\eta^\top)\beta} \leqslant \frac{\alpha^\top (M^\top M + \kappa \operatorname{Diag}(\lambda)^{-1})\alpha}{\alpha^\top (I - \eta\eta^\top)\alpha},$$

With the estimates of Lemma 11, we have for $mu^{2m} < \frac{1}{4aL}$:

$$\begin{aligned}
\mathcal{P}_\kappa^{-1} &\leqslant \frac{1 + Lm^2 u^{2m} + \kappa L m^{3/2}}{\frac{1}{4a} - Lmu^{2m}} \\
&\leqslant \mathcal{P}^{-1}(1 + Lm^2u^{2m} + \kappa L m^{3/2}).
\end{aligned}$$

Let us take $m = \frac{\ln(1/\kappa)}{2\ln(1/u)}$.Then

$$\begin{aligned}
\mathcal{P}_\kappa^{-1} &\leqslant \mathcal{P}^{-1}(1 + \kappa L \frac{\ln^2(1/\kappa)}{4\ln^2(1/u)} + \kappa L \frac{\ln^{3/2}(1/\kappa)}{2^{3/2}\ln^{3/2}(1/u)}) \\
&\leqslant \mathcal{P}^{-1}\left(1 + \kappa L \frac{\ln^2(1/\kappa)}{2\ln^2(1/u)}\right),
\end{aligned}$$

as soon as $\kappa \leqslant a^2$. Note also that the condition $mu^{2m} < \frac{1}{4aL}$ can be rewritten in terms of m as $\kappa \ln(1/\kappa) < \frac{\ln(1/u)}{2aL}$. The other conditions of Lemma 11 are $\kappa \leqslant \mathrm{e}^{-3/2} \sim 0.22$ and $\kappa \leqslant u^{1/(3c)}$

□

D.4 Facts about Hermite polynomials

Orthogonality. We have:

$$\int_{\mathbb{R}} \mathrm{e}^{-x^2} H_k(x) H_m(x) = 2^k k! \sqrt{\pi}\delta_{km}.$$

Recurrence relations. We have:

$$H_i'(x) = 2iH_{i-1}(x),$$

and

$$H_{i+1}(x) = 2xH_i(x) - 2iH_{i-1}(x).$$

Mehler's formula. We have:

$$\sum_{k=0}^{\infty} \frac{H_k(x)\mathrm{e}^{-x^2/2} H_k(y)\mathrm{e}^{-y^2/2}}{2^k k! \sqrt{\pi}} u^k = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(\frac{2u}{1+u} xy - \frac{u^2}{1-u^2}(x-y)^2 - \frac{x^2}{2} - \frac{y^2}{2} \Big).$$

This implies that the functions $x \mapsto \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(\frac{2u}{1+u} xy - \frac{u^2}{1-u^2}(x-y)^2 - \frac{x^2}{2} - \frac{y^2}{2} \Big)$ has coefficients $\frac{H_k(y)\mathrm{e}^{-y^2/2}}{\sqrt{2^k k! \sqrt{\pi}}} u^k$ in the orthonormal basis $(x \mapsto \frac{H_k(x)\mathrm{e}^{-x^2/2}}{\sqrt{2^k k! \sqrt{\pi}}})$ of $L_2(dx)$.

Thus

$$\int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(\frac{2u}{1+u} xy - \frac{u^2}{1-u^2}(x-y)^2 - \frac{x^2}{2} - \frac{y^2}{2} \Big) \frac{H_k(x)\mathrm{e}^{-x^2/2}}{\sqrt{2^k k! \sqrt{\pi}}} dx = \frac{H_k(y)\mathrm{e}^{-y^2/2}}{\sqrt{2^k k! \sqrt{\pi}}} u^k,$$

that is

$$\int_{\mathbb{R}} \exp\Big(\frac{2u}{1+u} xy - \frac{u^2}{1-u^2}(x-y)^2 - x^2 \Big) H_k(x) dx = \sqrt{\pi}\sqrt{1-u^2} H_k(y) u^k.$$

This implies:

$$\int_{\mathbb{R}} \exp\Big(\frac{2u}{1-u^2} xy - \frac{x^2}{1-u^2} \Big) H_k(x) dx = \sqrt{\pi}\sqrt{1-u^2} H_k(y) \exp(\frac{u^2}{1-u^2} y^2) u^k$$

For $y = 0$, we get

$$\int_{\mathbb{R}} \exp\Big(- \frac{x^2}{1-u^2} \Big) H_k(x) dx = \sqrt{\pi}\sqrt{1-u^2} H_k(0) u^k.$$

Another consequence is that

$$\begin{aligned}
\sum_{k=0}^{\infty} \frac{H_k(x)H_k(y)}{2^k k! \sqrt{\pi}} u^k &= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(\frac{2u(1-u)+2u^2}{1-u^2} xy - \frac{u^2}{1-u^2}(x^2+y^2) \Big) \\
&= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(\frac{2u}{1-u^2} xy - \frac{u}{1-u^2}(x^2+y^2) + \frac{u}{1+u}(x^2+y^2) \Big) \\
&= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1-u^2}} \exp\Big(- \frac{u}{1-u^2}(x-y)^2 \Big) \exp\Big(\frac{u}{1+u}(x^2+y^2) \Big) \\
&= \frac{1}{\sqrt{\pi}} \frac{\sqrt{u}}{\sqrt{1-u^2}} \exp\Big(- \frac{u}{1-u^2}(x-y)^2 \Big) \frac{1}{\sqrt{u}} \exp\Big(\frac{u}{1+u}(x^2+y^2) \Big).
\end{aligned}$$

Thus, when u tends to 1, as a function of x, this tends to a Dirac at y times e^{y^2}.