
Supplementary material for *Kernel Interpolation With Continuous Volume Sampling*

Ayoub Belhadji¹ Rémi Bardenet¹ Pierre Chainais¹

This document contains the detailed proofs of the results in the main paper. Links across documents should work if the two PDFs are placed in the same repository. Light blue links in this document refer to the main paper, while dark blue links refer to the current supplementary material.

1. Technical results borrowed from other papers

Throughout our proof, we use a few technical results from the literature, which we gather here for ease of reference.

1.1. The Jacobi identity

The following proposition is a direct consequence of the rank one-update for determinants, see e.g. (Marcus et al., 2015, Theorem 3.11).

Proposition S1 (Jacobi identity). *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$. If $\text{Det } \mathbf{A} \neq 0$, then*

$$\partial_t \text{Det}(\mathbf{A} + t\mathbf{B})|_{t=0} = \text{Det}(\mathbf{A}) \text{Tr}(\mathbf{A}^{-1}\mathbf{B}). \quad (1)$$

In particular, we have

$$\partial_t \text{Det}(\mathbf{A} + t\mathbf{B})|_{t=0^+} = \text{Det}(\mathbf{A}) \text{Tr}(\mathbf{A}^{-1}\mathbf{B}). \quad (2)$$

1.2. The Markov brothers' inequality

The following proposition is known as the Markov brother's inequality, see e.g. (Shadrin, 2004).

Proposition S2 (Markov brothers). *Let P be a polynomial of degree smaller than N . Then*

$$\max_{\tau \in [-1, 1]} |P'(\tau)| \leq N^2 \max_{\tau \in [-1, 1]} |P(\tau)|. \quad (3)$$

We shall actually use a straightforward corollary.

Corollary S3. *Let P be a polynomial of degree smaller than N . Then*

$$\max_{\tau \in [0, 1]} |P'(\tau)| \leq 2N^2 \max_{\tau \in [0, 1]} |P(\tau)|. \quad (4)$$

Proof. Define the polynomial $Q(x) = P((x+1)/2)$, so that

$$Q'(x) = \frac{1}{2}P'((x+1)/2), \quad x \in [-1, 1]. \quad (5)$$

In particular,

$$\max_{\tau \in [0, 1]} |P(\tau)| = \max_{\tau \in [-1, 1]} |Q(\tau)|,$$

¹Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, Villeneuve d'Ascq, France. Correspondence to: Ayoub Belhadji <ayoub.belhadji@centralelille.fr>.

so that

$$\max_{\tau \in [0,1]} |P'(\tau)| = \max_{\tau \in [-1,1]} 2|Q'(\tau)| \leq 2N^2 \max_{\tau \in [-1,1]} |Q(\tau)| \leq 2N^2 \max_{\tau \in [0,1]} |P(\tau)|. \quad (6)$$

□

1.3. An inequality on the ratio of symmetric polynomials

Recall that, for $d \in \mathbb{N}^*$, \mathbb{R}^d is naturally embedded in the set of sequences $\mathbb{R}^{\mathbb{N}^*}$.

Now, let $M \in \mathbb{N}^*$, and let $\boldsymbol{\lambda} \in \mathbb{R}_+^{\mathbb{N}^*}$ such that $\sum_{m \in \mathbb{N}^*} \lambda_m < +\infty$. By MacLaurin's inequality¹, see e.g. (Steele, 2004, Chapter 12),

$$\forall M \in \mathbb{N}^*, \sum_{U \in \mathcal{U}_M} \prod_{u \in U} \lambda_u \leq \frac{1}{M!} \left(\sum_{m \in \mathbb{N}^*} \lambda_m \right)^M < +\infty. \quad (7)$$

In the following, we denote by $p_M(\boldsymbol{\lambda})$ the elementary symmetric polynomial of order M on the sequence $\boldsymbol{\lambda}$,

$$p_M(\boldsymbol{\lambda}) = \sum_{U \in \mathcal{U}_M} \prod_{u \in U} \lambda_u. \quad (8)$$

In particular, the following identity relates p_M and p_{M+1} .

$$\forall M \geq 2, \forall m \in \mathbb{N}^*, p_M(\boldsymbol{\lambda}) = \lambda_m p_{M-1}(\boldsymbol{\lambda}^{\overline{\{m\}}}) + p_M(\boldsymbol{\lambda}^{\overline{\{m\}}}), \quad (9)$$

where we denote, for $S \subset \mathbb{N}^*$, $\boldsymbol{\lambda}^{\overline{S}} = (\lambda_m^{\overline{S}})_{m \in \mathbb{N}^*} = (\lambda_m \mathbb{1}_{m \notin S})_{m \in \mathbb{N}^*}$. Proposition S4 further relates two consecutive elementary polynomials.

Proposition S4 (Theorem 3.1 of Guruswami & Sinop, 2012). *Let $M \in \mathbb{N}^*$ and $L \geq M+1$. Let $\boldsymbol{\lambda} \in \mathbb{R}_+^L$ be a nonincreasing sequence*

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L. \quad (10)$$

Assume that $\lambda_L > 0$, then

$$\forall M' \leq M, \frac{p_{M+1}(\boldsymbol{\lambda})}{p_M(\boldsymbol{\lambda})} \leq \frac{\sum_{m \geq M'+1} \lambda_m}{M+1-M'}. \quad (11)$$

We will actually use an immediate consequence of Proposition S4.

Corollary S5. *Let $M \in \mathbb{N}^*$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^{\mathbb{N}^*}$ be a nonincreasing sequence such that $\sum \lambda_m < +\infty$ and $\lambda_m > 0$ for all $m \in \mathbb{N}^*$. Then (11) still holds.*

Proof. Define, for $L \in \mathbb{N}^*$,

$$\boldsymbol{\lambda}_L = (\lambda_\ell)_{\ell \in [L]} \in \mathbb{R}_+^L. \quad (12)$$

By Proposition S4,

$$\forall M' \leq M, \forall L \geq M+1, \frac{p_{M+1}(\boldsymbol{\lambda}_L)}{p_M(\boldsymbol{\lambda}_L)} \leq \frac{1}{M+1-M'} \sum_{m=M'+1}^L \lambda_m \quad (13)$$

$$\leq \frac{1}{M+1-M'} \sum_{m=M'+1}^{+\infty} \lambda_m. \quad (14)$$

Letting $L \rightarrow \infty$ allows us to conclude. □

For the last result, recall the definition of the (cross-)leverage scores τ_{m_1, m_2} in (32). We slightly adapt a result by (Belhadji et al., 2019).

¹The inequality is usually stated for $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ for some $d \in \mathbb{N}^*$. Taking limits immediately yields (7).

Lemma S6. Let $\mathbf{x} \in \mathcal{X}^N$ satisfy $\text{Det } \mathbf{K}(\mathbf{x}) > 0$. For $m, m_1, m_2 \in \mathbb{N}^*$ such that $m_1 \neq m_2$,

$$\tau_m^{\mathcal{F}}(\mathbf{x}) = \|\Pi_{\mathcal{T}(\mathbf{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 = e_m^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x}), \quad (15)$$

and

$$\tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = \langle \Pi_{\mathcal{T}(\mathbf{x})} e_{m_1}^{\mathcal{F}}, \Pi_{\mathcal{T}(\mathbf{x})} e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} = e_{m_1}^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^{\mathcal{F}}(\mathbf{x}). \quad (16)$$

In particular,

$$\tau_m^{\mathcal{F}}(\mathbf{x}) \text{ and } |\tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x})| \text{ are in } [0, 1]. \quad (17)$$

Proof. The proof of (15) is given in (Belhadji et al., 2019)[Lemma 4 of Appendix D]. The proof of (16) is straightforward following the same lines. $\Pi_{\mathcal{T}(\mathbf{x})}$ is an orthogonal projection with respect to $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ and

$$\|e_m^{\mathcal{F}}\|_{\mathcal{F}} = \|e_{m_1}^{\mathcal{F}}\|_{\mathcal{F}} = \|e_{m_2}^{\mathcal{F}}\|_{\mathcal{F}} = 1, \quad (18)$$

so that (17) follows from the Cauchy-Schwarz inequality. \square

2. Proofs

2.1. Proof of Proposition 2

Proposition 2 states that continuous volume sampling is a mixture of projection determinantal point processes. We adapt a result in (Kulesza & Taskar, 2012, Chapter 5) for finite volume sampling to the infinite-dimensional case. The idea of the proof is to apply the Cauchy-Binet identity to a sequence of kernels of finite rank that approximate k .

First, recall from Section 1 the Mercer decomposition of k ,

$$k(x, y) = \lim_{M \rightarrow \infty} \sum_{m \in [M]} \sigma_m e_m(x) e_m(y) = \lim_{M \rightarrow \infty} k_M(x, y), \quad \forall x, y \in \mathcal{X}. \quad (19)$$

where kernel k_M has rank M .

Now, let $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$, and define $\mathbf{K}_M(\mathbf{x}) = (k_M(x_i, x_j))_{i, j \in [N]}$. By continuity of the determinant and by (19), it comes

$$\lim_{M \rightarrow \infty} \text{Det } \mathbf{K}_M(\mathbf{x}) = \text{Det } \mathbf{K}(\mathbf{x}). \quad (20)$$

By construction,

$$\mathbf{K}_M(\mathbf{x}) = \mathbf{F}_M(\mathbf{x})^{\top} \Sigma_M \mathbf{F}_M(\mathbf{x}), \quad (21)$$

where $\mathbf{F}_M(\mathbf{x}) = (e_m(x_i))_{(m, i) \in [M] \times [N]}$ and Σ_M is a diagonal matrix containing the first M eigenvalues $(\sigma_m)_{m \in [M]}$ on its diagonal. The Cauchy-Binet identity yields

$$\text{Det } \mathbf{K}_M(\mathbf{x}) = \sum_{\substack{U \subset [M] \\ |U|=N}} \text{Det}^2(e_u(x_i))_{(u, i) \in U \times [N]} \prod_{u \in U} \sigma_u. \quad (22)$$

Let now $\lambda_u = \prod_{i \in [N]} \sigma_u$ and $\mathbf{E}_U(\mathbf{x}) = (e_u(x_i))_{(u, i) \in U \times [N]}$, we combine (20) and (22) to obtain

$$\text{Det } \mathbf{K}(\mathbf{x}) = \lim_{M \rightarrow \infty} \sum_{\substack{U \subset [M] \\ |U|=N}} \lambda_u \text{Det}^2(e_u(x_i))_{(u, i) \in U \times [N]} \quad (23)$$

$$= \sum_{U \in \mathcal{U}_N} \lambda_u \text{Det}^2(e_u(x_i))_{(u, i) \in U \times [N]} \quad (24)$$

$$= \sum_{U \in \mathcal{U}_N} \lambda_u \text{Det}(\mathbf{E}_U(\mathbf{x})^{\top} \mathbf{E}_U(\mathbf{x})) \quad (25)$$

$$= \sum_{U \in \mathcal{U}_N} \lambda_u \text{Det}(\mathfrak{K}_U(x_i, x_j))_{i, j \in [N]}, \quad (26)$$

where $\mathfrak{K}_U(x, y) \triangleq \sum_{u \in U} e_u(x)e_u(y)$. Since \mathfrak{K}_U is a projection kernel, writing the determinant as a sum over permutations easily yields, for all $U \in \mathcal{U}_N$,

$$\int_{\mathcal{X}^N} \text{Det}(\mathfrak{K}_U(x_i, x_j))_{i,j \in [N]} \otimes_{i \in [N]} d\omega(x_i) = N!, \quad (27)$$

see e.g. Lemma 21 in (Hough et al., 2006). Finally, the monotone convergence theorem allows us to conclude

$$\int_{\mathcal{X}^N} \text{Det} \mathbf{K}(\mathbf{x}) \otimes_{i \in [N]} d\omega(x_i) = N! \sum_{\substack{U \subset \mathbb{N}^* \\ |U|=N}} \prod_{u \in U} \sigma_u. \quad (28)$$

2.2. Proof of Lemma 3

Lemma 3 gives an upper bound on the biggest weight δ_N in the mixture of Proposition 2. The proof is straightforward, as

$$\begin{aligned} r_N \prod_{\ell \in [N]} \sigma_\ell &= \sigma_N \sum_{m \geq N+1} \sigma_m \prod_{\ell \in [N-1]} \sigma_\ell \\ &\leq \sigma_N \sum_{\substack{U \subset \mathbb{N}^* \\ |U|=N}} \prod_{u \in U} \sigma_u. \end{aligned} \quad (29)$$

This immediately yields $\delta_N \leq \sigma_N/r_N$.

2.3. Proof of Lemma 8

Lemma 8 decomposes the interpolation error in terms of (cross-)leverage scores. Let $g \in \mathbb{L}_2(d\omega)$ satisfy $\|g\|_{d\omega} \leq 1$. Since $\Pi_{\mathcal{T}(\mathbf{x})}$ is an orthogonal projection with respect to $\langle \cdot, \cdot \rangle_{\mathcal{F}}$, we have

$$\|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 = \|\mu_g\|_{\mathcal{F}}^2 - \|\Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 \quad (30)$$

Now, $\mu_g = \Sigma g = \sum_{m \in \mathbb{N}^*} \sqrt{\sigma_m} g_m e_m^{\mathcal{F}}$, so that (30) becomes

$$\begin{aligned} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 &= \sum_{m \in \mathbb{N}^*} \sigma_m g_m^2 - \left\| \sum_{m \in \mathbb{N}^*} \Pi_{\mathcal{T}(\mathbf{x})} \sqrt{\sigma_m} g_m e_m^{\mathcal{F}} \right\|_{\mathcal{F}}^2 \\ &= \sum_{m \in \mathbb{N}^*} \sigma_m g_m^2 - \sum_{m_1, m_2} g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \langle \Pi_{\mathcal{T}(\mathbf{x})} e_{m_1}^{\mathcal{F}}, \Pi_{\mathcal{T}(\mathbf{x})} e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}}. \end{aligned} \quad (31)$$

Lemma S6 allows us to recognize leverage scores in (31). Taking out of the second sum in (31) the terms for which $m_1 = m_2$ to put them in the first sum concludes the proof of Lemma 8.

2.4. Proof of Theorem 4

The proof of (24) relies on the identity

$$\mathbb{E}_{\text{VS}} \|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} g_m^2 \epsilon_m(N), \quad (32)$$

and the fact that $(\epsilon_m(N))$ is a non-increasing sequence. We prove these two results in turn, after what we prove (26).

2.4.1. PROOF OF (32)

Let $\mathbf{x} \in \mathcal{X}^N$ such that $\text{Det} \mathbf{K}(\mathbf{x}) > 0$. Lemma 8 yields

$$\|\mu_g - \Pi_{\mathcal{T}(\mathbf{x})}\mu_g\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} g_m^2 \sigma_m \left(1 - \tau_m^{\mathcal{F}}(\mathbf{x})\right) - \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}). \quad (33)$$

First, we prove that

$$\mathbb{E}_{\text{VS}} \sum_{m \in \mathbb{N}^*} g_m^2 \sigma_m \left(1 - \tau_m^{\mathcal{F}}(\mathbf{x}) \right) = \sum_{m \in \mathbb{N}^*} g_m^2 \sigma_m \left(1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \right). \quad (34)$$

By Lemma S6,

$$\forall m \in \mathbb{N}^*, \quad g_m^2 \sigma_m \left(1 - \tau_m^{\mathcal{F}}(\mathbf{x}) \right) \geq 0, \quad (35)$$

so that (34) follows from the Beppo Levi's monotone convergence theorem.

Second, it remains to prove that

$$\mathbb{E}_{\text{VS}} \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = 0. \quad (36)$$

Again, Lemma S6 guarantees that, for $m_1, m_2 \in \mathbb{N}^*$ such that $m_1 \neq m_2$,

$$|g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x})| \leq |g_{m_1} g_{m_2}| \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}}. \quad (37)$$

Since

$$\begin{aligned} \sum_{m_1 \neq m_2 \in \mathbb{N}^*} |g_{m_1} g_{m_2}| \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} &\leq \left(\sum_{m \in \mathbb{N}^*} |g_m| \sqrt{\sigma_m} \right)^2 \\ &\leq \sum_{m \in \mathbb{N}^*} g_m^2 \sum_{m \in \mathbb{N}^*} \sigma_m \\ &< +\infty, \end{aligned} \quad (38)$$

the dominated convergence theorem yields

$$\mathbb{E}_{\text{VS}} \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} g_{m_1} g_{m_2} \sqrt{\sigma_{m_1}} \sqrt{\sigma_{m_2}} \mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}),$$

but this is equal to zero by Proposition 9.

2.4.2. PROOF THAT $(\epsilon_m(N))$ IS NONINCREASING

Let $m \in \mathbb{N}^*$. By definition,

$$\epsilon_m(N) = \sigma_m \frac{\sum_{U \in \mathcal{U}_N^m} \prod_{u \in U} \sigma_u}{\sum_{U \in \mathcal{U}_N} \prod_{u \in U} \sigma_u} = \sigma_m \frac{p_N(\overline{\sigma^{\{m\}}})}{p_N(\overline{\sigma})}, \quad (39)$$

where we use a notation introduced in Section 1.3. This leads to

$$\epsilon_m(N) = \sigma_m \frac{\sigma_{m+1} p_{N-1}(\overline{\sigma^{\{m, m+1\}}}) + p_N(\overline{\sigma^{\{m, m+1\}}})}{p_N(\overline{\sigma})}, \quad (40)$$

and, similarly,

$$\epsilon_{m+1}(N) = \sigma_{m+1} \frac{\sigma_m p_{N-1}(\overline{\sigma^{\{m, m+1\}}}) + p_N(\overline{\sigma^{\{m, m+1\}}})}{p_N(\overline{\sigma})}. \quad (41)$$

Taking the ratio, it comes

$$\frac{\epsilon_m(N)}{\epsilon_{m+1}(N)} = \frac{\sigma_m \left(\sigma_{m+1} p_{N-1}(\overline{\sigma^{\{m, m+1\}}}) + p_N(\overline{\sigma^{\{m, m+1\}}}) \right)}{\sigma_{m+1} \left(\sigma_m p_{N-1}(\overline{\sigma^{\{m, m+1\}}}) + p_N(\overline{\sigma^{\{m, m+1\}}}) \right)} \quad (42)$$

$$\begin{aligned} &= \frac{1 + \frac{1}{\sigma_{m+1}} \frac{p_N(\overline{\sigma^{\{m, m+1\}}})}{p_{N-1}(\overline{\sigma^{\{m, m+1\}}})}}{1 + \frac{1}{\sigma_m} \frac{p_N(\overline{\sigma^{\{m, m+1\}}})}{p_{N-1}(\overline{\sigma^{\{m, m+1\}}})}} \geq 1, \end{aligned} \quad (43)$$

because $1/\sigma_{m+1} \geq 1/\sigma_m$.

2.4.3. PROOF OF (26)

We have $\epsilon_1(N) = \epsilon_N(N)\epsilon_1(N)/\epsilon_N(N) \leq \sigma_N\epsilon_1(N)/\epsilon_N(N)$ since a simple counting argument yields $\epsilon_N(N) \leq \sigma_N$. Along the lines of Section 2.4.2,

$$\frac{\epsilon_1(N)}{\epsilon_N(N)} = \frac{1 + \frac{1}{\sigma_N} \frac{p_N(\overline{\sigma^{\{1,N\}}})}{p_{N-1}(\overline{\sigma^{\{1,N\}}})}}{1 + \frac{1}{\sigma_1} \frac{p_N(\overline{\sigma^{\{1,N\}}})}{p_{N-1}(\overline{\sigma^{\{1,N\}}})}} \leq 1 + \frac{1}{\sigma_N} \frac{p_N(\overline{\sigma^{\{1,N\}}})}{p_{N-1}(\overline{\sigma^{\{1,N\}}})}. \quad (44)$$

Now, $\overline{\sigma^{\{1,N\}}}$ is a sequence of positive real numbers and the Σ is trace-class. Then, by Corollary S5, for $M \in [N-1]$,

$$\frac{p_N(\overline{\sigma^{\{1,N\}}})}{p_{N-1}(\overline{\sigma^{\{1,N\}}})} \leq \frac{1}{N-M} \sum_{m \geq M} \sigma_{m+2} = \frac{1}{N+1-(M+1)} \sum_{m+1 \geq M+1} \sigma_{m+2}. \quad (45)$$

Taking $M' = M+1$ concludes the proof of (26).

2.5. Proof of Proposition 5

2.5.1. THE CASE OF A POLYNOMIALLY-DECREASING SPECTRUM

Assume that $\sigma_m = m^{-2s}$ with $s > 1/2$. Let $N \in \mathbb{N}^*$ and $M_N = \lceil N/c \rceil \in \{2, \dots, N\}$, with $c \in [1, N[$. We have

$$\min_{M \in [2:N]} \frac{\sum_{m \geq M} \sigma_{m+1}}{(N-M+1)\sigma_N} \leq \frac{\sum_{m \geq M_N} \sigma_{m+1}}{(N-M_N+1)\sigma_N} \quad (46)$$

$$\leq \frac{\sum_{m \geq \lceil N/c \rceil} \sigma_{m+1}}{(N - \lceil N/c \rceil + 1)\sigma_N} \quad (47)$$

$$\leq \frac{\sum_{m \geq \lceil N/c \rceil} \sigma_{m+1}}{(N - N/c + 1)\sigma_N} \quad (48)$$

$$\leq \frac{\sum_{m \geq \lceil N/c \rceil} (m+1)^{-2s}}{(N - N/c + 1)N^{-2s}}. \quad (49)$$

$$(50)$$

Now,

$$\forall m \in \mathbb{N}^*, (m+1)^{-2s} \leq \int_m^{m+1} t^{-2s} dt = \frac{1}{2s-1} (m^{1-2s} - (m+1)^{1-2s}), \quad (51)$$

so that

$$\sum_{m \geq \lceil N/c \rceil} (m+1)^{-2s} \leq \frac{1}{2s-1} \lceil N/c \rceil^{1-2s}. \quad (52)$$

Recall that $2s > 1$, so that

$$\frac{1}{2s-1} \lceil N/c \rceil^{1-2s} \leq \frac{1}{2s-1} (N/c)^{1-2s}, \quad (53)$$

and

$$\min_{M \in [2:N]} \frac{\sum_{m \geq M} \sigma_{m+1}}{(N-M+1)\sigma_N} \leq \frac{1}{2s-1} \frac{(N/c)^{1-2s}}{(N - N/c + 1)N^{-2s}} \quad (54)$$

$$\leq \frac{c^{2s}}{2s-1} \frac{N}{(cN - N + c)}. \quad (55)$$

Note that c is a free parameter that belongs to $[1, N]$ ² that we can optimize in the upper bound: $\frac{c^{2s}}{2s-1} \frac{N}{(cN-N+c)}$. For this purpose, denote

$$\phi_N(c) = \frac{c^{2s}}{2s-1} \frac{N}{(cN-N+c)}. \quad (56)$$

For every $N \in \mathbb{N}^*$, ϕ_N is differentiable in $]0, +\infty[$ and

$$\phi'_N(c) = \frac{N}{2s-1} \frac{c^{2s-1}}{(cN-N+c)^2} ((2s-1)(N+1)c - 2sN), \quad (57)$$

so that ϕ'_N vanishes in $c_N^* = \frac{2s}{2s-1} \frac{N}{N+1}$; it is negative in $]0, c_N^*[$ and positive in $]c_N^*, +\infty[$. We distinguish three cases:

If $c_N^* < 1$, $N < 2s-1$ and ϕ'_N is positive on $[1, N]$ so that ϕ_N increases in $[1, N]$ and we take $c = 1$ in (55):

$$\phi_N(1) = \frac{N}{2s-1} < 1. \quad (58)$$

If $c_N^* \in [1, N]$, c_N^* is the unique minimizer of ϕ_N in $[1, N]$ and we take $c = c_N^*$ in (55) so that:

$$\phi_N(c_N^*) = \left(\frac{2s}{2s-1} \right)^{2s} \left(\frac{N}{N+1} \right)^{2s} \quad (59)$$

$$\leq \left(\frac{2s}{2s-1} \right)^{2s} \quad (60)$$

$$\leq \left(1 + \frac{1}{2s-1} \right) \left(1 + \frac{1}{2s-1} \right)^{2s-1}. \quad (61)$$

Finally, if $c_N^* > N$, $N < \frac{1}{2s-1}$, ϕ_N is decreasing in $[1, N]$ and we take $c = N$ in (55) so that:

$$\phi_N(N) = \frac{N^{2s-1}}{2s-1} \leq \frac{1}{2s-1} \left(\frac{1}{2s-1} \right)^{2s-1} \quad (62)$$

$$\leq \left(1 + \frac{1}{2s-1} \right) \left(1 + \frac{1}{2s-1} \right)^{2s-1}. \quad (63)$$

In the three cases, β_N is upper bounded by $\left(1 + \frac{1}{2s-1} \right) \left(1 + \frac{1}{2s-1} \right)^{2s-1}$. The artificial two-factor form of (61) and (63) is there to make limits clearer. In particular, the RHS goes to e as $s \rightarrow \infty$.

2.5.2. THE CASE OF AN EXPONENTIALLY DECREASING SPECTRUM

Assume that $\sigma_m = \alpha^m$ with $\alpha \in [0, 1]$. Let $N \in \mathbb{N}^*$, and $M_N = N \in \{2, \dots, N\}$. We have

$$\min_{M \in [2: N]} \frac{\sum_{m \geq M} \sigma_{m+1}}{(N-M+1)\sigma_N} \leq \frac{\sum_{m \geq M_N} \sigma_{m+1}}{(N-M_N+1)\sigma_N} \quad (64)$$

$$\leq \frac{\sum_{m \geq N} \sigma_{m+1}}{\sigma_N} \quad (65)$$

$$\leq \frac{\sum_{m \geq N} \alpha^{m+1}}{\alpha^N} \quad (66)$$

$$\leq \alpha^{N+1} \frac{\sum_{m \geq 0} \alpha^m}{\alpha^N} \quad (67)$$

$$\leq \frac{\alpha}{1-\alpha}. \quad (68)$$

²The inequality in (55) is valid for $c = N$ by continuity.

2.6. Proof of Proposition 11

We start with deriving the spectrum³ of the trace-class, self-adjoint operator

$$\Sigma_t = \Sigma + te_m^{\mathcal{F}} \otimes e_m^{\mathcal{F}}, \quad (69)$$

where $e_m^{\mathcal{F}} \otimes e_m^{\mathcal{F}}$ is defined by

$$\forall g \in \mathbb{L}_2(d\omega), \quad e_m^{\mathcal{F}} \otimes e_m^{\mathcal{F}} g(\cdot) = e_m^{\mathcal{F}}(\cdot) \int_{\mathcal{X}} g(y) e_m^{\mathcal{F}}(y) d\omega(y). \quad (70)$$

The two operators Σ and $e_m^{\mathcal{F}} \otimes e_m^{\mathcal{F}}$ are co-diagonalizable in the basis $(e_m)_{m \in \mathbb{N}^*}$, thus their linear combination Σ_t diagonalizes in this basis too. In other words, for $u \in \mathbb{N}^*$, e_u is an eigenfunction of Σ_t and

$$\Sigma_t e_u = \Sigma e_u + te_m^{\mathcal{F}} \otimes e_m^{\mathcal{F}}(e_u) = (\sigma_u + t\delta_{u,m}\sigma_u)e_u. \quad (71)$$

Therefore, the set $\{\sigma_u(1 + t\delta_{u,m}), u \in \mathbb{N}^*\}$ is included in the spectrum of Σ_t . Since $(e_m)_{m \in \mathbb{N}^*}$ is an orthonormal basis of $\mathbb{L}_2(d\omega)$ and correspond to the eigenfunctions of Σ_t associated to the elements of $\{\sigma_u(1 + t\delta_{u,m}), u \in \mathbb{N}^*\}$, then the spectrum of Σ_t is exactly the set $\{\sigma_u(1 + t\delta_{u,m}), u \in \mathbb{N}^*\}$.⁴ We now turn to deriving the spectrum of the trace-class, self-adjoint operator Σ_t^+ ; the case of Σ_t^- follows the same lines and will be omitted for brevity. We will prove that there exists an orthonormal basis $(f_m)_{m \in \mathbb{N}^*}$ of $\mathbb{L}_2(d\omega)$ such that every f_m is an eigenfunction of Σ_t^+ . If $t = 0$, $\Sigma_t^+ = \Sigma$ and $(e_m)_{m \in \mathbb{N}^*}$ is already an orthonormal basis of $\mathbb{L}_2(d\omega)$. We assume in the following that $t > 0$.

Consider the operator Δ_t^+ defined on $\mathbb{L}_2(d\omega)$ by

$$\Delta_t^+ g(\cdot) = t \left(e_{m_1}^{\mathcal{F}}(\cdot) + e_{m_2}^{\mathcal{F}}(\cdot) \right) \int_{\mathcal{X}} g(y) \left(e_{m_1}^{\mathcal{F}}(y) + e_{m_2}^{\mathcal{F}}(y) \right) d\omega(y). \quad (72)$$

We can write $\Sigma_t^+ = \Sigma + \Delta_t^+$, but this time, if $t > 0$, Σ and Δ_t^+ do not commute. In particular, they are not co-diagonalizable, and a more detailed analysis is necessary. First, by construction of Δ_t^+ ,

$$\Delta_t^+ e_m = 0, \quad m \notin \{m_1, m_2\},$$

so that for any $m \notin \{m_1, m_2\}$, Σ_t^+ and Σ have e_m for eigenfunction, with the same eigenvalue σ_m . Observe that

$$\mathbb{L}_2(d\omega) = \text{Span}(e_{m_1}, e_{m_2}) \oplus \text{Span}(e_m)_{m \notin \{m_1, m_2\}}. \quad (73)$$

Therefore, the rest of the proof consists in completing $(e_m)_{m \notin \{m_1, m_2\}}$ into an orthonormal basis of $\mathbb{L}_2(d\omega)$, by finding two orthonormal eigenfunctions of Σ_t^+ in $\text{Span}(e_{m_1}, e_{m_2})$. Since we assumed in Section 1 that the eigenvalues of Σ are nonzero, we note that $\text{Span}(e_{m_1}, e_{m_2}) = \text{Span}(e_{m_1}^{\mathcal{F}}, e_{m_2}^{\mathcal{F}})$. Expressing the new eigenfunctions in terms of $e_{m_1}^{\mathcal{F}}$ and $e_{m_2}^{\mathcal{F}}$ will turn out to be more convenient.

First, note that

$$\Sigma_t^+ e_{m_1}^{\mathcal{F}}(\cdot) = \Sigma e_{m_1}^{\mathcal{F}}(\cdot) + t \int_{\mathcal{X}} (e_{m_1}^{\mathcal{F}}(\cdot) + e_{m_2}^{\mathcal{F}}(\cdot)) (e_{m_1}^{\mathcal{F}}(y) + e_{m_2}^{\mathcal{F}}(y)) e_{m_1}^{\mathcal{F}}(y) d\omega(y) \quad (74)$$

$$= \sigma_{m_1} e_{m_1}^{\mathcal{F}}(\cdot) + t\sigma_{m_1} (e_{m_1}^{\mathcal{F}}(\cdot) + e_{m_2}^{\mathcal{F}}(\cdot)) \quad (75)$$

$$= (1 + t)\sigma_{m_1} e_{m_1}^{\mathcal{F}} + t\sigma_{m_1} e_{m_2}^{\mathcal{F}}. \quad (76)$$

Similarly,

$$\Sigma_t^+ e_{m_2}^{\mathcal{F}}(\cdot) = t\sigma_{m_2} e_{m_1}^{\mathcal{F}} + (1 + t)\sigma_{m_2} e_{m_2}^{\mathcal{F}}. \quad (77)$$

³ All the integration operators in this article, and specifically in this section, are self-adjoint and compact. The spectrum of such operators is the union of $\{0\}$ (the essential spectrum) and the set of eigenvalues (Brezis, 2010)[Theorem 6.8]. Yet, the proof of the Mercer decomposition, only involves the set of eigenvalues (Steinwart & Scovel, 2012). For this reason, we use the term ‘‘spectrum’’ to refer to the set of eigenvalues.

⁴ Σ_t is self-adjoint, and has no zero eigenvalue by assumption. Thus, any new eigenfunction that is not in our basis needs to be orthogonal to all basis elements, and is thus zero.

Now, let $v = \lambda_1 e_{m_1}^{\mathcal{F}} + \lambda_2 e_{m_2}^{\mathcal{F}}$, so that, by (76) and (77),

$$\begin{aligned}\Sigma_t^+ v &= \lambda_1 \left((1+t)\sigma_{m_1} e_{m_1}^{\mathcal{F}} + t\sigma_{m_1} e_{m_2}^{\mathcal{F}} \right) + \lambda_2 \left((1+t)\sigma_{m_2} e_{m_2}^{\mathcal{F}} + t\sigma_{m_2} e_{m_1}^{\mathcal{F}} \right) \\ &= \left(\lambda_1(1+t)\sigma_{m_1} + \lambda_2 t\sigma_{m_2} \right) e_{m_1}^{\mathcal{F}} + \left(\lambda_2(1+t)\sigma_{m_2} + \lambda_1 t\sigma_{m_1} \right) e_{m_2}^{\mathcal{F}},\end{aligned}\quad (78)$$

Solving for eigenvalues, we look for $\mu \in \mathbb{R}$ such that $\Sigma_t^+ v = \mu v$, or equivalently

$$\begin{cases} (1+t)\sigma_{m_1}\lambda_1 + t\sigma_{m_2}\lambda_2 &= \mu\lambda_1, \\ t\sigma_{m_1}\lambda_1 + (1+t)\sigma_{m_2}\lambda_2 &= \mu\lambda_2. \end{cases}$$

This is just saying that μ should be an eigenvalue of the matrix

$$\begin{pmatrix} (1+t)\sigma_{m_1} & t\sigma_{m_2} \\ t\sigma_{m_1} & (1+t)\sigma_{m_2} \end{pmatrix}, \quad (79)$$

which yields two solutions,

$$\mu_1^+ = (1+t) \frac{\sigma_{m_1} + \sigma_{m_2}}{2} + \frac{1}{2} \sqrt{(1+t)^2(\sigma_{m_1} - \sigma_{m_2})^2 + 4\sigma_{m_1}\sigma_{m_2}t^2}, \quad (80)$$

and

$$\mu_2^+ = (1+t) \frac{\sigma_{m_1} + \sigma_{m_2}}{2} - \frac{1}{2} \sqrt{(1+t)^2(\sigma_{m_1} - \sigma_{m_2})^2 + 4\sigma_{m_1}\sigma_{m_2}t^2}. \quad (81)$$

These solutions are distinct since $t > 0$, and the corresponding normalized eigenfunctions v_1^+ and v_2^+ are orthogonal with respect to $\langle \cdot, \cdot \rangle_{d\omega}$ since Σ_t^+ is self-adjoint. Finally, we define the set of eigenfunctions of Σ_t^+ by the system $(e_m)_{m \notin \{m_1, m_2\}} \cup (v_1^+, v_2^+)$ that is an orthonormal basis of $\mathbb{L}_2(d\omega)$. Therefore, the spectrum of the compact operator Σ_t^+ is exactly the set

$$\{\sigma_m, m \notin \{m_1, m_2\}\} \cup \{\mu_1^+, \mu_2^+\}. \quad (82)$$

Along the same lines, one can show that the eigenvalues of Σ_t^- restricted to $\text{Span}(e_{m_1}^{\mathcal{F}}, e_{m_2}^{\mathcal{F}})$ satisfy

$$\lambda^2 - (1+t)(\sigma_{m_1} + \sigma_{m_2})\lambda - \sigma_{m_1}\sigma_{m_2}t^2 = 0. \quad (83)$$

For $t > 0$, this equation again admits two distinct solutions

$$\hat{\mu}_1^- = (1+t) \frac{\sigma_{m_1} + \sigma_{m_2}}{2} + \frac{1}{2} \sqrt{(1+t)^2(\sigma_{m_1} - \sigma_{m_2})^2 + 4\sigma_{m_1}\sigma_{m_2}t^2}, \quad (84)$$

and

$$\hat{\mu}_2^- = (1+t) \frac{\sigma_{m_1} + \sigma_{m_2}}{2} - \frac{1}{2} \sqrt{(1+t)^2(\sigma_{m_1} - \sigma_{m_2})^2 + 4\sigma_{m_1}\sigma_{m_2}t^2}. \quad (85)$$

so that the spectrum of Σ_t^- is exactly the set

$$\{\sigma_m, m \notin \{m_1, m_2\}\} \cup \{\mu_1^-, \mu_2^-\} = \{\sigma_m, m \notin \{m_1, m_2\}\} \cup \{\mu_1^+, \mu_2^+\}. \quad (86)$$

In other words, the two operators Σ_t^+ and Σ_t^- share the same eigenvalues.

2.7. Proof of Proposition 10

2.7.1. THE EXPECTED VALUE OF THE m -TH LEVERAGE SCORE

Let $m \in \mathbb{N}^*$. On the one hand, recall that $\tau_m^{\mathcal{F}}(\mathbf{x}) = e_m^{\mathcal{F}}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x})$, so that, by Definition 1,

$$\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) = \left(N! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \sigma_u \right)^{-1} \int_{\mathcal{X}^N} e_m^{\mathcal{F}}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x}) \text{Det } \mathbf{K}(\mathbf{x}) \otimes_{i \in [N]} d\omega(x_i). \quad (87)$$

We have

$$\begin{aligned}
 \text{Det } \mathbf{K}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x}) &= \text{Det } \mathbf{K}(\mathbf{x}) \text{Tr} (e_m^{\mathcal{F}}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x})) \\
 &= \text{Det } \mathbf{K}(\mathbf{x}) \text{Tr} (\mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top) \\
 &= \partial_t \text{Det}(\mathbf{K}(\mathbf{x}) + t e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top)|_{t=0^+}, \tag{88}
 \end{aligned}$$

where the last line follows from the Jacobi identity of Theorem S1.

On the other hand, for $t > 0$ and with the notation of Section 5.2, let $\mathbf{K}_t(\mathbf{x}) := (k_t(x_i, x_j))_{i,j \in [N]} = \mathbf{K}(\mathbf{x}) + t e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top$. Since

$$\int_{\mathcal{X}} k_t(x, x) d\omega(x) = \int_{\mathcal{X}} k(x, x) d\omega(x) + t \int_{\mathcal{X}} e_m^{\mathcal{F}}(x)^2 d\omega(x) = \sum_{n \in \mathbb{N}^*} \sigma_n + t \sigma_m < \infty, \tag{89}$$

Hadamard's inequality yields the integrability of $\psi(\cdot, t) : \mathbf{x} \mapsto \text{Det } \mathbf{K}_t(\mathbf{x})$. Finally, observe that⁵

$$\phi_m(t) := Z_N(k_t) = \int_{\mathcal{X}^N} \psi(\mathbf{x}, t) \otimes_{i \in [N]} d\omega(x_i). \tag{90}$$

If we prove that ϕ_m is right differentiable in zero, and that we can justify the interchange of the derivation and the integration operations, we will have equated the right derivative of ϕ_m in zero and (87) using (88); this will achieve proving the first equation in Proposition 10. To this purpose, we need to prove that $t \mapsto \psi(\mathbf{x}, t)$ is right differentiable at zero, it is locally dominated by an integrable function and its derivative is locally dominated by an integrable function. Now, observe that $t \mapsto \psi(\mathbf{x}, t)$ is a polynomial of degree smaller than N , so that it is differentiable, and Corollary S3 yields

$$\max_{\tau \in [0,1]} |\partial_t \psi(\mathbf{x}, \tau)| \leq 2N^2 \max_{\tau \in [0,1]} |\psi(\mathbf{x}, \tau)|. \tag{91}$$

In other words, to dominate $\tau \mapsto |\partial_t \psi(\mathbf{x}, \tau)|$ uniformly on $[0, 1]$, it is sufficient to dominate $\tau \mapsto |\psi(\mathbf{x}, \tau)|$ uniformly there. Now, let $\tau \in [0, 1]$, we have

$$\mathbf{K}_1(\mathbf{x}) - \mathbf{K}_\tau(\mathbf{x}) = \mathbf{K}(\mathbf{x}) + e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top - \mathbf{K}(\mathbf{x}) - \tau e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top \tag{92}$$

$$= (1 - \tau) e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top \in \mathcal{S}_N^+. \tag{93}$$

Thus

$$0 \preceq \mathbf{K}_\tau(\mathbf{x}) \preceq \mathbf{K}_1(\mathbf{x}) \tag{94}$$

in the Loewner order, so that for any $\tau \in [0, 1]$,

$$|\psi(\mathbf{x}, \tau)| = \psi(\mathbf{x}, \tau) = \text{Det } \mathbf{K}_\tau(\mathbf{x}) \leq \text{Det } \mathbf{K}_1(\mathbf{x}) = \psi(\mathbf{x}, 1). \tag{95}$$

We conclude by observing that $\mathbf{x} \mapsto \psi(\mathbf{x}, 1)$ is integrable on \mathcal{X}^N by Proposition 2, and the fact that

$$\int_{\mathcal{X}} k_1(x, x) d\omega(x) < +\infty. \tag{96}$$

2.7.2. THE EXPECTED VALUE OF CROSS-LEVERAGE SCORES

Let $m_1, m_2 \in \mathbb{N}^*$ such that $m_1 \neq m_2$. We have

$$\begin{aligned}
 \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) &= e_{m_1}^{\mathcal{F}}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^{\mathcal{F}}(\mathbf{x}) \\
 &= \frac{1}{4} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \mathbf{K}(\mathbf{x})^{-1} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \\
 &\quad - \frac{1}{4} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \mathbf{K}(\mathbf{x})^{-1} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top. \tag{97}
 \end{aligned}$$

⁵In the main paper, a mistake has crept into the definition of $\phi_m(t)$, $\phi_{m_1, m_2}^+(t)$ and $\phi_{m_1, m_2}^-(t)$. The correct definition of these quantities is the following: $\phi_m(t) = Z_N(k_t)$, $\phi_{m_1, m_2}^+(t) = Z_N(k_t^+)$, and $\phi_{m_1, m_2}^-(t) = Z_N(k_t^-)$.

Thus

$$\mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = \frac{1}{4Z_N(k)} \int_{\mathcal{X}^N} (\Psi^+(\mathbf{x}) - \Psi^-(\mathbf{x})) \otimes_{i \in [N]} d\omega(x_i), \quad (98)$$

where

$$\Psi^+(\mathbf{x}) = (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \mathbf{K}(\mathbf{x})^{-1} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x})) \text{Det } \mathbf{K}(\mathbf{x}), \quad (99)$$

and

$$\Psi^-(\mathbf{x}) = (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \mathbf{K}(\mathbf{x})^{-1} (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x})) \text{Det } \mathbf{K}(\mathbf{x}). \quad (100)$$

We proceed as in Section 2.7.1 and we use Proposition S1 to prove that

$$\begin{aligned} \Psi^+(\mathbf{x}) &= \partial_t \text{Det} \left(\mathbf{K}(\mathbf{x}) + t (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x})) (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \right) \Big|_{t=0^+} \\ &= \partial_t \text{Det} (\mathbf{K}_t^+(\mathbf{x})) \Big|_{t=0^+}. \end{aligned} \quad (101)$$

and

$$\begin{aligned} \Psi^-(\mathbf{x}) &= \partial_t \text{Det} \left(\mathbf{K}(\mathbf{x}) + t (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x})) (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \right) \Big|_{t=0^+} \\ &= \partial_t \text{Det} (\mathbf{K}_t^-(\mathbf{x})) \Big|_{t=0^+}. \end{aligned} \quad (102)$$

In order to prove that ϕ_{m_1, m_2}^+ and ϕ_{m_1, m_2}^- are right differentiable in zero along with the second equation in Proposition 10, one can follow the same steps as in the end of Section 2.7.1. In particular, the interchange of the derivation and the integration operations follows from the same arguments, upon noting that both $\int_{\mathcal{X}} k_t^+(x, x) d\omega(x)$ and $\int_{\mathcal{X}} k_t^-(x, x) d\omega(x)$ are finite.

2.8. Proof of Proposition 9

The proof is a straightforward computation now that we have Proposition 10 and Proposition 11.

2.8.1. THE EXPECTED VALUE OF THE m -TH LEVERAGE SCORE

Let $m \in \mathbb{N}^*$. We have by Proposition 10 and Proposition 2,

$$\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) = \frac{1}{N! \sum_{\substack{U \subset \mathbb{N}^* \\ |U|=N}} \prod_{u \in U} \sigma_u} \frac{\partial \phi_m}{\partial t} \Big|_{t=0^+}, \quad (103)$$

where

$$\phi_m(t) = \int_{\mathcal{X}^N} \text{Det} \left(\mathbf{K}(\mathbf{x}) + t e_m^{\mathcal{F}}(\mathbf{x}) e_m^{\mathcal{F}}(\mathbf{x})^\top \right) \otimes_{i=1}^N d\omega(x_i). \quad (104)$$

Now by Proposition 11 and Proposition 2⁶,

$$\phi_m(t) = N! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \tilde{\sigma}_u(t), \quad (105)$$

where for $u \in \mathbb{N}^*$, $\tilde{\sigma}_u(t) = \sigma_u + t \delta_{m, u} \sigma_u$. Therefore,

$$\phi_m(t) = N! \sum_{\substack{U \in \mathcal{U}_N \\ m \in U}} \prod_{u \in U} \tilde{\sigma}_u(t) + N! \sum_{\substack{U \in \mathcal{U}_N \\ m \notin U}} \prod_{u \in U} \tilde{\sigma}_u(t) \quad (106)$$

$$= N! \sigma_m (t+1) \sum_{\substack{U \in \mathcal{U}_{N-1} \\ m \notin U}} \prod_{u \in U} \sigma_u + N! \sum_{\substack{U \in \mathcal{U}_N \\ m \notin U}} \prod_{u \in U} \sigma_u. \quad (107)$$

Thus,

$$\frac{\partial \phi_m}{\partial t} \Big|_{t=0} = N! \sigma_m \sum_{\substack{U \in \mathcal{U}_{N-1} \\ m \notin U}} \prod_{u \in U} \sigma_u, \quad (108)$$

⁶In the main paper, a mistake has crept into the definition of $\phi_m(t)$, $\phi_{m_1, m_2}^+(t)$ and $\phi_{m_1, m_2}^-(t)$. The correct definition of these quantities is the following: $\phi_m(t) = Z_N(k_t)$, $\phi_{m_1, m_2}^+(t) = Z_N(k_t^+)$, and $\phi_{m_1, m_2}^-(t) = Z_N(k_t^-)$.

so that (103) becomes

$$\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) = \left(\mathcal{N}! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \sigma_u \right)^{-1} \mathcal{N}! \sigma_m \sum_{\substack{U \in \mathcal{U}_{N-1} \\ m \notin U}} \prod_{u \in U} \sigma_u, \quad (109)$$

which concludes the proof.

2.8.2. THE EXPECTED VALUE OF CROSS-LEVERAGE SCORES

Let $m_1, m_2 \in \mathbb{N}^*$ such that $m_1 \neq m_2$. We have by Proposition 10 and Proposition 2,

$$\mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = \frac{1}{4N! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \sigma_u} \left(\frac{\partial \phi_{m_1, m_2}^+}{\partial t} - \frac{\partial \phi_{m_1, m_2}^-}{\partial t} \right) \Big|_{t=0^+}, \quad (110)$$

where

$$\phi_{m_1, m_2}^+(t) = \int_{\mathcal{X}^N} \text{Det} \left(\mathbf{K}(\mathbf{x}) + t (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x})) (e_{m_1}^{\mathcal{F}}(\mathbf{x}) + e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \right) \otimes_{i=1}^N d\omega(x_i), \quad (111)$$

and

$$\phi_{m_1, m_2}^-(t) = \int_{\mathcal{X}^N} \text{Det} \left(\mathbf{K}(\mathbf{x}) + t (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x})) (e_{m_1}^{\mathcal{F}}(\mathbf{x}) - e_{m_2}^{\mathcal{F}}(\mathbf{x}))^\top \right) \otimes_{i=1}^N d\omega(x_i). \quad (112)$$

Now by Proposition 11, for $t \geq 0$,

$$\phi_{m_1, m_2}^+(t) = N! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \tilde{\sigma}_u^+(t) = N! \sum_{U \in \mathcal{U}_N} \prod_{u \in U} \tilde{\sigma}_u^-(t) = \phi_{m_1, m_2}^-(t). \quad (113)$$

Plugging this back into (110) yields $\mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = 0$.

2.9. Proof of Theorem 6

2.9.1. A DECOMPOSITION RESULT FOR THE ERROR

We start with a lemma.

Lemma S7. *Let $\mu \in \mathcal{F}$ such that $\|\mu\|_{\mathcal{F}} \leq 1$. Under Assumption B,*

$$\mathbb{E}_{\text{VS}} \mathcal{E}(\mu; x)^2 \leq (1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 + \sum_{m \geq N+1} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2. \quad (114)$$

Proof. Using the same arguments as in the proof of Lemma 8 in Section 2.3, it comes that, for $\mathbf{x} \in \mathcal{X}^N$ such that $\text{Det } \mathbf{K}(\mathbf{x}) > 0$,

$$\|\mu - \Pi_{\mathcal{T}(\mathbf{x})} \mu\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \tau_m^{\mathcal{F}}(\mathbf{x}) \right) - \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}). \quad (115)$$

We want to take expectations in both sides of (115). For the first term in the RHS, we prove, using the same arguments as for the proof of Theorem 4 in Section 2.4, that

$$\mathbb{E}_{\text{VS}} \sum_{m \in \mathbb{N}^*} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \tau_m^{\mathcal{F}}(\mathbf{x}) \right) = \sum_{m \in \mathbb{N}^*} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \right). \quad (116)$$

For the second term in the RHS of (115), we need to justify that

$$\begin{aligned} \mathbb{E}_{\text{VS}} \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) \\ = \sum_{\substack{m_1, m_2 \in \mathbb{N}^* \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = 0. \end{aligned} \quad (117)$$

This can be done using dominated convergence. Indeed, let $M \in \mathbb{N}^*$. We have

$$\begin{aligned} \mathbb{E}_{\text{VS}} \sum_{\substack{m_1, m_2 \in [M] \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) \\ = \sum_{\substack{m_1, m_2 \in [M] \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \mathbb{E}_{\text{VS}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) = 0. \end{aligned} \quad (118)$$

Moreover,

$$\begin{aligned} & \left| \sum_{\substack{m_1, m_2 \in [M] \\ m_1 \neq m_2}} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) \right| \\ &= \left| \sum_{m_1, m_2 \in [M]} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) - \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \tau_m^{\mathcal{F}}(\mathbf{x}) \right| \\ &\leq \left| \sum_{m_1, m_2 \in [M]} \langle \mu, e_{m_1}^{\mathcal{F}} \rangle_{\mathcal{F}} \langle \mu, e_{m_2}^{\mathcal{F}} \rangle_{\mathcal{F}} \tau_{m_1, m_2}^{\mathcal{F}}(\mathbf{x}) \right| + \left| \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \tau_m^{\mathcal{F}}(\mathbf{x}) \right| \\ &= \left\| \Pi_{\mathcal{T}(\mathbf{x})} \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m^{\mathcal{F}} \right\|_{\mathcal{F}}^2 + \left| \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \tau_m^{\mathcal{F}}(\mathbf{x}) \right| \\ &\leq \left\| \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m^{\mathcal{F}} \right\|_{\mathcal{F}}^2 + \sum_{m \in [M]} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \\ &= 2 \|\mu\|_{\mathcal{F}}^2 < +\infty. \end{aligned} \quad (119)$$

Combining (118) and (119), we deduce (117) by the dominated convergence theorem.

Finally, we combine (116) and (117) to get

$$\begin{aligned} \mathbb{E}_{\text{VS}} \|\mu - \Pi_{\mathcal{T}(\mathbf{x})} \mu\|_{\mathcal{F}}^2 &= \sum_{m \in \mathbb{N}^*} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \right) \\ &= \sum_{n \in [N]} \langle \mu, e_n^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \mathbb{E}_{\text{VS}} \tau_n^{\mathcal{F}}(\mathbf{x}) \right) + \sum_{m \geq N+1} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \left(1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \right). \end{aligned} \quad (120)$$

On the one hand,

$$\forall m \geq N+1, \quad 1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \leq 1, \quad (121)$$

and on the other hand, remember that by Theorem 4, the sequence $\epsilon_m(N)$ is non-increasing, so that

$$\forall n \in [N], \quad \sigma_n (1 - \mathbb{E}_{\text{VS}} \tau_n^{\mathcal{F}}(\mathbf{x})) = \mathbb{E}_{\text{VS}} \|\mu_{e_n} - \Pi_{\mathcal{T}(\mathbf{x})} \mu_{e_n}\|_{\mathcal{F}}^2 \quad (122)$$

$$= \epsilon_n(N) \quad (123)$$

$$\leq \epsilon_1(N), \quad (124)$$

and by (26) in the same theorem one gets

$$\sigma_n (1 - \mathbb{E}_{\text{VS}} \tau_n^{\mathcal{F}}(\mathbf{x})) \leq (1 + \beta_N) \sigma_N, \quad (125)$$

so that

$$(1 - \mathbb{E}_{\text{VS}} \tau_n^{\mathcal{F}}(\mathbf{x})) \leq (1 + \beta_N) \frac{\sigma_N}{\sigma_n}. \quad (126)$$

Assumption B yields

$$\forall n \in [N], \quad 1 - \mathbb{E}_{\text{VS}} \tau_n^{\mathcal{F}}(\mathbf{x}) \leq (1 + B) \frac{\sigma_N}{\sigma_n}. \quad (127)$$

This concludes the proof of the lemma. \square

2.9.2. THE EXPECTED VALUE OF THE INTERPOLATION ERROR

If there exists $r \in [0, 1/2]$ such that $\|\Sigma^{-r}\mu\|_{\mathcal{F}} < +\infty$, we have

$$\sum_{m \geq N+1} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 = \sum_{m \geq N+1} \sigma_m^{2r} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (128)$$

$$\leq \sigma_{N+1}^{2r} \sum_{m \geq N+1} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (129)$$

$$\leq \sigma_{N+1}^{2r} \|\Sigma^{-r}\mu\|_{\mathcal{F}}^2, \quad (130)$$

and

$$(1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 = (1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m^{1-2r+2r}} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \quad (131)$$

$$= (1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m^{1-2r}} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (132)$$

$$\leq (1+B) \sigma_N^{2r} \sum_{m \in [N]} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (133)$$

$$= (1+B) \sigma_N^{2r} \|\Sigma^{-r}\mu\|_{\mathcal{F}}^2. \quad (134)$$

By Lemma S7, $\mathbb{E}_{\text{VS}} \|\mu - \Pi_{\mathcal{T}(x)}\mu\|_{\mathcal{F}}^2$ converges at the slow rate $\mathcal{O}(\sigma_N^{2r})$.

On the other hand, if there exists $r > 1/2$ such that $\|\Sigma^{-r}\mu\|_{\mathcal{F}} < +\infty$, we have

$$(1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 = (1+B) \sum_{m \in [N]} \frac{\sigma_N}{\sigma_m^{1-2r+2r}} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 \quad (135)$$

$$\leq (1+B) \sigma_N \sigma_1^{2r-1} \sum_{m \in [N]} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (136)$$

$$\leq (1+B) \sigma_N \sigma_1^{2r-1} \|\Sigma^{-r}\mu\|_{\mathcal{F}}^2, \quad (137)$$

and

$$\sum_{m \geq N+1} \langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2 = \sum_{m \geq N+1} \sigma_m^{2r} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (138)$$

$$\leq \sigma_{N+1}^{2r} \sum_{m \geq N+1} \frac{\langle \mu, e_m^{\mathcal{F}} \rangle_{\mathcal{F}}^2}{\sigma_m^{2r}} \quad (139)$$

$$\leq \sigma_{N+1}^{2r} \|\Sigma^{-r}\mu\|_{\mathcal{F}}^2. \quad (140)$$

This time, the bound in Lemma S7 is dominated by its first term, so that $\mathbb{E}_{\text{VS}} \|\mu - \Pi_{\mathcal{T}(x)}\mu\|_{\mathcal{F}}^2$ converges at the faster rate $\mathcal{O}(\sigma_N)$.

2.10. Proof of Theorem 7

2.10.1. PROOF OF THE BIAS IDENTITY

First, recall that, as f and g belong to $\mathbb{L}_2(d\omega)$, we have

$$\int_{\mathcal{X}} f(x)g(x) d\omega(x) = \sum_{m \in \mathbb{N}^*} \langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega}, \quad (141)$$

thus, in order to prove the result, it is enough to prove that

$$\mathbb{E}_{\text{VS}} \sum_{i \in [N]} \hat{w}_i f(x_i) = \sum_{m \in \mathbb{N}^*} \langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega} \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}). \quad (142)$$

Let $\mathbf{x} \in \mathcal{X}^N$ such that $\text{Det } \mathbf{K}(\mathbf{x}) > 0$. The optimal kernel quadrature weights satisfy

$$\hat{\mathbf{w}} = \mathbf{K}(\mathbf{x})^{-1} \mu_g(\mathbf{x}), \quad (143)$$

so that

$$\sum_{i \in N} \hat{w}_i f(x_i) = \hat{\mathbf{w}}^\top f(\mathbf{x}) \quad (144)$$

$$= \mu_g(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} f(\mathbf{x}) \quad (145)$$

$$= \sum_{m_1, m_2 \in \mathbb{N}^*} \sigma_{m_1} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_1}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^\mathcal{F}(\mathbf{x}) \quad (146)$$

$$= \sum_{m_1, m_2 \in \mathbb{N}^*} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_1}^\mathcal{F}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^\mathcal{F}(\mathbf{x}). \quad (147)$$

We want to use the dominated convergence theorem to take expectations in (147). Let $M \in \mathbb{N}^*$. By Lemma S6 and by the fact that $\Pi_{\mathcal{T}(\mathbf{x})}$ is an $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ -orthogonal projection, it comes

$$\left| \sum_{m_1, m_2 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_1}^\mathcal{F}(\mathbf{x})^\top \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^\mathcal{F}(\mathbf{x}) \right| \quad (148)$$

$$= \left| \sum_{m_1, m_2 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} \langle \Pi_{\mathcal{T}(\mathbf{x})} e_{m_1}^\mathcal{F}, \Pi_{\mathcal{T}(\mathbf{x})} e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} \right| \quad (149)$$

$$= \left| \left\langle \Pi_{\mathcal{T}(\mathbf{x})} \sum_{m_1 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1}, \Pi_{\mathcal{T}(\mathbf{x})} \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\rangle_{\mathcal{F}} \right| \quad (150)$$

$$\leq \left| \left\langle \sum_{m_1 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1}, \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\rangle_{\mathcal{F}} \right| \quad (151)$$

$$\leq \left\| \sum_{m_1 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1} \right\|_{\mathcal{F}} \left\| \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\|_{\mathcal{F}}. \quad (152)$$

Now,

$$\left\| \sum_{m_1 \in [M]} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1} \right\|_{\mathcal{F}} \left\| \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\|_{\mathcal{F}} \quad (153)$$

$$= \left\| \sum_{m_1 \in [M]} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1}^\mathcal{F} \right\|_{\mathcal{F}} \left\| \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\|_{\mathcal{F}} \quad (154)$$

$$= \left\| \sum_{m_1 \in [M]} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1} \right\|_{d\omega} \left\| \sum_{m_2 \in [M]} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\|_{\mathcal{F}} \quad (155)$$

$$\leq \left\| \sum_{m_1 \in \mathbb{N}^*} \langle g, e_{m_1} \rangle_{d\omega} e_{m_1} \right\|_{d\omega} \left\| \sum_{m_2 \in \mathbb{N}^*} \langle f, e_{m_2}^\mathcal{F} \rangle_{\mathcal{F}} e_{m_2}^\mathcal{F} \right\|_{\mathcal{F}} \quad (156)$$

$$< +\infty, \quad (157)$$

since $\sum_{m \in \mathbb{N}^*} \sqrt{\sigma_m} \langle g, e_m \rangle_{d\omega} e_m \in \mathcal{F}$. Dominated convergence thus yields

$$\mathbb{E}_{\text{VS}} \sum_{m_1, m_2 \in \mathbb{N}^*} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2} \rangle_{\mathcal{F}} e_{m_1}^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^{\mathcal{F}}(\mathbf{x}) \quad (158)$$

$$= \sum_{m_1, m_2 \in \mathbb{N}^*} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2} \rangle_{\mathcal{F}} \mathbb{E}_{\text{VS}} e_{m_1}^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^{\mathcal{F}}(\mathbf{x}). \quad (159)$$

Using Proposition 9, we continue our derivation as

$$\mathbb{E}_{\text{VS}} \sum_{m_1, m_2 \in \mathbb{N}^*} \sqrt{\sigma_{m_1}} \langle g, e_{m_1} \rangle_{d\omega} \langle f, e_{m_2} \rangle_{\mathcal{F}} e_{m_1}^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_{m_2}^{\mathcal{F}}(\mathbf{x}) \quad (160)$$

$$= \sum_{m \in \mathbb{N}^*} \sqrt{\sigma_m} \langle g, e_m \rangle_{d\omega} \langle f, e_m \rangle_{\mathcal{F}} \mathbb{E}_{\text{VS}} e_m^{\mathcal{F}}(\mathbf{x})^{\top} \mathbf{K}(\mathbf{x})^{-1} e_m^{\mathcal{F}}(\mathbf{x}) \quad (161)$$

$$= \sum_{m \in \mathbb{N}^*} \langle g, e_m \rangle_{d\omega} \sqrt{\sigma_m} \langle f, e_m \rangle_{\mathcal{F}} \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}). \quad (162)$$

Finally, (142) is obtained upon noting that

$$\forall m \in \mathbb{N}^*, \langle f, e_m \rangle_{d\omega} = \sqrt{\sigma_m} \langle f, e_m \rangle_{\mathcal{F}}. \quad (163)$$

2.10.2. PROOF OF THE ASYMPTOTIC UNBIASEDNESS OF THE QUADRATURE

The expected value of the bias writes

$$\mathbb{E}_{\text{VS}} \left(\int_{\mathcal{X}} f(x)g(x)d\omega(x) - \sum_{i \in [N]} \hat{w}_i f(x_i) \right) = \sum_{m \in \mathbb{N}^*} \langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega} (1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})). \quad (164)$$

Now, by Theorem 4, for $m \in \mathbb{N}^*$,

$$\mathbb{E}_{\text{VS}} \|\mu_{e_m} - \Pi_{\mathcal{T}(\mathbf{x})} \mu_{e_m}\|_{\mathcal{F}}^2 \leq \epsilon_1(N) \leq \sigma_N (1 + \beta_N) \leq \sigma_N + \sum_{n \geq N} \sigma_n. \quad (165)$$

Thus

$$0 \leq 1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) = \sigma_m^{-1} \mathbb{E}_{\text{VS}} \|\mu_{e_m} - \Pi_{\mathcal{T}(\mathbf{x})} \mu_{e_m}\|_{\mathcal{F}}^2 \leq \sigma_m^{-1} \sigma_N + \sum_{n \geq N} \sigma_n, \quad (166)$$

so that

$$\lim_{N \rightarrow \infty} \langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega} (1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})) = \langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega} (1 - \lim_{N \rightarrow \infty} \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})) = 0. \quad (167)$$

To conclude, it is thus enough to apply the dominated convergence theorem to (164). By Lemma S6, $\tau_m^{\mathcal{F}}(\mathbf{x}) \in [0, 1]$, so that $1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}) \in [0, 1]$. In particular, for all $N \in \mathbb{N}^*$,

$$|\langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega} (1 - \mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x}))| \leq |\langle f, e_m \rangle_{d\omega} \langle g, e_m \rangle_{d\omega}| \quad (168)$$

$$\leq \frac{1}{2} \left(\langle f, e_m \rangle_{d\omega}^2 + \langle g, e_m \rangle_{d\omega}^2 \right), \quad (169)$$

which is the generic term of a convergent series as $f, g \in \mathbb{L}_2(d\omega)$. This concludes the proof.

3. More concrete examples of RKHSs

In this section, we illustrate the bound of Theorem 4 and the constants of Proposition 5 on more examples.

3.1. The uni-dimensional periodic Sobolev spaces

Consider the uni-dimensional periodic Sobolev space of smoothness parameter $s \in \{1, 2, 3, 4, 5\}$. The eigenvalues have a polynomial decay; see (Wahba, 1990). We take for $m \in \mathbb{N}^*$, $\sigma_m = m^{-2s}$ ⁷. For different values of m , Figure 3.1 illustrates the expected value of the m -th leverage score $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ (left panels) and the expected interpolation error $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ (right panels), both as functions of N . Remember that by Theorem 4:

$$\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2 = \sigma_m \left(\sum_{U \in \mathcal{U}_N} \prod_{u \in U} \sigma_u \right)^{-1} \sum_{U \in \mathcal{U}_N^m} \prod_{u \in U} \sigma_u. \quad (170)$$

For numerical simulations, we make the following approximation

$$\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2 \approx \sigma_m \left(\sum_{\substack{U \subset [M] \\ |U|=N}} \prod_{u \in U} \sigma_u \right)^{-1} \sum_{\substack{U \subset [M] \\ |U|=N, m \notin U}} \prod_{u \in U} \sigma_u, \quad (171)$$

for an $M \geq N$ sufficiently large. The numerator and denominator of the right hand side of (171) can be calculated using an efficient algorithm for the calculation of the elementary symmetric polynomials (Kulesza & Taskar, 2012)[Algorithm 7].

We observe that for low values of s , $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ depends smoothly on N . On the other hand, $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ undergoes a sharp transition at $N = m$ for high values of s : the reconstruction of the m -th eigenfunction is almost perfect for N slightly larger than m . Moreover, $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ respects the upper bound of Theorem 4; the constant B of Proposition 5 is small for high values of s and converges to e when $s \rightarrow +\infty$.

3.2. The uni-dimensional Gaussian spaces

Consider now the RKHS generated by Gaussian kernel and the Gaussian measure. We take for all $m \in \mathbb{N}^*$, $\sigma_m = \alpha^N$ (Zhu et al., 1997), for some $\alpha \in [0, 1[$. Figure 3.2 illustrates the expected value of the m -th leverage score $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ (left panels) and the expected interpolation error $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ (right panels), both as functions of N , for different values of m and $\alpha \in \{0.7, 0.5, 0.2\}$. The numerical simulation of $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ uses again the approximation (171).

We make the same observations on the dependency of $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ on N as in the Sobolev case. The rougher the kernel (i.e., the lower the value of α), the smoother the transition of $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ as a function of N . Moreover, $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ respects the upper bound of Theorem 4; the constant B of Proposition 5 is small for low values of α and converges to 0 when $\alpha \rightarrow 0$.

⁷We drop the potential multiplicities of the eigenvalues by simplicity.

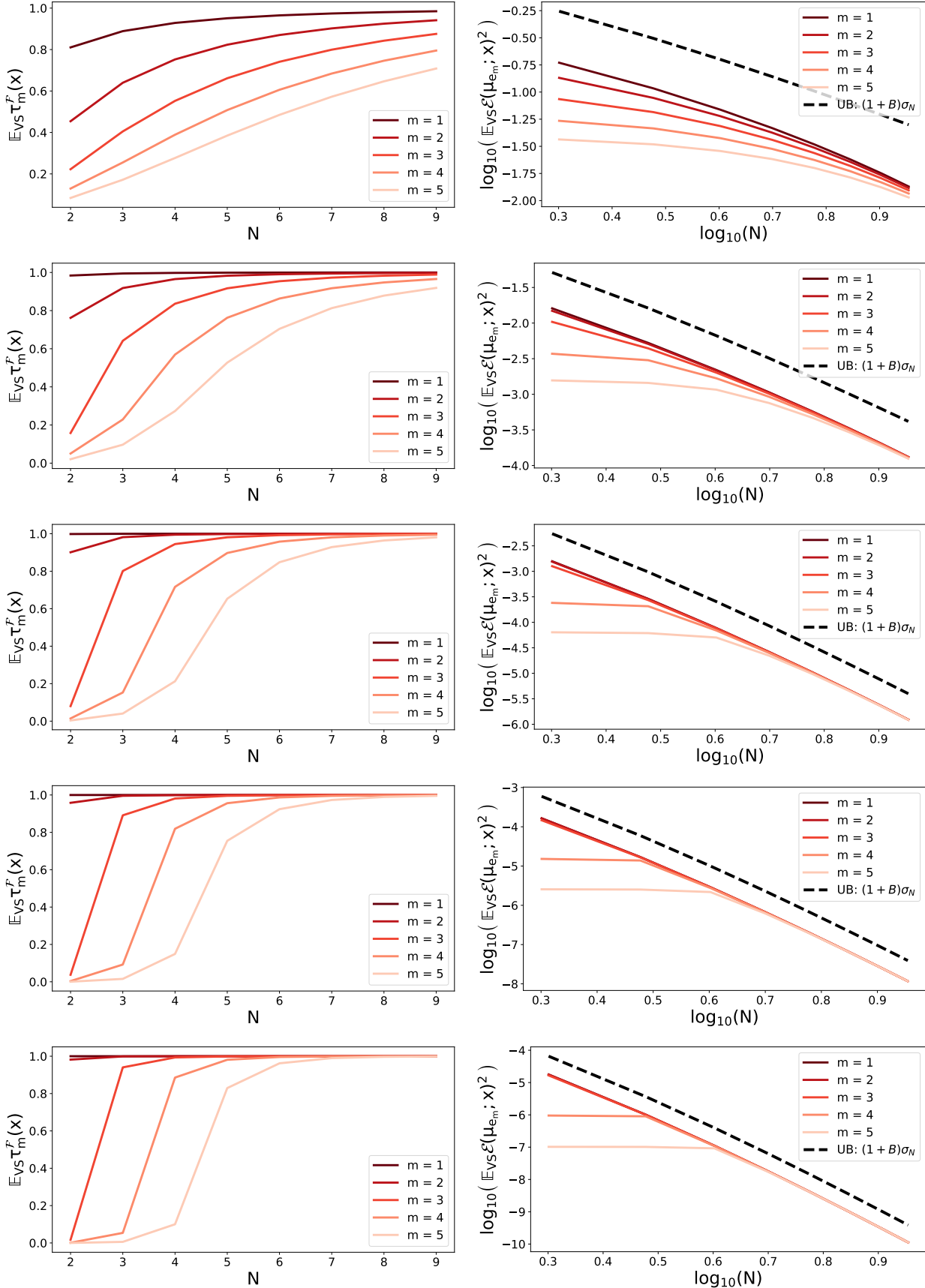


Figure 1. The expected value of the m -th leverage score $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ (left panels) and the expected interpolation error $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ (right panels), under the distribution of continuous volume sampling, for $m \in \{1, 2, 3, 4, 5\}$ and the uni-variate periodic Sobolev kernel. Rows correspond to increasing values of the smoothness parameter $s = 1, 2, 3, 4, 5$.

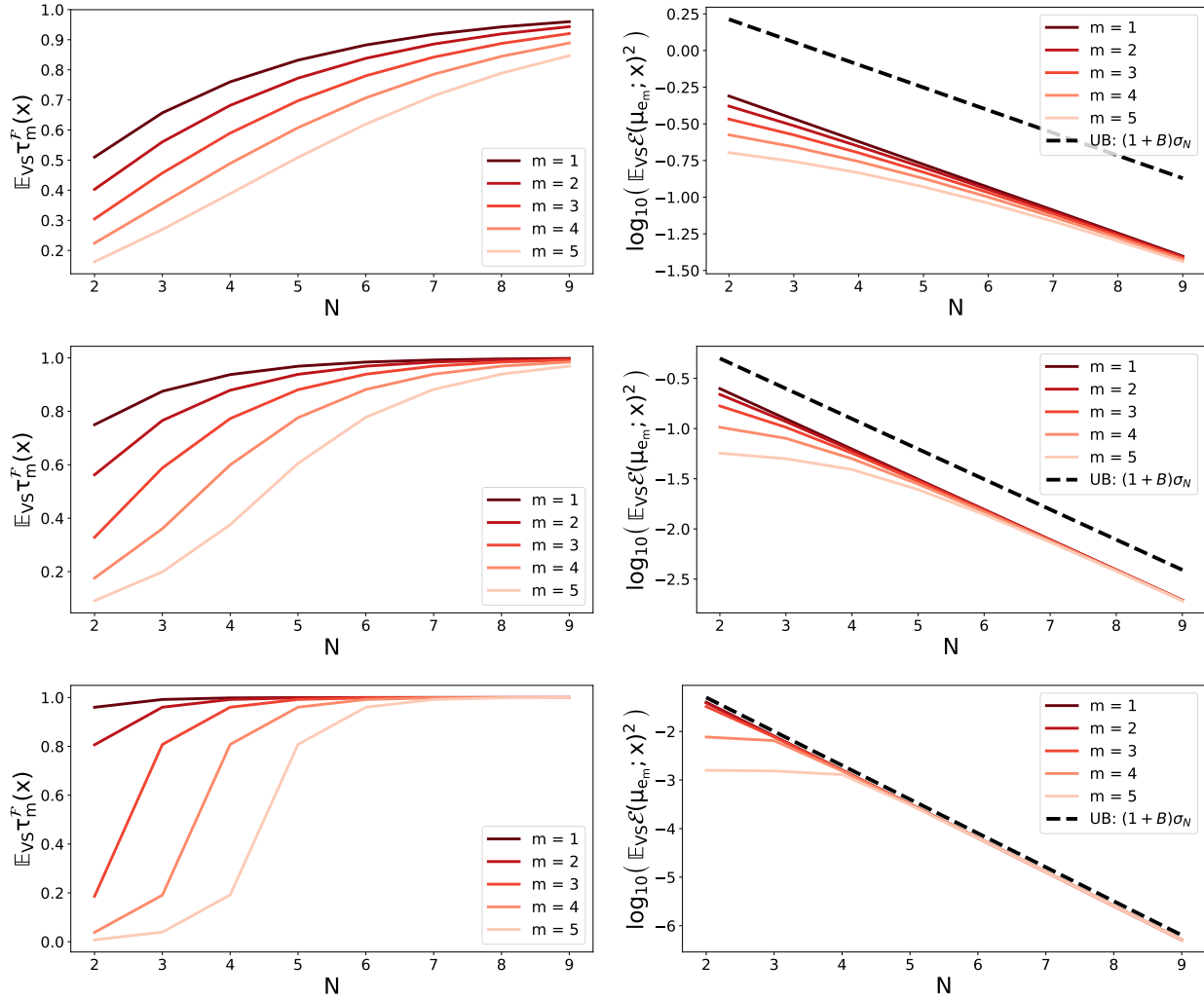


Figure 2. The expected value of the m -th leverage score $\mathbb{E}_{\text{VS}} \tau_m^{\mathcal{F}}(\mathbf{x})$ and the expected interpolation error $\mathbb{E}_{\text{VS}} \mathcal{E}(\mu_{e_m}; \mathbf{x})^2$ under the distribution of continuous volume sampling for $m \in \{1, 2, 3, 4, 5\}$. Every row corresponds to a uni-dimensional Gaussian space ($\sigma_m = \alpha^m$) with a parameter $\alpha \in \{0.7, 0.5, 0.2\}$.

References

- Belhadji, A., Bardenet, R., and Chainais, P. Kernel quadrature with DPPs. In *Advances in Neural Information Processing Systems* 32, pp. 12907–12917. 2019.
- Brezis, H. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- Guruswami, V. and Sinop, A. K. Optimal column-based low-rank matrix reconstruction. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pp. 1207–1214. SIAM, 2012.
- Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. Determinantal processes and independence. *Probability surveys*, 3: 206–229, 2006.
- Kulesza, A. and Taskar, B. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.
- Marcus, A. W., Spielman, D. A., and Srivastava, N. Interlacing families II: Mixed characteristic polynomials and the Kadison—Singer problem. *Annals of Mathematics*, pp. 327–350, 2015.
- Shadrin, A. Twelve proofs of the Markov inequality. *Approximation theory: a volume dedicated to Borislav Bojanov*, pp. 233–298, 2004.
- Steele, J. M. *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, New York, NY, USA, 2004. ISBN 052154677X.
- Steinwart, I. and Scovel, C. Mercer’s theorem on general domains: on the interaction between measures, kernels, and RKHSs. *Constructive Approximation*, 35(3):363–417, 2012.
- Wahba, G. *Spline Models for Observational Data*, volume 59. SIAM, 1990.
- Zhu, H., Williams, C. K. I., Rohwer, R., and Morciniec, M. Gaussian regression and optimal finite dimensional linear models. 1997.