

## SUPPLEMENT TO “AN OPERATOR THEORETIC APPROACH TO NONPARAMETRIC MIXTURE MODELS”

BY ROBERT A. VANDERMEULEN\* AND CLAYTON D. SCOTT†

*Technische Universität Kaiserslautern: Department of Computer Science\**  
*University of Michigan: Electrical and Computer Engineering, Statistics†*

### APPENDIX A: PROOFS OMITTED IN MAIN TEXT

This section contains all proofs omitted in the main text. The first subsection introduces some tools necessary for many of the proofs. The second subsection contains proofs of technical lemmas. The final subsection contains proofs of the major results.

**A.1. A Few Prerequisites from the Main Text.** In Section 8 of the main text we introduced bounded linear operators, Hilbert-Schmidt operators, and lemmas related to linear operators and tensor products. While these results were only mentioned in Section 8 of the main text we need these tools to prove results from earlier sections. We will quickly review them here for convenience. To begin we introduce the set of Hilbert-Schmidt operators which are a subspace of the bounded linear operators of a Hilbert space.

**DEFINITION A.1.** *Let  $H, H'$  be Hilbert spaces and  $T \in \mathcal{L}(H, H')$ .  $T$  is called a Hilbert-Schmidt operator if  $\sum_{x \in J} \|Tx\|^2 < \infty$  for an orthonormal basis  $J \subset H$ . We denote the set of Hilbert-Schmidt operators in  $\mathcal{L}(H, H')$  by  $\mathcal{HS}(H, H')$ .*

This definition does not depend on the choice of orthonormal basis: the sum  $\sum_{x \in J} \|T(x)\|^2$  will always yield the same value regardless of the choice of orthonormal basis  $J$ . The set of Hilbert-Schmidt operators is itself a Hilbert space when equipped with the inner product

$$(1) \quad \sum_{x \in J} \langle Tx, Sx \rangle$$

where  $J$  is an orthonormal basis. Again this value does not depend on the choice of  $J$ . The Hilbert-Schmidt norm will be denoted as  $\|\cdot\|_{\mathcal{HS}}$  and the standard operator norm will have no subscript. There is a well known bound relating the two norms: for a Hilbert-Schmidt operator  $T$  we have that

$$(2) \quad \|T\| \leq \|T\|_{\mathcal{HS}}.$$

We will also need to make use of tensor products of bounded linear operators. The following lemma is exactly Proposition 2.6.12 from [3] and is Lemma 8.1 from the main text.

LEMMA A.1. *Let  $H_1, \dots, H_n, H'_1, \dots, H'_n$  be Hilbert spaces and let  $U_i \in \mathcal{L}(H_i, H'_i)$  for all  $i \in [n]$ . There exists a unique*

$$(3) \quad U \in \mathcal{L}(H_1 \otimes \dots \otimes H_n, H'_1 \otimes \dots \otimes H'_n),$$

*such that  $U(h_1 \otimes \dots \otimes h_n) = U_1(h_1) \otimes \dots \otimes U_n(h_n)$  for all  $h_1 \in H_1, \dots, h_n \in H_n$ .*

DEFINITION A.2. *The operator constructed in Lemma A.1 is called the tensor product of  $U_1, \dots, U_n$  and is denoted  $U_1 \otimes \dots \otimes U_n$ .*

We will also need to utilize the equivalence between tensor products and linear operators ([3] Proposition 2.6.9). The following is Lemma 8.2 from the main text.

LEMMA A.2. *Let  $H, H'$  be Hilbert spaces. There exists a unitary operator  $U : H \otimes H' \rightarrow \mathcal{H}\mathcal{S}(H, H')$  such that, for any simple tensor  $h \otimes h' \in H \otimes H'$ ,  $U(h \otimes h') = \langle h, \cdot \rangle h'$ .*

## A.2. Proofs of Technical Lemmas.

PROOF OF LEMMA 3.1. Because both representations are minimal it follows that  $\alpha'_i \neq 0$  for all  $i$  and  $\nu'_i \neq \nu'_j$  for all  $i \neq j$ . From this we know  $\mathcal{Q}(\{\nu'_i\}) \neq 0$  for all  $i$ . Because  $\mathcal{Q}(\{\nu'_i\}) \neq 0$  for all  $i$  it follows that for any  $i$  there exists some  $j$  such that  $\nu'_i = \nu_j$ . Let  $\psi : [r] \rightarrow [r]$  be a function satisfying  $\nu'_i = \nu_{\psi(i)}$ . Because the elements  $\nu_1, \dots, \nu_r$  are also distinct,  $\psi$  must be injective and thus a permutation. Again from this distinctness we get that, for all  $i$ ,  $\mathcal{Q}(\{\nu'_i\}) = \alpha'_i = \alpha_{\psi(i)}$  and we are done.  $\square$

PROOF OF LEMMA 4.1 AND 4.3. We will proceed by contradiction. Let  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$  be  $n$ -identifiable/determined, let  $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$  be a different mixture of measures, with  $l \leq m$  for the  $n$ -identifiable case, and

$$(4) \quad \sum_{i=1}^m a_i \mu_i^{\times q} = \sum_{j=1}^l b_j \nu_j^{\times q}$$

for some  $q > n$ . Let  $A \in \mathcal{F}^{\times n}$  be arbitrary. We have

$$(5) \quad \sum_{i=1}^m a_i \mu_i^{\times q} = \sum_{j=1}^l b_j \nu_j^{\times q}$$

$$(6) \quad \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times q} (A \times \Omega^{\times q-n}) = \sum_{j=1}^l b_j \nu_j^{\times q} (A \times \Omega^{\times q-n})$$

$$(7) \quad \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times n} (A) = \sum_{j=1}^l b_j \nu_j^{\times n} (A).$$

This implies that  $\mathcal{P}$  is not  $n$ -identifiable/determined, a contradiction.  $\square$

PROOF OF LEMMA 4.2 AND 4.4. Let a mixture of measures  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$  not be  $n$ -identifiable/determined. It follows that there exists a different mixture of measures  $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$ , with  $l \leq m$  for the  $n$ -identifiability case, such that

$$(8) \quad \sum_{i=1}^m a_i \mu_i^{\times n} = \sum_{j=1}^l b_j \nu_j^{\times n}.$$

Let  $A \in \mathcal{F}^{\times q}$  be arbitrary, we have

$$(9) \quad \sum_{i=1}^m a_i \mu_i^{\times n} (A \times \Omega^{\times n-q}) = \sum_{j=1}^l b_j \nu_j^{\times n} (A \times \Omega^{\times n-q})$$

$$(10) \quad \Rightarrow \sum_{i=1}^m a_i \mu_i^{\times q} (A) = \sum_{j=1}^l b_j \nu_j^{\times q} (A)$$

and therefore  $\mathcal{P}$  is not  $q$ -identifiable/determined.  $\square$

PROOF OF LEMMA 5.1. Lemma A.1 states that there exists a continuous linear operator  $\tilde{U} : H_1 \otimes \cdots \otimes H_n \rightarrow H'_1 \otimes \cdots \otimes H'_n$  such that  $\tilde{U}(h_1 \otimes \cdots \otimes h_n) = U_1(h_1) \otimes \cdots \otimes U_n(h_n)$  for all  $h_1 \in H_1, \dots, h_n \in H_n$ . Let  $\hat{H}$  be the set of simple tensors in  $H_1 \otimes \cdots \otimes H_n$  and  $\hat{H}'$  be the set of simple tensors in  $H'_1 \otimes \cdots \otimes H'_n$ . Because  $U_i$  is surjective for all  $i$ , clearly  $\tilde{U}(\hat{H}) = \hat{H}'$ . The linearity of  $\tilde{U}$  implies that  $\tilde{U}(\text{span}(\hat{H})) = \text{span}(\hat{H}')$ . Because  $\text{span}(\hat{H}')$  is dense in  $H'_1 \otimes \cdots \otimes H'_n$  the continuity of  $\tilde{U}$  implies that  $\tilde{U}(H_1 \otimes \cdots \otimes H_n) = H'_1 \otimes \cdots \otimes H'_n$  so  $\tilde{U}$  is surjective. All that remains to be shown is that  $\tilde{U}$  preserves the inner product (see Theorem 4.18 in [5]). By the continuity of inner product we

need only show that  $\langle h, g \rangle = \langle \tilde{U}(h), \tilde{U}(g) \rangle$  for  $h, g \in \text{span}(\widehat{H})$ . With this in mind let  $h_1, \dots, h_N, g_1, \dots, g_M$  be simple tensors in  $H_1 \otimes \dots \otimes H_n$ . We have the following

$$(11) \quad \left\langle \tilde{U} \left( \sum_{i=1}^N h_i \right), \tilde{U} \left( \sum_{j=1}^M g_j \right) \right\rangle = \left\langle \sum_{i=1}^N \tilde{U}(h_i), \sum_{j=1}^M \tilde{U}(g_j) \right\rangle$$

$$(12) \quad = \sum_{i=1}^N \sum_{j=1}^M \langle \tilde{U}(h_i), \tilde{U}(g_j) \rangle$$

$$(13) \quad = \sum_{i=1}^N \sum_{j=1}^M \langle h_i, g_j \rangle$$

$$(14) \quad = \left\langle \sum_{i=1}^N h_i, \sum_{j=1}^M g_j \right\rangle.$$

We have now shown that  $\tilde{U}$  is unitary which completes our proof.  $\square$

**PROOF OF LEMMA 5.2.** Example 2.6.11 in [3] states that for any two  $\sigma$ -finite measure spaces  $(S, \mathcal{S}, m), (S', \mathcal{S}', m')$  there exists a unitary operator  $U : L^2(S, \mathcal{S}, m) \otimes L^2(S', \mathcal{S}', m') \rightarrow L^2(S \times S', \mathcal{S} \times \mathcal{S}', m \times m')$  such that, for all  $f, g$ ,

$$(15) \quad U(f \otimes g) = f(\cdot)g(\cdot).$$

Because  $(\Psi, \mathcal{G}, \gamma)$  is a  $\sigma$ -finite measure space it follows that  $(\Psi^{\times m}, \mathcal{G}^{\times m}, \gamma^{\times m})$  is a  $\sigma$ -finite measure space for all  $m \in \mathbb{N}$ . We will now proceed by induction. Clearly the lemma holds for  $n = 1$ . Suppose the lemma holds for  $n - 1$ . From the induction hypothesis we know that there exists a unitary transform  $U_{n-1} : L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1} \rightarrow L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1})$  such that for all simple tensors  $f_1 \otimes \dots \otimes f_{n-1} \in L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1}$  we have  $U_{n-1}(f_1 \otimes \dots \otimes f_{n-1}) = f_1(\cdot) \dots f_{n-1}(\cdot)$ . Combining  $U_{n-1}$  with the identity map via Lemma 5.1 we can construct a unitary operator  $T_n : L^2(\Psi, \mathcal{G}, \gamma)^{\otimes n-1} \otimes L^2(\Psi, \mathcal{G}, \gamma) \rightarrow L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma)$ , which maps  $f_1 \otimes \dots \otimes f_{n-1} \otimes f_n \mapsto f_1(\cdot) \dots f_{n-1}(\cdot) \otimes f_n$ .

From the aforementioned example there exists a unitary transform

$$(16) \quad \begin{aligned} K_n : L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma) \\ \rightarrow L^2(\Psi^{\times n-1} \times \Psi, \mathcal{G}^{\times n-1} \times \mathcal{G}, \gamma^{\times n-1} \times \gamma) \end{aligned}$$

which maps simple tensors  $g \otimes g' \in L^2(\Psi^{\times n-1}, \mathcal{G}^{\times n-1}, \gamma^{\times n-1}) \otimes L^2(\Psi, \mathcal{G}, \gamma)$  as  $K_n(g \otimes g') = g(\cdot)g'(\cdot)$ . Defining  $U_n(\cdot) = K_n(T_n(\cdot))$  yields our desired unitary transform.  $\square$

PROOF OF LEMMA 5.3. We will proceed by induction. For  $n = 2$  the lemma clearly holds. Suppose the lemma holds for  $n - 1$  and let  $h_1, \dots, h_n$  satisfy the assumptions in the lemma statement. Let  $\alpha_1, \dots, \alpha_n$  satisfy

$$(17) \quad \sum_{i=1}^n \alpha_i h_i^{\otimes n-1} = 0.$$

To finish the proof we will show that  $\alpha_1$  must be zero which can be generalized to any  $\alpha_i$ . Applying Lemma A.2 to (17) we get

$$(18) \quad \sum_{i=1}^n \alpha_i h_i^{\otimes n-2} \langle h_i, \cdot \rangle = 0.$$

Because  $h_1$  and  $h_n$  are linearly independent we can choose  $z$  such that  $\langle h_1, z \rangle \neq 0$  and  $z \perp h_n$ . Plugging  $z$  into (18) yields

$$(19) \quad \sum_{i=1}^{n-1} \alpha_i h_i^{\otimes n-2} \langle h_i, z \rangle = 0$$

and therefore  $\alpha_1 = 0$  by the inductive hypothesis.  $\square$

PROOF OF LEMMA 5.4. Let  $\dim(\text{span}(h_1, \dots, h_m)) = l$  and let  $h = \sum_{i=1}^m h_i^{\otimes 2}$ . Without loss of generality assume that  $h_1, \dots, h_l$  are linearly independent and nonzero. From Lemma A.2 there exists a unitary transform  $U : H \otimes H \rightarrow \mathcal{H}\mathcal{S}(H, H)$  which, for any simple tensor  $x \otimes y$ , we have  $U(x \otimes y) = x \langle y, \cdot \rangle$ .

First we will show that the rank is greater than or equal to  $l$  by contradiction. Suppose that  $g = \sum_{i=1}^{l'} x_i \otimes y_i = h$  with  $l' < l$ . Since  $l' < l$  there must exist some  $j$  such that  $h_j \notin \text{span}(x_1, \dots, x_{l'})$ . Let  $z \perp x_1, \dots, x_{l'}$  and  $z \not\perp h_j$ . Now we have

$$(20) \quad \langle z \otimes z, h \rangle = \sum_{i=1}^m \langle z, h_i \rangle^2 \geq \langle z, h_j \rangle^2 > 0,$$

but

$$(21) \quad \langle z \otimes z, g \rangle = \sum_{i=1}^{l'} \langle z, x_i \rangle \langle z, y_i \rangle = 0,$$

a contradiction.

For the other direction, observe that  $U(h)$  is a compact Hermitian operator and thus admits an spectral decomposition ([5] Theorem 8.15). From

this we have that  $U(h) = \sum_{i=1}^m h_i \langle h_i, \cdot \rangle = \sum_{i=1}^{\infty} \lambda_i \langle \psi_i, \cdot \rangle \psi_i$  with  $(\psi_i)_{i=1}^{\infty}$  orthonormal and  $\lambda_i \geq 0$  for all  $i$  since  $U(h)$  is PSD. Clearly the dimension of the span of  $U(h)$  is less than or equal to  $l$  and thus this decomposition has exactly  $l$  nonzero terms. From this we can let  $U(h) = \sum_{i=1}^l \lambda_i \langle \psi_i, \cdot \rangle \psi_i$  and applying  $U^{-1}$  we have that  $h = \sum_{i=1}^l \lambda_i \psi_i^{\otimes 2}$ . From this it follows that the rank of  $h$  is less than or equal to  $l$  and we are done.  $\square$

**PROOF OF LEMMA 6.1.** The lemma is obvious when  $n = n'$ . Assume that  $n' < n$ . Let  $A \in \mathcal{G}^{\times n'}$  be arbitrary. We have that

$$(22) \quad \sum_{i=1}^m a_i \gamma_i^{\times n} (A \times \Psi^{\times n-n'}) = \sum_{j=1}^l b_j \pi_j^{\times n} (A \times \Psi^{\times n-n'})$$

$$(23) \quad \Rightarrow \sum_{i=1}^m a_i \gamma_i^{\times n'} (A) \gamma_i^{\times n-n'} (\Psi^{\times n-n'}) = \sum_{j=1}^l b_j \pi_j^{\times n'} (A) \pi_j^{\times n-n'} (\Psi^{\times n-n'})$$

$$(24) \quad \Rightarrow \sum_{i=1}^m a_i \gamma_i^{\times n'} (A) = \sum_{j=1}^l b_j \pi_j^{\times n'} (A).$$

Since  $A$  was chosen arbitrarily we have that  $\sum_{i=1}^m a_i \gamma_i^{\times n'} = \sum_{j=1}^l b_j \pi_j^{\times n'}$ .  $\square$

**PROOF OF LEMMA 6.2.** Let  $\pi = \sum_{i=1}^n \gamma_i$ . Because  $\pi$  is  $\sigma$ -finite for all  $i$  we can define  $f_i = \frac{d\gamma_i}{d\pi}$ , where the derivatives are Radon-Nikodym derivatives. Let  $f_k$  be arbitrary. We will first show that  $f_k \leq 1$   $\pi$ -almost everywhere. Suppose there exists a non  $\pi$ -null set  $A \in \mathcal{G}$  such that  $f_i(A) > 1$ . Then we would have

$$(25) \quad \gamma_k(A) = \int_A f_k d\pi$$

$$(26) \quad > \int_A 1 d\pi$$

$$(27) \quad = \sum_{i=1}^n \gamma_i(A)$$

$$(28) \quad \geq \gamma_k(A)$$

a contradiction. From this we have

$$(29) \quad \int f_k^2 d\pi \leq \int 1 d\pi$$

$$(30) \quad \leq \sum_{i=1}^n \gamma_i(\Psi)$$

$$(31) \quad < \infty.$$

From our construction it is clear that  $f_i \geq 0$   $\xi$ -almost everywhere so we can assert  $f_i \geq 0$  without issue.  $\square$

**PROOF OF LEMMA 6.3.** The fact that  $f$  is non-negative and integrable implies that the map  $S \mapsto \int_S f^{\times n} d\pi^{\times n}$  is a bounded measure on  $(\Psi^{\times n}, \mathcal{G}^{\times n})$  (see [2] Exercise 2.12). Let  $R = R_1 \times \cdots \times R_n$  be a rectangle in  $\mathcal{G}^{\times n}$ . Let  $\mathbb{1}_S$  be the indicator function for a set  $S$ . Integrating over  $R$  and using Tonelli's theorem we get

$$(32) \quad \int_R f^{\times n} d\pi^{\times n} = \int \mathbb{1}_R f^{\times n} d\pi^{\times n}$$

$$(33) \quad = \int \left( \prod_{i=1}^n \mathbb{1}_{R_i}(x_i) \right) \left( \prod_{j=1}^n f(x_j) \right) d\pi^{\times n}(x_1, \dots, x_n)$$

$$(34) \quad = \int \cdots \int \left( \prod_{i=1}^n \mathbb{1}_{R_i}(x_i) \right) \left( \prod_{j=1}^n f(x_j) \right) d\pi(x_1) \cdots d\pi(x_n)$$

$$(35) \quad = \int \cdots \int \left( \prod_{i=1}^n \mathbb{1}_{R_i}(x_i) f(x_i) \right) d\pi(x_1) \cdots d\pi(x_n)$$

$$(36) \quad = \prod_{i=1}^n \left( \int \mathbb{1}_{R_i}(x_i) f(x_i) d\pi(x_i) \right)$$

$$(37) \quad = \prod_{i=1}^n \gamma(R_i)$$

$$(38) \quad = \gamma^{\times n}(R).$$

Any product probability measure is uniquely determined by its measure over the rectangles (this is a consequence of Lemma 1.17 in [4] and the definition of product  $\sigma$ -algebra) therefore, for all  $B \in \mathcal{G}^{\times n}$ ,

$$(39) \quad \gamma^{\times n}(B) = \int_B f^{\times n} d\pi^{\times n}.$$

$\square$

### A.3. Proofs of Major Results in Paper.

PROOF OF THEOREM 4.5. Let  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$  be a mixture of measures with linearly independent components. Let  $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$  be a mixture of measures with  $V_3(\mathcal{P}) = V_3(\mathcal{P}')$  and  $l \leq m$ . From Lemma 6.2 there exists a finite measure  $\xi$  and non-negative functions  $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$  such that, for all  $B \in \mathcal{F}$ ,  $\int_B p_i d\xi = \mu_i(B)$  and  $\int_B q_j d\xi = \nu_j(B)$  for all  $i, j$ . Using Lemma 6.1, 6.3, and 6.4 as we did in Theorems 4.1 through 4.4 it follows that

$$(40) \quad \sum_{i=1}^m a_i p_i^{\times 2} = \sum_{j=1}^l b_j q_j^{\times 2}.$$

From Lemma 5.2 we have

$$(41) \quad \sum_{i=1}^m a_i p_i^{\otimes 2} = \sum_{j=1}^l b_j q_j^{\otimes 2}.$$

By Lemma 5.4 we now know that the rank of the LHS of the previous equation is  $m$  and thus  $l = m$  and  $q_1, \dots, q_m$  are linearly independent. We will now show that  $q_j \in \text{span}(\{p_1, \dots, p_m\})$  for all  $j$ . Suppose that  $q_t \notin \text{span}(\{p_1, \dots, p_m\})$ . Then there exists  $z \in L^2(\Omega, \mathcal{F}, \xi)$  such that  $z \perp p_1, \dots, p_m$  but  $z \not\perp q_t$ . Now we have

$$(42) \quad \sum_{i=1}^m a_i p_i^{\otimes 2} = \sum_{j=1}^m b_j q_j^{\otimes 2}$$

$$(43) \quad \Rightarrow \left\langle \sum_{i=1}^m a_i p_i \otimes p_i, z \otimes z \right\rangle = \left\langle \sum_{j=1}^m b_j q_j \otimes q_j, z \otimes z \right\rangle$$

$$(44) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i \otimes p_i, z \otimes z \rangle = \sum_{j=1}^m b_j \langle q_j \otimes q_j, z \otimes z \rangle$$

$$(45) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i, z \rangle^2 = \sum_{j=1}^m b_j \langle q_j, z \rangle^2.$$

We know that the LHS of the last equation is zero but the RHS is not, a contradiction.

Because  $p_1, \dots, p_m$  are linearly independent we can do the following: for each  $k \in [m]$  let  $z_k \in \text{span}(\{p_1, \dots, p_m\})$  so that  $z_k \perp \{p_i : i \neq k\}$  and  $\langle z_k, p_k \rangle = 1$ . By considering elements of  $L^2(\Omega, \mathcal{F}, \xi)^{\otimes 3}$  as elements of

$L^2(\Omega, \mathcal{F}, \xi) \otimes L^2(\Omega, \mathcal{F}, \xi)^{\otimes 2}$ , we can use Lemma A.2 to transform elements in  $L^2(\Omega, \mathcal{F}, \xi)^{\otimes 3}$  into elements of  $\mathcal{HS}(L^2(\Omega, \mathcal{F}, \xi), L^2(\Omega, \mathcal{F}, \xi)^{\otimes 2})$ ,

$$(46) \quad \sum_{i=1}^m a_i p_i^{\otimes 3} = \sum_{j=1}^m b_j q_j^{\otimes 3}$$

$$(47) \quad \Rightarrow \sum_{i=1}^m a_i p_i^{\otimes 2} \langle p_i, \cdot \rangle = \sum_{j=1}^m b_j q_j^{\otimes 2} \langle q_j, \cdot \rangle.$$

It now follows that

$$(48) \quad \begin{aligned} \sum_{i=1}^m a_i p_i^{\otimes 2} \langle p_i, z_k \rangle &= \sum_{j=1}^m b_j q_j^{\otimes 2} \langle q_j, z_k \rangle \\ \Rightarrow a_k p_k^{\otimes 2} &= \sum_{j=1}^m b_j q_j^{\otimes 2} \langle q_j, z_k \rangle. \end{aligned}$$

Using Lemma A.2 we have

$$(49) \quad a_k p_k \langle p_k, \cdot \rangle = \sum_{j=1}^m b_j \langle q_j, z_k \rangle q_j \langle q_j, \cdot \rangle.$$

The LHS of (49) is a rank one operator and thus the RHS must have exactly one nonzero summand, since  $q_1, \dots, q_m$  are linearly independent. Let  $\varphi : [m] \rightarrow [m]$  be a function such that, for all  $k$ ,

$$(50) \quad a_k p_k^{\otimes 2} = \langle q_{\varphi(k)}, z_k \rangle b_{\varphi(k)} q_{\varphi(k)}^{\otimes 2}.$$

From Lemmas 5.2 and 6.3 we have

$$(51) \quad a_k \mu_k^{\times 2} = \langle q_{\varphi(k)}, z_k \rangle b_{\varphi(k)} \nu_{\varphi(k)}^{\times 2},$$

for all  $k$ . By Lemma 6.1 we have that  $a_k \mu_k = \langle q_{\varphi(k)}, z_k \rangle b_{\varphi(k)} \nu_{\varphi(k)}$  for all  $k$  and thus  $\mu_k = \nu_{\varphi(k)}$  since  $\mu_k$  and  $\nu_{\varphi(k)}$  are collinear probability measures. Because  $\mu_i \neq \mu_j$  for all  $i, j$  we have that  $\varphi$  must be a bijection. Let  $\sigma = \varphi^{-1}$ . By Lemma 6.1 we have that

$$(52) \quad \sum_{i=1}^m a_i \mu_i = \sum_{j=1}^m b_j \nu_{\sigma(j)}.$$

Since  $\mu_1, \dots, \mu_m$  are linearly independent the last equation only has one solution for  $b_1, \dots, b_m$ , which is  $b_k = a_{\sigma(k)}$ , for all  $k$ . Thus

$$(53) \quad \mathcal{P}' = \sum_{i=1}^m a_{\sigma(i)} \delta_{\mu_{\sigma(i)}}$$

which is equal to  $\mathcal{P}$ . □

PROOF OF THEOREM 4.6. Let  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$  be a mixture of measures with linearly independent components. We will proceed by contradiction: let  $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j} \neq \mathcal{P}$  be a mixture of measures with  $V_4(\mathcal{P}) = V_4(\mathcal{P}')$ . From Theorem 4.1 we know that  $\mathcal{P}$  is 3-identifiable. By Lemma 4.1 it follows that  $\mathcal{P}$  is 4-identifiable and thus  $l > m$ . From Lemma 6.2 there exists a finite measure  $\xi$  and non-negative functions  $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$  such that, for all  $B \in \mathcal{F}$ ,  $\int_B p_i d\xi = \mu_i(B)$  and  $\int_B q_j d\xi = \nu_j(B)$  for all  $i, j$ .

Proceeding as we did in the previous theorem proofs we have that

$$(54) \quad \sum_{i=1}^m a_i p_i^{\otimes 4} = \sum_{j=1}^l b_j q_j^{\otimes 4}.$$

Suppose that there exists  $k$  such that  $\nu_k \notin \text{span}(\{\mu_1, \dots, \mu_m\})$ . From this it would follow that there exists  $z$  such that  $z \perp \{p_1, \dots, p_m\}$  and  $z \not\perp q_k$ . Then we would have that

$$(55) \quad \left\langle \sum_{i=1}^m a_i p_i^{\otimes 4}, z^{\otimes 4} \right\rangle = \left\langle \sum_{j=1}^l b_j q_j^{\otimes 4}, z^{\otimes 4} \right\rangle$$

$$(56) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i, z \rangle^4 = \sum_{j=1}^l b_j \langle q_j, z \rangle^4,$$

but the LHS of the last equation is 0 and the RHS is positive, a contradiction. Thus we have that  $q_k \in \text{span}(\{p_1, \dots, p_m\})$  for all  $k$ .

Since  $l > m$  and no pair of elements in  $q_1, \dots, q_m$  are collinear, there must exist a vector in  $q_1, \dots, q_l$  which is a nontrivial linear combination of  $p_1, \dots, p_m$ . Without loss of generality we will assume that  $q_1 = \sum_{i=1}^m c_i p_i$  with  $c_1$  and  $c_2$  nonzero. By the linear independence of  $p_1, \dots, p_m$  there must exist vectors  $z_1, z_2$  such that  $\langle z_1, p_1 \rangle = 1$ ,  $z_1 \perp \{p_i : i \neq 1\}$ ,  $\langle z_2, p_2 \rangle = 1$ , and  $z_2 \perp \{p_i : i \neq 2\}$ . Now consider

$$(57) \quad \left\langle \sum_{i=1}^m a_i p_i^{\otimes 4}, z_1^{\otimes 2} \otimes z_2^{\otimes 2} \right\rangle = \left\langle \sum_{j=1}^l b_j q_j^{\otimes 4}, z_1^{\otimes 2} \otimes z_2^{\otimes 2} \right\rangle$$

$$(58) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i, z_1 \rangle^2 \langle p_i, z_2 \rangle^2 = \sum_{j=1}^l b_j \langle q_j, z_1 \rangle^2 \langle q_j, z_2 \rangle^2.$$

The LHS of the last equation is 0 and the RHS is positive, a contradiction.  $\square$

PROOF OF THEOREM 4.7. Let  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\mu_i}$  be a mixture of measures with jointly irreducible components. Consider a mixture of measures  $\mathcal{P}' = \sum_{j=1}^l b_j \delta_{\nu_j}$  with  $V_2(\mathcal{P}) = V_2(\mathcal{P}')$ . From Lemma 6.2 there exists a finite measure  $\xi$  and non-negative functions  $p_1, \dots, p_m, q_1, \dots, q_l \in L^1(\Omega, \mathcal{F}, \xi) \cap L^2(\Omega, \mathcal{F}, \xi)$  such that, for all  $B \in \mathcal{F}$ ,  $\int_B p_i d\xi = \mu_i(B)$  and  $\int_B q_j d\xi = \nu_j(B)$  for all  $i, j$ . Proceeding as we have done in the previous theorem proofs we have

$$(59) \quad \sum_{i=1}^m a_i p_i \times p_i = \sum_{j=1}^l b_j q_j \times q_j.$$

From Lemma 5.2 we have

$$(60) \quad \sum_{i=1}^m a_i p_i \otimes p_i = \sum_{j=1}^l b_j q_j \otimes q_j.$$

Suppose for a moment that  $\mathcal{P}'$  contains a mixture component which does not lie in  $\text{span}(\{\mu_1, \dots, \mu_m\})$ . Without loss of generality we will assume that  $\nu_1 \notin \text{span}(\{\mu_1, \dots, \mu_m\})$ . Recall that joint irreducibility implies linear independence so  $\nu_1, \mu_1, \dots, \mu_m$  are a linearly independent set of measures and thus  $q_1, p_1, \dots, p_m$  are linearly independent. It follows that we can find some  $z \in L^2(\Omega, \mathcal{F}, \xi)$  such that  $\langle z, q_1 \rangle \neq 0$  and  $z \perp \{p_i : i \in [m]\}$ . From (60) we have the following

$$(61) \quad \left\langle \sum_{i=1}^m a_i p_i \otimes p_i, z \otimes z \right\rangle = \left\langle \sum_{j=1}^l b_j q_j \otimes q_j, z \otimes z \right\rangle$$

$$(62) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i \otimes p_i, z \otimes z \rangle = \sum_{j=1}^l b_j \langle q_j \otimes q_j, z \otimes z \rangle$$

$$(63) \quad \Rightarrow \sum_{i=1}^m a_i \langle p_i, z \rangle^2 = \sum_{j=1}^l b_j \langle q_j, z \rangle^2.$$

All the summands on both sides of the last equation are nonnegative. By our construction of  $z$  the LHS of the previous equation is zero and the first summand on the RHS is positive, a contradiction. Thus, each component in  $\mathcal{P}'$  must lie in the span of the components of  $\mathcal{P}$ .

Now we have, for all  $j$ ,  $q_j = \sum_{i=1}^m c_i^j p_i$ . From joint irreducibility we have that  $c_i^j \geq 0$  for all  $i$  and  $j$ . We will now consider the situation where a component of  $\mathcal{P}'$  is a nontrivial linear combination of components in  $\mathcal{P}$ . Suppose that there exists  $r, s, s'$ , with  $s \neq s'$ , such that  $c_s^r, c_{s'}^r > 0$ . From the

linear independence of  $p_1, \dots, p_m$  we can find a  $z$  such that  $\langle p_s, z \rangle = 1$  and  $z \perp \{p_q : q \in [m] \setminus \{s\}\}$ . Applying Lemma A.2 to (60) we have

$$(64) \quad \sum_{i=1}^m a_i p_i \langle p_i, \cdot \rangle = \sum_{j=1}^l b_j q_j \langle q_j, \cdot \rangle$$

$$(65) \quad \Rightarrow \sum_{i=1}^m a_i p_i \langle p_i, z \rangle = \sum_{j=1}^l b_j q_j \langle q_j, z \rangle$$

$$(66) \quad \Rightarrow a_s p_s = \sum_{j=1}^l b_j \left[ \sum_{t=1}^m c_t^j p_t \right] \left\langle \sum_{u=1}^m c_u^j p_u, z \right\rangle$$

$$(67) \quad \Rightarrow a_s p_s = \sum_{j=1}^l b_j \left[ \sum_{t=1}^m c_t^j p_t \right] c_s^j$$

$$(68) \quad = \sum_{t=1}^m \sum_{j=1}^l b_j c_t^j c_s^j p_t$$

$$(69) \quad = \sum_{t=1}^m p_t \sum_{j=1}^l b_j c_t^j c_s^j.$$

Let  $\alpha_t = \sum_{j=1}^l b_j c_t^j c_s^j$  for all  $t$  and note that each summand is nonnegative. Now we have

$$(70) \quad a_s p_s = \sum_{t=1}^m \alpha_t p_t.$$

We know that  $\alpha_{s'} > 0$  since  $b_r c_s^r c_{s'}^r > 0$ . This violates the linear independence of  $p_1, \dots, p_m$ . Now we have that for all  $i$  there exists  $j$  such that  $p_i = q_j$ . From the minimality of the representation of mixtures of measures it follows that  $l = m$  and without loss of generality we can assert that  $p_i = q_i$  for all  $i$  and thus  $\mu_i = \nu_i$  for all  $i$ . Because  $p_1, \dots, p_m$  are linearly independent it follows that  $p_1 \otimes p_1, \dots, p_m \otimes p_m$  are linearly independent. We can show this by the contrapositive, suppose  $p_1 \otimes p_1, \dots, p_m \otimes p_m$  are not linearly independent then there exists a nontrivial linear combination such that  $\sum_{i=1}^m \kappa_i p_i \otimes p_i = 0$ . Assume without loss of generality that  $\kappa_1 \neq 0$ . Applying Lemma A.2 we get

that

$$(71) \quad \sum_{i=1}^m \kappa_i p_i \langle p_i, \cdot \rangle = 0$$

$$(72) \quad \Rightarrow \sum_{i=1}^m \kappa_i p_i \langle p_i, p_1 \rangle = 0$$

$$(73) \quad \Rightarrow \kappa_1 p_1 \|p_1\|_{L^2}^2 + \sum_{i=2}^m \kappa_i p_i \langle p_i, p_1 \rangle = 0$$

$$(74)$$

and thus  $p_1, \dots, p_m$  are not linearly independent.

Since  $p_1 \otimes p_1, \dots, p_m \otimes p_m$  are linearly independent it follows that  $a_i = b_i$  for all  $i$  and thus  $\mathcal{P} = \mathcal{P}'$ . □

PROOF OF LEMMA 7.1. For brevity's sake let  $Q = T_{n,q}(Q_{n,p,q})$  and  $R = V_n(\delta_{\sum_{i=1}^q p_i \delta_i})$ . Let  $y \in [q]^{\times n}$  be arbitrary. We will prove that  $Q(\{y\}) = R(\{y\})$  which, since  $y$  is arbitrary, clearly generalizes to  $Q = R$ . Let  $\tilde{y} \in C_{n,q}$  be the element such that  $\tilde{y}_i = |\{j : y_j = i\}|$  for all  $i$ , i.e. the  $i$ th index of  $\tilde{y}$  contains the number of times the value  $i$  occurs in  $y$ . From the definition of  $V_n$  we have that

$$(75) \quad R(\{y\}) = \left( \sum_{i=1}^q p_i \delta_i \right)^{\times n} (\{y\}) = \prod_{i=1}^n p_{y_i} = \prod_{j=1}^q p_j^{\tilde{y}_j}.$$

We define  $\chi$  to be the indicator function, which is equal to 1 if its subscript is true and 0 otherwise. Consider some  $z \neq \tilde{y}$ . We have

$$(76) \quad T_{n,q}(\delta_z)(\{y\}) = \frac{1}{n!} \sum_{\sigma \in S_n} \delta_{\sigma(F_{n,q}(z))}(\{y\})$$

$$(77) \quad = \frac{1}{n!} \sum_{\sigma \in S_n} \chi_{\sigma(F_{n,q}(z))=y}.$$

From our definition of  $F_{n,q}$  and  $\tilde{y}$  it is clear that there must exist some  $r$  such that the number of entries of  $F_{n,q}(z)$  which equal  $r$  is different from the number of indices of  $y$  which equal  $r$ . Because of this no permutation of  $F_{n,q}(z)$  can equal  $y$  and thus  $T_{n,q}(\delta_z)(\{y\}) = 0$ . From this it follows that  $T_{n,q}(\delta_z)(\{y\}) = 0$  for all  $z \neq \tilde{y}$ .

Now we will consider  $T_{n,q}(\delta_{\check{y}})(\{y\})$ . Again we have

$$(78) \quad T_{n,q}(\delta_{\check{y}})(\{y\}) = \frac{1}{n!} \sum_{\sigma \in S_n} \chi_{\sigma(F_{n,q}(\check{y}))=y},$$

so we need only determine how many permutations of  $F_{n,q}(\check{y})$  are equal to  $y$ . Basic combinatorics tells us that there are  $\check{y}_1! \cdots \check{y}_q!$  such permutations. The coefficient of  $\delta_{\check{y}}$  in  $Q_{n,p,q}$  is  $\frac{n!}{\check{y}_1! \cdots \check{y}_q!} p_1^{\check{y}_1} \cdots p_n^{\check{y}_n}$  so we have that  $Q(\{y\}) = R(\{y\})$  by direct evaluation.  $\square$

PROOF OF COROLLARY 7.1. We will proceed by contradiction and assume that there exists two mixtures of the form in the corollary statement,

$$(79) \quad \sum_{i=1}^m a_i Q_{n,p_i,q} = \sum_{j=1}^s b_j Q_{n,r_j,q}$$

but  $s \neq m$  or  $s = m$  and there exists no permutation such that  $a_i Q_{n,p_i,q} = b_{\sigma(i)} Q_{n,r_{\sigma(i)},q}$ . If we apply  $T_{n,q}$  defined earlier, from Lemma 7.1 it follows that

$$(80) \quad V_n \left( \sum_{i=1}^m a_i \delta_{\sum_{k=1}^q p_{i,k} \delta_k} \right) = V_n \left( \sum_{j=1}^s b_j \delta_{\sum_{l=1}^q r_{j,l} \delta_l} \right).$$

We have that  $\mathcal{P} = \sum_{i=1}^m a_i \delta_{\sum_{k=1}^q p_{i,k} \delta_k}$  and  $\mathcal{P}' = \sum_{j=1}^s b_j \delta_{\sum_{l=1}^q r_{j,l} \delta_l}$  are mixtures of measures which are not  $n$ -identifiable. Our contradiction hypothesis implies that  $\mathcal{P} \neq \mathcal{P}'$ . From Lemma 4.2 we have that

$$(81) \quad V_{2m-1} \left( \sum_{i=1}^m a_i \delta_{\sum_{k=1}^q p_{i,k} \delta_k} \right) = V_{2m-1} \left( \sum_{j=1}^s b_j \delta_{\sum_{l=1}^q r_{j,l} \delta_l} \right),$$

which contradicts Theorem 4.1.  $\square$

PROOF OF COROLLARY 7.2. The proof of this corollary is virtually identical to the proof of Corollary 7.1, it simply uses Theorem 4.3 instead of Theorem 4.1 and related notions of determinedness.  $\square$

## APPENDIX B: “RANDOM DOMINATING MEASURE” TRICK

Here we introduce the technique mentioned in Sections 8 and 9 of the main text which ensures that mixture components over a finite sample space have distinct norms, the “random dominating measure” trick. Let  $(\Omega, 2^\Omega)$

be a finite sample space with  $\Omega = \{\omega_1, \dots, \omega_d\}$ . Let  $\mu_1, \dots, \mu_m$  be distinct measures on this space. Let  $y_1, \dots, y_d \stackrel{iid}{\sim} \text{unif}(1, 2)$  and let  $\xi$  be a random measure on  $(\Omega, 2^\Omega)$  defined by  $\xi(\{\omega_i\}) = y_i$  for all  $i$ . Clearly  $\xi$  dominates all  $\mu_1, \dots, \mu_m$  and thus we can define Radon-Nikodym derivatives  $p_i = \frac{d\mu_i}{d\xi}$  for all  $i$ . We will treat these Radon-Nikodym derivatives as being elements in  $L^2(\Omega, 2^\Omega, \xi)$ . We have the following lemma.

LEMMA B.1. *With probability one*

$$(82) \quad \int p_i(\omega)^2 d\xi(\omega) \neq \int p_j(\omega)^2 d\xi(\omega)$$

for all  $i \neq j$ .

PROOF. Observe that, for all  $i, j$ ,

$$(83) \quad \int_{\{\omega_j\}} p_i d\xi = p_i(\omega_j) \xi(\{\omega_j\}) = p_i(\omega_j) y_j = \mu_i(\{\omega_j\})$$

and thus  $p_i(\omega_j) = \frac{\mu_i(\{\omega_j\})}{y_j}$ . We will show that  $\|p_1\|_{\ell^2(\mathbb{R}^d)}^2 \neq \|p_2\|_{\ell^2(\mathbb{R}^d)}^2$  with probability one, which implies  $\|p_i\|_{\ell^2(\mathbb{R}^d)} \neq \|p_j\|_{\ell^2(\mathbb{R}^d)}$  for all  $i \neq j$  with probability one (here and for the rest of this supplement  $\|\cdot\|_{\ell^2(\mathbb{R}^d)}$  will denote the standard Euclidean norm on  $\mathbb{R}^d$  and  $\langle \cdot, \cdot \rangle_{\ell^2(\mathbb{R}^d)}$  the standard inner product).

Because  $\mu_1 \neq \mu_2$  it follows that there exists some  $j$  such that  $\mu_1(\{\omega_j\}) \neq \mu_2(\{\omega_j\})$ . Without loss of generality we will assume that  $j = 1$  in the previous statement. Now we have

$$(84) \quad P\left(\int p_1(\omega)^2 d\xi(\omega) = \int p_2(\omega)^2 d\xi(\omega)\right)$$

$$(85) \quad = P\left(\sum_{i=1}^d \frac{\mu_1(\{\omega_i\})^2}{y_i} = \sum_{j=1}^d \frac{\mu_2(\{\omega_j\})^2}{y_j}\right)$$

$$(86) \quad = P\left(\left(\frac{\mu_1(\{\omega_1\})^2}{y_1} - \frac{\mu_2(\{\omega_1\})^2}{y_1}\right) = \left(\sum_{i=2}^d \frac{\mu_1(\{\omega_i\})^2}{y_i} - \sum_{j=2}^d \frac{\mu_2(\{\omega_j\})^2}{y_j}\right)\right)$$

which is clearly zero since  $(\mu_1(\{\omega_1\}))^2 - (\mu_2(\{\omega_1\}))^2 \neq 0$  and  $y_1, \dots, y_d$  are all independent random variables and from a non-atomic measure.  $\square$

### APPENDIX C: RECOVERY ALGORITHM FOR FINITE SAMPLE SPACES

Let  $(\Omega, 2^\Omega)$  be a finite measurable space with  $|\Omega| = d$ . To simplify exposition we will assume that  $\Omega$  is simply the set of  $d$  dimensional indicator

vectors in  $\mathbb{R}^d$ ,  $e_1, \dots, e_d$ . Note that Euclidean space with the standard inner product is  $L^2\left(\Omega, 2^\Omega, \sum_{i=1}^d \delta_{e_i}\right) = \ell^2(\mathbb{R}^d)$ . Let  $\mu_1, \dots, \mu_m$  be distinct probability measures on  $\Omega$ . Let  $\mathcal{P} = \sum_{i=1}^m w_i \delta_{\mu_i}$  be a mixture of measures. Let  $\tilde{p}_i \triangleq \mathbb{E}_{x \sim \mu_i}[x]$  for all  $i$ . Note that  $\tilde{p}_{i,j} = \mu_i(\{e_j\})$  for all  $i, j$ . Let  $\mathbf{X}_1, \mathbf{X}_2, \dots \stackrel{iid}{\sim} V_{2m-1}(\mathcal{P})$  with  $\mathbf{X}_i = [X_{i,1}, \dots, X_{i,2m-1}]$ .

To begin we construct the random dominating measure described in Appendix B. Let  $y_1, \dots, y_d \stackrel{iid}{\sim} \text{unif}(1, 2)$ . The random dominating measure  $\xi$  is defined by  $\xi(\{e_i\}) = y_i$  for all  $i$ . Let  $p_i = \frac{d\mu_i}{d\xi}$ , i.e.  $p_i(e_j) = \frac{\tilde{p}_{i,j}}{y_j}$  for all  $i$  and  $j$ . There is a bit of a computational issue with this representation for the densities  $p_1, \dots, p_m$  since the new dominating measure changes the inner product from the standard inner product. We can remedy this with the following lemma.

LEMMA C.1. *Let  $x, v \in \ell^2(\mathbb{R}^d)$ ,  $\xi$  be as above, and*

$$(87) \quad B = \begin{bmatrix} \frac{1}{\sqrt{y_1}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{y_2}} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\sqrt{y_d}} \end{bmatrix}.$$

Then  $\langle Bx, Bv \rangle_{L^2(\Omega, 2^\Omega, \xi)} = \langle x, v \rangle_{\ell^2(\mathbb{R}^d)}$ .

PROOF OF LEMMA C.1. We have

$$(88) \quad \langle Bx, Bv \rangle_{L^2(\Omega, 2^\Omega, \xi)} = \int (Bx)(i)(Bv)(i) d\xi(i)$$

$$(89) \quad = \sum_{i=1}^d (Bx)(i)(Bv)(i) y_i$$

$$(90) \quad = \sum_{i=1}^d \frac{x(i)}{\sqrt{y_i}} \frac{v(i)}{\sqrt{y_i}} y_i$$

$$(91) \quad = \sum_{i=1}^d x(i) y(i)$$

$$(92) \quad = \langle x, y \rangle_{\ell^2(\mathbb{R}^d)}.$$

□

From this lemma we have that  $B$ , when considered as an operator in  $\mathcal{L}(\ell^2(\mathbb{R}^d), L^2(\Omega, 2^\Omega, \xi))$ , is a unitary transform. We are interested in estimating the tensor  $\sum_{i=1}^m w_i p_i^{\otimes 2m-1}$ , but in order to keep the algorithm operating in standard Euclidean space we will instead transform it into  $\ell^2(\mathbb{R}^d)$ . To this end consider an arbitrary  $i$ . We have

$$\begin{aligned} (93) \quad B^{-1}p_i &= B^{-1}[p_{i,1}, \dots, p_{i,d}]^T \\ (94) \quad &= B^{-1}\left[\frac{\tilde{p}_{i,1}}{y_1}, \dots, \frac{\tilde{p}_{i,d}}{y_d}\right]^T \\ (95) \quad &= \left[\frac{\tilde{p}_{i,1}}{\sqrt{y_1}}, \dots, \frac{\tilde{p}_{i,d}}{\sqrt{y_d}}\right]^T, \end{aligned}$$

and thus  $B^{-1}p_j = B\tilde{p}_j$  for all  $j$ .

We will use the following lemma to find the expected value of

$$(96) \quad \mathbb{E}[BX_{i,1} \otimes \dots \otimes BX_{i,2m-1}].$$

LEMMA C.2. *Let  $n > 1$  and  $Z_1, \dots, Z_n$  be independent random vectors in  $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$  such that  $\mathbb{E}[Z_i]$  exists for all  $i$ . Then  $\mathbb{E}[Z_1 \otimes \dots \otimes Z_n] = \mathbb{E}[Z_1] \otimes \dots \otimes \mathbb{E}[Z_n]$ .*

PROOF OF LEMMA C.2. Let  $[i_1, \dots, i_n] \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_n}$  be arbitrary. We have that

$$\begin{aligned} (97) \quad \mathbb{E}[Z_1 \otimes \dots \otimes Z_n]_{i_1, \dots, i_n} &= \mathbb{E}[Z_{1,i_1} \dots Z_{n,i_n}] \\ (98) \quad &= \mathbb{E}[Z_{1,i_1}] \dots \mathbb{E}[Z_{n,i_n}]. \end{aligned}$$

Since  $i_1, \dots, i_n$  were arbitrary it implies that all entries of  $\mathbb{E}[Z_1 \otimes \dots \otimes Z_n]$  and  $\mathbb{E}[Z_1] \otimes \dots \otimes \mathbb{E}[Z_n]$  are equal.  $\square$

Recall that  $X_{i,1}, \dots, X_{i,2m-1} \stackrel{iid}{\sim} \mu$  with  $\mu \sim \mathcal{P}$ . From the previous lemma and the definition of  $\tilde{p}_i$  it follows that

$$\begin{aligned} (99) \quad &\mathbb{E}[BX_{i,1} \otimes \dots \otimes BX_{i,2m-1}] \\ (100) \quad &= \mathbb{E}_{\mu \sim \mathcal{P}}[\mathbb{E}[BX_{i,1} \otimes \dots \otimes BX_{i,2m-1} | \mu]] \\ (101) \quad &= \mathbb{E}_{\mu \sim \mathcal{P}}[\mathbb{E}[BX_{i,1} | \mu] \otimes \dots \otimes \mathbb{E}[BX_{i,2m-1} | \mu]] \\ (102) \quad &= \mathbb{E}_{\mu \sim \mathcal{P}}[B\mathbb{E}[X_{i,1} | \mu] \otimes \dots \otimes B\mathbb{E}[X_{i,2m-1} | \mu]] \\ (103) \quad &= \sum_{i=1}^m w_i B\mathbb{E}[X_{i,1} | \mu = \mu_i] \otimes \dots \otimes B\mathbb{E}[X_{i,2m-1} | \mu = \mu_i] \\ (104) \quad &= \sum_{i=1}^m w_i (B\tilde{p}_i)^{\otimes 2m-1}. \end{aligned}$$

Let  $Y_{i,j} = BX_{i,j}$  and recall that  $S_n$  is the symmetric group over  $n$  symbols. Now we will construct the whitening operator. To do this first construct the operator

$$(105) \quad \widehat{C} = \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} Y_{i,\sigma(1)} \otimes \cdots \otimes Y_{i,\sigma(m-1)} \langle Y_{i,\sigma(m)} \otimes \cdots \otimes Y_{i,\sigma(2m-2)}, \cdot \rangle.$$

There are some repeated terms in the previous summation, which is not an issue. Instead we could have set  $\widehat{C}$  to be equal to

$$\frac{1}{(2m-2)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-2}} Y_{i,\sigma(1)} \otimes \cdots \otimes Y_{i,\sigma(m-1)} \langle Y_{i,\sigma(m)} \otimes \cdots \otimes Y_{i,\sigma(2m-2)}, \cdot \rangle,$$

but this would not utilize all the data, specifically  $Y_{1,2m-1}, \dots, Y_{n,2m-1}$ . In the second operator, the average over  $S_{2m-2}$  functions as a projection onto the space of symmetric tensors [1] and the summation over  $S_{2m-1}$  in the definition of  $\widehat{C}$  serves a similar purpose. Viewed alternatively, the distribution of  $[Y_{i,1}, \dots, Y_{i,2m-1}]^T$  does not change if we reorder the entries of the vector, so the summation is considering all possible orderings of random groups. This symmetrization conveniently assures that  $\widehat{C}$  is a Hermitian operator. This  $\widehat{C}$  is estimating the  $C$  mentioned in the Section 8 of the main text. Let  $\lambda_{\widehat{C},1}, \dots, \lambda_{\widehat{C},m}$  be the top  $m$  eigenvalues of  $\widehat{C}$  and  $v_{\widehat{C},1}, \dots, v_{\widehat{C},m}$  be their associated eigenvectors. We can now construct the whitening operator

$$(106) \quad \widehat{W} = \sum_{i=1}^m \lambda_{\widehat{C},i}^{-\frac{1}{2}} v_{\widehat{C},i} \langle v_{\widehat{C},i}, \cdot \rangle.$$

Now construct the tensor

$$(107) \quad \widehat{A} = \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} Y_{i,\sigma(1)} \otimes \widehat{W} (Y_{i,\sigma(2)} \otimes \cdots \otimes Y_{i,\sigma(m)}) \otimes \cdots \widehat{W} (Y_{i,\sigma(m+1)} \otimes \cdots \otimes Y_{i,\sigma(2m-1)}).$$

Using simple unfolding techniques we can transform  $\widehat{A}$  in to the operator  $\widehat{T}$ :

$$(108) \quad \widehat{T} = \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} Y_{i,\sigma(1)} \otimes \widehat{W} (Y_{i,\sigma(2)} \otimes \cdots \otimes Y_{i,\sigma(m)}) \cdots \langle \widehat{W} (Y_{i,\sigma(m+1)} \otimes \cdots \otimes Y_{i,\sigma(2m-1)}), \cdot \rangle,$$

as well as its Hermitian,  $\widehat{T}^H$ :

$$(109) \quad \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} \widehat{W} (Y_{i,\sigma(m+1)} \otimes \cdots \otimes Y_{i,\sigma(2m-1)}) \cdots \langle Y_{i,\sigma(1)} \otimes \widehat{W} (Y_{i,\sigma(2)} \otimes \cdots \otimes Y_{i,\sigma(m)}), \cdot \rangle.$$

Let  $v_1, \dots, v_m$  be the top  $m$  eigenvectors of  $\widehat{T}\widehat{T}^H$ , which will be elements of  $\ell^2(\mathbb{R}^d)^{\otimes m}$ . These vectors are estimates of  $\|B\tilde{p}_1\|_2^{-1} B\tilde{p}_1 \otimes \widehat{W}\sqrt{w_1}(B\tilde{p}_1)^{\otimes m-1}, \dots, \|B\tilde{p}_m\|_2^{-1} B\tilde{p}_m \otimes \widehat{W}\sqrt{w_m}(B\tilde{p}_m)^{\otimes m-1}$  (possibly multiplied by  $-1$ ). The factors in front of the tensors normalize the tensors to have norm 1.

Using a transform of the form in Lemma A.2, we can implement a transform

$$(110) \quad U : \ell^2(\mathbb{R}^d)^{\otimes m} \rightarrow \mathcal{HS} \left( \ell^2(\mathbb{R}^d)^{\otimes m-1}, \ell^2(\mathbb{R}^d) \right)$$

which maps simple tensors  $x_1 \otimes \cdots \otimes x_m$  to  $x_1 \langle x_2 \otimes \cdots \otimes x_m, \cdot \rangle$ . Applying this transform to  $v_1, \dots, v_m$  yields estimates of  $\|B\tilde{p}_i\|_{\ell^2(\mathbb{R}^d)}^{-1} B\tilde{p}_i \langle \widehat{W}\sqrt{w_i}(B\tilde{p}_i)^{\otimes m-1}, \cdot \rangle$ , for all  $i$ . At this point one simply needs to find vectors  $q_1, \dots, q_m$  which are not orthogonal to  $\widehat{W}\sqrt{w_1}(B\tilde{p}_1)^{\otimes m-1}, \dots, \widehat{W}\sqrt{w_m}(B\tilde{p}_m)^{\otimes m-1}$  to get  $\|B\tilde{p}_i\|_{\ell^2(\mathbb{R}^d)}^{-1} B\tilde{p}_i \langle \widehat{W}\sqrt{w_i}(B\tilde{p}_i)^{\otimes m-1}, q_i \rangle$ , which is  $B\tilde{p}_i, \dots, B\tilde{p}_i$  up to scaling. Such vectors can be found by simply using a tensor populated by iid standard normal random variables. After this we can recover  $\tilde{p}_1, \dots, \tilde{p}_m$ , up to scaling, by simply applying  $B^{-1}$ , which we would then want to normalize to sum to one. Alternatively we could take the largest left singular vector of these operators. We will call these estimates  $\widehat{p}_1, \dots, \widehat{p}_m$ .

Using the data we can estimate the tensor  $\sum_{i=1}^m w_i \tilde{p}_i^{\otimes m-1}$  with the estimator

$$(111) \quad \widehat{E} = \frac{1}{2m-1} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} X_{i,\sigma(1)} \otimes \cdots \otimes X_{i,\sigma(m-1)}.$$

To estimate the mixture proportions we find the value of  $\alpha = (\alpha_1, \dots, \alpha_m)$  which minimizes

$$(112) \quad \left\| \widehat{E} - \sum_{i=1}^m \alpha_i \widehat{p}_i^{\otimes m-1} \right\|.$$

## APPENDIX D: CONSISTENCY OF RECOVERY ALGORITHM

Here we show that the recovery algorithm for categorical distributions described in Appendix C is consistent. Let  $C, \widehat{C}, T, \widehat{T}, W,$  and  $\widehat{W}$  be as they were defined Appendix C and Section 8 of the main text. The crux of our algorithm is the recovery of the eigenvectors of  $TT^H$ , from which we then recover the mixture components through the application of linear and continuous transforms to the eigenvectors. In order to simplify the notation in our explanation we will assume that the norms of  $\tilde{p}_1, \dots, \tilde{p}_m$  are distinct. We do this so that there are gaps in the spectral decomposition of  $TT^H$  thus making the random dominating measure trick unnecessary. Were this not the case, we could simply represent the probability vectors as densities with respect to some dominating measure which makes their norms distinct, as we did in the previous section. Because of this assumption we can simply set  $B$  to be the identity operator. From this we have that  $p_i = \tilde{p}_i$  for all  $i$  and  $X_{i,j} = Y_{i,j}$  for all  $i$  and  $j$ . The following theorem demonstrates that the algorithm does indeed recover the eigenvectors of  $TT^H$ .

THEOREM D.1. *With  $T$  and  $\widehat{T}$  defined as above, as  $n \rightarrow \infty$  then*

$$(113) \quad \left\| TT^H - \widehat{T}\widehat{T}^H \right\|_{\mathcal{H}\mathcal{S}} \xrightarrow{p} 0.$$

PROOF OF THEOREM D.1. Let

$$(114) \quad Q = \sum_{i=1}^m w_i p_i^{\otimes 2m-1}$$

and

$$(115) \quad \widehat{Q} = \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} X_{i,\sigma(1)} \otimes \cdots \otimes X_{i,\sigma(2m-1)}.$$

Note that

$$(116) \quad (I \otimes W \otimes W)(Q) = \sum_{i=1}^m w_i p_i \otimes W(p_i^{\otimes m-1}) \otimes W(p_i^{\otimes m-1})$$

and

$$\begin{aligned} (I \otimes \widehat{W} \otimes \widehat{W})(\widehat{Q}) &= \frac{1}{(2m-1)!} \frac{1}{n} \sum_{i=1}^n \sum_{\sigma \in S_{2m-1}} X_{i,\sigma(1)} \otimes \widehat{W}(X_{i,\sigma(2)} \otimes \cdots \otimes X_{i,\sigma(m)}) \otimes \cdots \\ &\quad \widehat{W}(X_{i,\sigma(m+1)} \otimes \cdots \otimes X_{i,\sigma(2m-1)}). \end{aligned}$$

Since the transform in Lemma A.2 is unitary, we have that

$$(117) \quad \left\| T - \widehat{T} \right\|_{\mathcal{H}\mathcal{S}} = \left\| (I \otimes W \otimes W)(Q) - (I \otimes \widehat{W} \otimes \widehat{W})(\widehat{Q}) \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}.$$

We will now show that  $\left\| T - \widehat{T} \right\| \xrightarrow{P} 0$ .

$$(118) \quad \left\| T - \widehat{T} \right\| \leq \left\| T - \widehat{T} \right\|_{\mathcal{H}\mathcal{S}}$$

$$(119) \quad = \left\| (I \otimes W \otimes W)(Q) - (I \otimes \widehat{W} \otimes \widehat{W})(\widehat{Q}) \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}$$

$$(120) \quad \leq \left\| (I \otimes W \otimes W)(Q) - (I \otimes W \otimes W)(\widehat{Q}) \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}$$

$$(121) \quad + \left\| (I \otimes W \otimes W)(\widehat{Q}) - (I \otimes \widehat{W} \otimes \widehat{W})(\widehat{Q}) \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}$$

$$(122) \quad \leq \|I \otimes W \otimes W\| \left\| Q - \widehat{Q} \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}$$

$$(123) \quad + \left\| I \otimes W \otimes W - I \otimes \widehat{W} \otimes \widehat{W} \right\| \left\| \widehat{Q} \right\|_{\ell^2(\mathbb{R}^d)^{\otimes 2m-1}}.$$

We have that  $\mathbb{E}[\widehat{Q}] = Q$  so the first summand goes to zero in probability by the law of large numbers and Lemma C.2. All we need to show is that  $\left\| I \otimes W \otimes W - I \otimes \widehat{W} \otimes \widehat{W} \right\| \xrightarrow{P} 0$ .

The following equality is mentioned in [3].

LEMMA D.1. *Let  $U_1, \dots, U_n$  be defined as in Lemma A.1. Then*

$$(124) \quad \|U_1 \otimes \dots \otimes U_n\| = \|U_1\| \|U_2\| \dots \|U_n\|.$$

From Lemma D.1 we have that

$$(125) \quad \left\| I \otimes W \otimes W - I \otimes \widehat{W} \otimes \widehat{W} \right\|$$

$$(126) \quad \leq \|I\| \left\| W \otimes W - \widehat{W} \otimes \widehat{W} \right\|$$

$$(127) \quad = \left\| W \otimes W - \widehat{W} \otimes \widehat{W} \right\|$$

$$(128) \quad \leq \left\| W \otimes W - W \otimes \widehat{W} \right\| + \left\| W \otimes \widehat{W} - \widehat{W} \otimes \widehat{W} \right\|$$

$$(129) \quad = \|W\| \left\| W - \widehat{W} \right\| + \left\| \widehat{W} \right\| \left\| W - \widehat{W} \right\|$$

$$(130) \quad = \left( \|W\| + \left\| \widehat{W} \right\| \right) \left\| W - \widehat{W} \right\|.$$

The left factor converges in probability to  $2\|W\|$  and the right factor converges to 0 in probability and so we have that  $\|T - \hat{T}\| \xrightarrow{p} 0$ . From this we also have that  $\|\hat{T}\hat{T}^H - TT^H\| \xrightarrow{p} 0$ .  $\square$

#### APPENDIX E: ADDITIONAL EXPERIMENTAL DETAILS

For the “random dominating measure” algorithm, the random dominating measure was generated using the square of iid Gaussian random variables with mean 0 and standard deviation 0.03. We used the Gaussian random variables instead of the uniform distribution described in Section B simply because the Gaussian random measure performed better. Finally we made one minor adjustment to the algorithm proposed in the main text. If an estimator yielded a component which has a negative entry, we simply set the negative entry to zero and renormalise.

## REFERENCES

- [1] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008.
- [2] Gerald B. Folland. *Real analysis: modern techniques and their applications*. Pure and applied mathematics. Wiley, 1999.
- [3] R.V. Kadison and J.R. Ringrose. *Fundamentals of the theory of operator algebras. V1: Elementary theory*. Pure and Applied Mathematics. Elsevier Science, 1983.
- [4] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer New York, 2002.
- [5] N. Young. *An Introduction to Hilbert Space*. Cambridge mathematical textbooks. Cambridge University Press, 1988.

FACHBEREICH INFORMATIK  
TU KAISERSLAUTERN  
POSTFACH 3049  
67653 KAISERSLAUTERN  
GERMANY  
E-MAIL: [vandermeulen@cs.uni-kl.de](mailto:vandermeulen@cs.uni-kl.de)

EECS BUILDING  
1301 BEAL AVENUE  
ANN ARBOR MI, 48109  
E-MAIL: [clayscot@umich.edu](mailto:clayscot@umich.edu)