HAL
open science

Convex geometry of finite exchangeable laws and de Finetti style representation with universal correlated corrections

Guillaume Carlier, Gero Friesecke, Daniela Vögler

To cite this version:

Guillaume Carlier, Gero Friesecke, Daniela Vögler. Convex geometry of finite exchangeable laws and de Finetti style representation with universal correlated corrections. 2021. hal-03504025

HAL Id: hal-03504025

https://hal.science/hal-03504025v1

Preprint submitted on 28 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Convex geometry of finite exchangeable laws and de Finetti style representation with universal correlated corrections

Guillaume Carlier* Gero Friesecke† Daniela Vögler‡

June 16, 2021

Abstract

We present a novel analogue for finite exchangeable sequences of the de Finetti, Hewitt and Savage theorem and investigate its implications for multi-marginal optimal transport (MMOT) and Bayesian statistics. If $(Z_1, ..., Z_N)$ is a finitely exchangeable sequence of N random variables taking values in some Polish space X, we show that the law μ_k of the first k components has a representation of the form

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda)\, \mathrm{d}\alpha(\lambda)$$

for some probability measure α on the set of $\frac{1}{N}$-quantized probability measures on X and certain universal polynomials $F_{N,k}$. The latter consist of a leading term $N^{k-1}/\prod_{j=1}^{k-1}(N-j)\lambda^{\otimes k}$ and a finite, exponentially decaying series of correlated corrections of order N^{-j} $(j = 1, ..., k)$. The $F_{N,k}(\lambda)$ are precisely the extremal such laws, expressed via an explicit polynomial formula in terms of their one-point marginals λ. Applications include novel approximations of MMOT via polynomial convexification and the identification of the remainder which is estimated in the celebrated error bound of Diaconis-Freedman [11] between finite and infinite exchangeable laws.

Keywords: N-representability, finite exchangeability, de Finetti, Hewitt-Savage theorem, multi-marginal optimal transport, Bayesian statistics, extremal measures, Choquet theory

*CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

†Faculty of Mathematics, Technische Universität München, gf@ma.tum.de

‡Faculty of Mathematics, Technische Universität München, voegler@ma.tum.de

1 Introduction

Multi-marginal optimal transport (MMOT) has attracted a great deal of attention in recent years. The relevance of MMOT to tackle challenging problems arising from electronic density functional theory was established in [7, 5]. In this context, one has to find the joint density of N electrons with fixed one-point marginal so as to minimize a total repulsive Coulombian cost. Even though the problem is difficult for large N, it is symmetric (invariant under permutations of the electrons) and only depends on the two-point marginal of the joint law of the N electrons (2-body interaction). Whether symmetries and few-body interactions are helpful to analyze such MMOT problems is a natural question. An interesting result from [8] relying on the fact that the Coulomb potential has a positive Fourier transform and the de Finetti, Hewitt and Savage theorem is that when one lets N go to $+\infty$, the optimal plan is the independent (infinite product) measure. This is in striking contrast with the more standard two-marginal optimal transport where, for typical costs including the Coulomb cost, optimal plans are sparse and concentrate on low-dimensional subsets of the product space [4, 17, 7]. The present paper is motivated by MMOT for a possibly large but finite number of marginals N and symmetric k-body (with $k \leq N$) interaction cost. We present a novel explicit analogue of the de Finetti, Hewitt and Savage theorem and investigate its implications for such problems. We also briefly indicate implications for Bayesian statistics.

The main technical novelty in our work is the construction of an explicit polynomial inverse of the marginal map from extremal N-representable k-point probability measures (see below for terminology) to 1-point probability measures. This extends previous results for 2-point [16] and 3-point [26] measures on finite state spaces to arbitrary k and general Polish spaces.

Our ensuing finite version of de Finetti yields a complete, finite, exponentially decaying series of correlated corrections which need to be added to the independent measure in the case of finite N. This explicitly identifies the remainder estimated in the celebrated error bound of Diaconis-Freedman [11] between finite and infinite exchangeable laws.

In the remainder of this introduction we first recall the celebrated de Finetti-Hewitt-Savage theorem, then describe in more detail what changes in the finite exchangeable case.

De Finetti-Hewitt-Savage. Recall that a sequence $(Z_i)_{i\in\mathbb{N}}$ of random variables taking values in a Polish space X is called exchangeable if the law of $(Z_1, Z_2, ...)$ equals that of $(Z_{\sigma(1)}, Z_{\sigma(2)}, ...)$ for each finite permutation σ of $\mathbb{N}$, that is each permutation which leaves all but finitely many elements unchanged. The de Finetti-Hewitt-Savage theorem says that any such sequence

is a convex mixture of i.i.d. sequences. In other words, the law of $(Z_1, Z_2, ...)$ is a convex combination of independent measures,

$$\mu = \int_{\mathcal{P}(X)} \lambda^{\otimes\infty} d\alpha(\lambda) \tag{1.1}$$

for some probability measure (or prior) α on the set $\mathcal{P}(X)$ of probability measures on X. In Bayesian language, this says that the general infinite exchangeable sequence (Z_i) is obtained by first picking some distribution λ on X at random from some prior, then taking (Z_i) to be i.i.d. with distribution λ. For comprehensive reviews of exchangeability we refer the reader to Aldous [1] and Kallenberg [23].

Finite exchangeability; finite extendibility. A sequence $(Z_1, ..., Z_N)$ is called finitely exchangeable if its law equals that of $(Z_{\sigma(1)}, ..., Z_{\sigma(N)})$ for any permutation σ of $\{1, ..., N\}$. For $k \leq N$, a sequence $(Z_1, ..., Z_k)$ is called finitely extendible if its law equals that of the first k elements of some finitely exchangeable sequence $(\tilde{Z}_1, ..., \tilde{Z}_N)$.[1]

For finite exchangeable sequences $(Z_1, ..., Z_N)$ it is well-known that the analogous representation to (1.1) with $\lambda^{\otimes\infty}$ replaced by $\lambda^{\otimes N}$ does not hold, see Diaconis [12],Diaconis and Freedman [11], Jaynes [22]; the error is known to be of order $\frac{1}{N}$ in total variation [11, 3].

The main approach for describing finite exchangeable sequences which has been introduced in the probability literature is to write such a sequence as a superposition of i.i.d. sequences but drop the requirement that the superposition of the laws be convex, i.e. allows signed measures α in (1.1), see Dellacherie and Meyer [10], Jaynes [22], Kerns and Székely [25], Janson, Konstantopoulos and Yuan [21]. For applications, this approach has limited appeal, for two reasons. First, the signed measure representation is not unique. Second, the superposition does not yield a probability measure for an arbitrary signed α, but remaining within probability measures is mandatory for recovering an exchangeable law by sampling (see below) and for our application to MMOT.

A very interesting second picture of finite exchangeability which appears not to have received the attention it deserves can be found in Kerns and Székely [25] and Kallenberg [23]. Namely, finite exchangeable sequences are convex mixtures of "urn sequences", or equivalently, finitely exchangeable laws are convex superpositions of symmetrized Dirac measures, the latter being the laws of urn sequences (as described further below). See Kerns and Székely ([25], top of p.600), where such a representation appears as an

[1] Analogously, $(Z_1, ..., Z_k)$ is called infinitely extendible if its law equals that of the first k elements of an infinite exchangeable sequence $(\tilde{Z}_1, \tilde{Z}_2, ...)$.

intermediate step in the proof of the signed-measure representation. The laws of urn sequences are known to be the extreme points of the convex set of finitely exchangeable laws, see [25] for an elementary proof for finite state spaces and Kallenberg ([23] Proposition 1.8) for a general proof using advanced probability methods. Although not given in [23], the Kerns-Székely representation could be deduced from the statement of Proposition 1.8 via disintegration of measures.

The present work builds upon this picture, which turns out to be very useful for the applications we have in mind. Thus we view finite exchangeable laws as convex mixtures of urns. But we re-instate the idea from original de Finetti, kept in the signed-measure approach, that the parameter space of the superposition should consist of probability measures on the *original* Polish space X, not its N-fold product. In principle, this is possible by parametrizing urn laws by their one-point marginals, which are easily seen to be in 1-1 correspondence with these laws. In practice, to arrive at an explicit representation one needs an explicit formula for the inverse of this marginal map. By deriving such a formula, we obtain a unique representation of finitely exchangeable laws which sheds some light on their universal correlation structure and is useful for applications.

Main results. In terms of laws, N-extendibility turns into what has been called N-representability [15]: for $k \leq N$, a k-point probability measure μ_k on X^k, or k-plan for short, is called N-representable if it is the k-point marginal of a symmetric N-point probability measure μ_N on X^N (see Definition 2.1).

As a first main result, we explicitly determine the extremal N-representable k-plans, that is, those that cannot be written as strict convex combinations of any other N-representable k-plans, and give a polynomial parametrization in terms of their one-point marginals. Focusing in this introduction for simplicity on the case $k = 4$, these are the probability measures

$$F_{N,4}(\lambda) = \frac{N^3}{(N-1)(N-2)(N-3)} \Bigg[\lambda^{\otimes 4} - \frac{6}{N} S_4 \, \mathrm{id}_{\#}^{\otimes 2} \lambda \otimes \lambda^{\otimes 2} + \frac{8 \, S_4 \, \mathrm{id}_{\#}^{\otimes 3} \lambda \otimes \lambda + 3 \, S_4 \, \mathrm{id}_{\#}^{\otimes 2} \lambda \otimes \mathrm{id}_{\#}^{\otimes 2} \lambda}{N^2} - \frac{6}{N^3} \, \mathrm{id}_{\#}^{\otimes 4} \lambda \Bigg]$$

where λ is a $\frac{1}{N}$-quantized probability measure on X, i.e. an empirical measure of the form $\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ for some – not necessarily distinct – points $x_i \in X$. It is not obvious, but part of our result, that these measures are nonnegative, and different for different λ.

The above expression can be viewed as a degree-4 symmetric polynomial in λ. Besides an overall positive prefactor, the polynomial has leading term

$\lambda^{\otimes 4}$ which is homogeneous of degree 4 and uncorrelated, and alternating corrections of order $\frac{1}{N^j}$ which are homogeneous of degree 4–j and more and more strongly correlated. As N tends to infinity $F_{N,4}(\lambda)$ approaches the independent measure $\lambda^{\otimes 4}$, recovering the basis in the de Finetti representation for infinitely representable 4-plans implied by (1.1). The correlated corrections are of significant size even when N is quite large; see Figure 1. All these findings persist for general k; see Theorem 4.5 for the general expression $F_{N,k}(\lambda)$ for extremal N-representable k-plans. Qualitatively, the corrections to independence form a finite exponentially decaying series; quantitatively the (rational) coefficients which appear can be related to the analytic continuation of the Ewens function from genetics, which we introduce for this purpose.

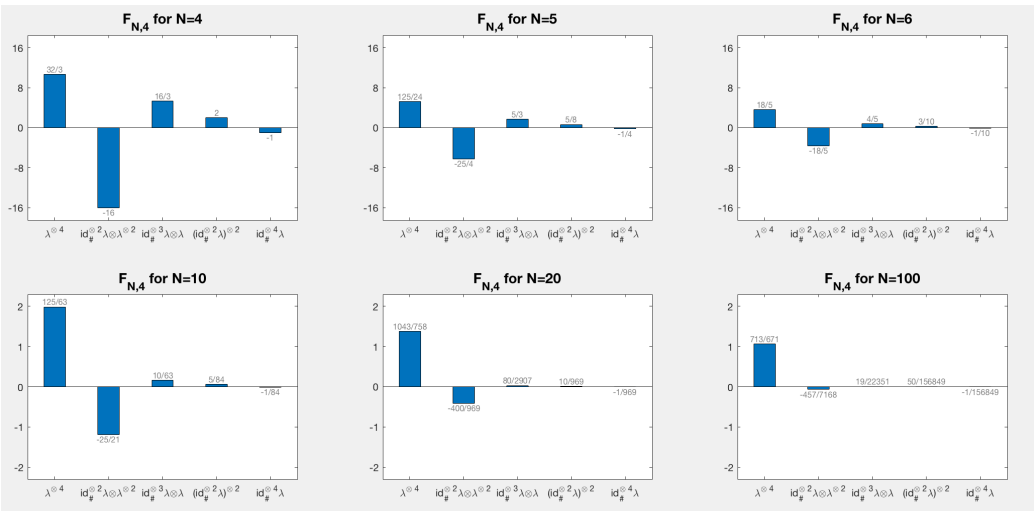

Figure 1: Coefficients of the universal polynomial $F_{N,4}$ for different N. For $N = 5$ and 6, the second (correlated) term is bigger respectively equal in absolute value to the first (independent) term; for $N = 20$ its size is about 30% that of the first term. For large N, $F_{N,4}$ converges to the independent measure $\lambda^{\otimes 4}$, but even for $N = 100$ the deviation from the latter is still visible.

Our second contribution is to cast the abstract insight [25, 23] that finite exchangeables are convex mixtures of urn sequences into a quantitative polynomial formula. We show that any N-representable k-plan is a convex mixture of the $F_{N,k}(\lambda)$. More precisely, a k-plan is N-representable if and

only if it is of the form

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda)\, \mathrm{d}\alpha(\lambda) \tag{1.2}$$

for some probability measure α on the set $\mathcal{P}_{\frac{1}{N}}(X)$ of $\frac{1}{N}$-quantized probability measures on X. Moreover if $N = k$ the measure α is unique, giving a one-to-one parametrization of the laws of finitely exchangeable sequences. By contrast, the signed measure representation of such laws is not unique [21], and not all signed measures give rise to such a law.

Formula (1.2) generalizes de Finetti-Hewitt-Savage, (1.1), from infinitely to finitely representable measures, or from infinite to finite exchangeable sequences of random variables. Formally, in the limit $N \to \infty$ the domain of integration in (1.2) tends to all of $\mathcal{P}(X)$, and the integrand tends to the independent measure $\lambda^{\otimes k}$, recovering de Finetti (see Section 6 for a rigorous account). In Bayesian language, formula (1.2) says that the general N-extendible sequence $(Z_1, ..., Z_k)$ of X-valued random variables is obtained by first picking some $\frac{1}{N}$-quantized distribution λ on X at random from some prior, then sampling $(Z_1, ..., Z_k)$ from the correlated distribution $F_{N,k}(\lambda)$. In particular, by setting $k{=}N$ we conclude that *the general finite exchangeable sequence* $(Z_1, ..., Z_N)$ *is obtained by picking* λ *at random from its – in the case* $k{=}N$ *unique – prior* α *in* (1.2)*, then sampling* $(Z_1, ..., Z_N)$ *from* $F_{N,N}(\lambda)$.

Sampling from $F_{N,k}(\lambda)$ has a transparent probabilistic meaning which we now explain in the language of urns. Write a given $\frac{1}{N}$-quantized probability measure λ as $\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$ for N not necessarily distinct points $x_1, ..., x_N \in X$. Now pick, in turn, k of these points at random *without replacement*, and denote the so-obtained sequence by $(Z_1, ..., Z_k)$. By construction, the law of this sequence is the k-point marginal μ_k of the symmetrization of the Dirac measure $\delta_{(x_1,...,x_N)}$ on X^N. But the polynomial $F_{N,k}$ is precisely constructed as the inverse of the marginal map $\mu_k \mapsto \lambda$ (see eq. (4.6), eq. (4.8), and Theorem 4.5 below). Hence $F_{N,k}(\lambda) = \mu_k$, and so $(Z_1, ..., Z_k)$ is the sought-after finite N-extendible sequence. We find it quite remarkable that the extremal N-representable k-plans $F_{N,k}(\lambda)$ – which emerge purely from convex geometric considerations – have such a simple probabilistic meaning, being the laws of classical examples [1] of finite exchangeable sequences which are not infinitely extendible.

Recovering the prior from sampling. A nice aspect of our representation of the law of a general finitely exchangeable sequence $(Z_1, ..., Z_N)$ (eq. (1.2) with $k = N$) is that the prior α, which is unique when $k = N$, can be determined by sampling, as follows. Let

$$\left(Z_1^{(\nu)}, ..., Z_N^{(\nu)}\right)_{\nu=1}^{n}$$

be a sequence of n independent samples in X^N. Form the $\mathcal{P}_{\frac{1}{N}}(X)$-valued sequence

$$\lambda^{(\nu)} = \frac{1}{N}\sum_{i=1}^{N} \delta_{Z_i^{(\nu)}}.$$

Then the empirical measure

$$\frac{1}{n}\sum_{\nu=1}^{n} \delta_{\lambda^{(\nu)}}$$

converges almost surely to α. See Corollary 5.2.

The paper is organized as follows. After introducing some notations and preliminaries in Section 2, Section 3 deals with $\frac{1}{N}$-quantized measures. In Section 4, we focus on the finite case; we first recall the results of [16] and then identify the universal correlated polynomials $F_{N,k}(\lambda)$. Section 5 extends these findings to the case of a Polish state space X, and show in addition that the $F_{N,k}(\lambda)$ are in fact exposed N-representable k-plans. Section 6 discusses connections with the Hewitt and Savage theorem and the Diaconis-Freedman error bounds. Section 7 gives an unexpected connection with the Ewens sampling formula from genetics. Finally, Section 8 is devoted to applications to MMOT emphasizing the connection with convexification of polynomials.

2 Preliminaries and notations

In the sequel X will denote a Polish (i.e., complete and separable metric) space. The principal example we have in mind is $X = \mathbb{R}^d$, in which case all of our results are already new and interesting. In this case the metric is the usual Euclidean metric $d(x, y) = |x - y| = (\sum_{i=1}^{d}(x_i - y_i)^2))^{1/2}$. We denote by $\mathcal{P}(X)$ the set of Borel probability measures on X. Probability measures on X^k will be called k-plans. From now on, we fix two integers k and N with $1 \leq k \leq N$. Given $\gamma \in \mathcal{P}(X^N)$ we denote by $M_k\gamma$ the k-point-marginal of γ, i.e.,

$$(M_k\gamma)(A) := \gamma(A \times X^{N-k}) \text{ for every Borel subset } A \text{ of } X^k \tag{2.1}$$

(with the convention $M_N\gamma = \gamma$).

We denote by $C_b(X^N)$ the space of bounded and continuous functions on X^N, and by $\mathcal{S}_N$ the group of permutations of $\{1, \ldots, N\}$. For $\gamma \in \mathcal{P}(X^N)$ and $\sigma \in \mathcal{S}_N$, the measure $\gamma^\sigma \in \mathcal{P}(X^N)$ is defined by

$$\int_{X^N} \varphi \mathrm{d}\gamma^\sigma = \int_{X^N} \varphi(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) \mathrm{d}\gamma(x_1, \ldots, x_N)$$

for every test-function $\varphi \in C_b(X^N)$. A measure $\gamma \in \mathcal{P}(X^N)$ is called symmetric if $\gamma = \gamma^\sigma$ for every $\sigma \in \mathcal{S}_N$. If $\gamma \in \mathcal{P}(X^N)$ is arbitrary, its symmetrization $S_N\gamma$ is given by

$$S_N\gamma := \frac{1}{N!} \sum_{\sigma \in S_N} \gamma^\sigma. \tag{2.2}$$

The symmetrization operator $S_N : \gamma \mapsto S_N\gamma$ is a linear projection operator on $\mathcal{P}(X^N)$, i.e. it maps $\mathcal{P}(X^N)$ linearly into itself and satisfies $(S_N)^2 = S_N$; and γ is symmetric if and only if $S_N\gamma = \gamma$. The set of symmetric N-plans is denoted by $\mathcal{P}_{\text{sym}}(X^N)$:

$$\mathcal{P}_{\text{sym}}(X^N) := \{\gamma \in \mathcal{P}(X^N) \ : \ \gamma = S_N\gamma\}. \tag{2.3}$$

We shall use the notation # to denote the push-forward measure, that is, given two Polish spaces Y and Z, a Borel map T from Y to Z and a Borel probability measure μ on Y, then $T_\#\mu$ is the probability measure on Z defined by $T_\#\mu(B) = \mu(T^{-1}(B))$ for every Borel subset B of Z; equivalently, for every real-valued, continuous and bounded function φ on Z:

$$\int_Z \varphi \, \mathrm{d}T_\#\mu = \int_Y \varphi \circ T \ \ \mathrm{d}\mu.$$

We recall the following definition from [15].

Definition 2.1. *For $N \in \mathbb{N}$ and $k \in \{1, ..., N\}$, a k-plan $\mu_k \in \mathcal{P}(X^k)$ is said to be N-representable if it is the k-point marginal of a symmetric N-plan, that is to say if there exists $\gamma \in \mathcal{P}_{\text{sym}}(X^N)$ such that $\mu_k = M_k\gamma$. We denote by $\mathcal{P}_{N-\text{rep}}(X^k)$ the set of N-representable k-plans, i.e.:*

$$\mathcal{P}_{N-\text{rep}}(X^k) = \{M_k S_N \tilde{\gamma} \ : \ \tilde{\gamma} \in \mathcal{P}(X^N)\} = \{M_k\gamma \ : \ \gamma \in \mathcal{P}_{\text{sym}}(X^N)\}.$$

In probabilistic terms, a symmetric N-plan $\gamma \in \mathcal{P}_{\text{sym}}(X^N)$ is the law of a finite exchangeable random sequence $(Z_1, \ldots, Z_N)$ with values in X^N, whereas $\mu_k = M_k\gamma \in \mathcal{P}_{N-\text{rep}}(X^k)$ is the law of its first k-components $(Z_1, \ldots, Z_k)$.

We will work with the following standard notion of convergence in $\mathcal{P}(X)$ (as well as $\mathcal{P}(X^k)$, $\mathcal{P}(X^N)$, ...). Recall that $C_b(X)$ denotes the space of bounded continuous functions on X.

Definition 2.2. *A sequence $(\mu_\nu)_{\nu \in \mathbb{N}}$ of probability measures in $\mathcal{P}(X)$ is said to converge narrowly to $\mu \in \mathcal{P}(X)$ if*

$$\lim_{\nu \to \infty} \int_X \varphi d\mu_\nu = \int_X \varphi d\mu \ \ \textit{for all } \varphi \in C_b(X).$$

Thus, in applications to statistical physics where the probability measures live on a space of particle configurations, narrow convergence corresponds to convergence of bounded continuous observables.

We note the following basic topological property of the set of N-representable k-plans.

Lemma 2.3. *The set $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is closed under narrow convergence.*

Proof Let $\{\mu_\nu\}_{\nu\in\mathbb{N}}$ be a narrowly convergent sequence in $\mathcal{P}_{N-\mathrm{rep}}(X^k)$. Write μ_ν as $M_k\gamma_\nu$ for $\gamma_\nu \in \mathcal{P}_{\mathrm{sym}}(X^N)$. Since $\mu_\nu = M_k\gamma_\nu$ is narrowly convergent, it is tight and hence so is γ_ν. By Prokhorov's theorem, γ_ν has a narrowly convergent subsequence which converges to some $\gamma \in \mathcal{P}_{\mathrm{sym}}(X^N)$. Since M_k is narrowly continuous, μ_ν converges to $M_k\gamma \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$.

We recall that narrow convergence on $\mathcal{P}(X)$ is metrizable. For instance one may start from the metric d on X, truncate it to the (topologically equivalent) bounded metric $\tilde{d}(x,y) := \min\{d(x,y),1\}$, and use the associated 1-Wasserstein metric on $\mathcal{P}(X)$

$$W_1(\lambda_1,\lambda_2) := \inf_{\theta\in\Pi(\lambda_1,\lambda_2)} \Big\{ \int_{X\times X} \tilde{d}(x,y)\ \mathrm{d}\theta(x,y) \Big\} \text{ for all } (\lambda_1,\lambda_2) \in \mathcal{P}(X)^2, \tag{2.4}$$

where $\Pi(\lambda_1,\lambda_2)$ is the set of transport plans between λ_1 and λ_2, i.e., the set of Borel probability measures on $X\times X$ having λ_1 and λ_2 as marginals. Then the (bounded) metric W_1 metrizes narrow convergence on $\mathcal{P}(X)$ (that is, μ_ν converges narrowly to μ if and only if $W_1(\mu_\nu,\mu)$ tends to zero) and $(\mathcal{P}(X), W_1)$ is itself a Polish space.

Also we recall the definition of the total variation distance between two signed measures μ and ν on X:

$$\|\mu-\nu\|_{\mathrm{TV}} := \sup\Big\{ |\mu(A)-\nu(A)| \ :\ A \text{ Borel subset of } X \Big\}. \tag{2.5}$$

3 $\frac{1}{\mathrm{N}}$-quantized probability measures

An important role will be played by the set of $\frac{1}{N}$-quantized probability measures on the Polish space X,

$$\mathcal{P}_{\frac{1}{N}}(X) := \left\{ \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \ :\ x_1,\dots,x_N \in X^N \text{ (not necessarily distinct)} \right\}. \tag{3.1}$$

It is easy to see that this set can also be written as

$$\mathcal{P}_{\frac{1}{N}}(X) = \left\{ \lambda \in \mathcal{P}(X) \ : \ \lambda(A) \in \left\{ 0, \frac{1}{N}, ..., 1 \right\} \text{ for every Borel subset } A \text{ of } X \right\}. \tag{3.2}$$

In the special case of finite state spaces X, this set was introduced – and utilized to parametrize extremal N-representable measures – in [16]. Let us collect two basic properties of this set which hold for general state spaces.

Lemma 3.1. *(Quadratic constraint characterization and closedness of $\frac{1}{N}$-quantized probability measures)*
a) $\lambda \in \mathcal{P}(X)$ belongs to $\mathcal{P}_{\frac{1}{N}}(X)$ if and only if

$$\lambda \otimes \lambda - \frac{2k+1}{N} \operatorname{id}_{\#}^{\otimes 2} \lambda \geq -\frac{k(k+1)}{N^2} \text{ , for } k = 0, \ldots, N-1. \tag{3.3}$$

b) $\mathcal{P}_{\frac{1}{N}}(X)$ is closed under narrow convergence.

In formula (3.3) and in the sequel, $\lambda^{\otimes \ell}$ denotes the ℓ-fold tensor product of λ with itself and $\operatorname{id}_{\#}^{\otimes \ell} \lambda$ is defined by

$$\int_{X^\ell} \varphi \, \mathrm{d} \operatorname{id}_{\#}^{\otimes \ell} \lambda := \int_X \varphi(x, \ldots, x) \mathrm{d}\lambda(x), \quad \text{for all } \varphi \in C_b(X^\ell).$$

Proof a): The nontrivial implication is that (3.3) implies that $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$. Let A be any Borel subset of X. Applying the measure on the left hand side of (3.3) to $A \times A$ gives $f(\lambda(A)) \geq 0$, where f is the scalar function $f(t) = t^2 - \frac{2k+1}{N} t + \frac{k(k+1)}{N^2} = (t - \frac{k}{N})(t - \frac{k+1}{N})$. But f is negative precisely in the open interval $(\frac{k}{N}, \frac{k+1}{N})$, whence $\lambda(A)$ does not lie in this interval. Since this holds for all $k = 0, ..., N-1$, it follows that λ belongs to the set (3.2).
b): The maps $\lambda \in \mathcal{P}(X) \mapsto \lambda^{\otimes 2} \in \mathcal{P}(X^2)$ and $\lambda \in \mathcal{P}(X) \mapsto \operatorname{id}_{\#}^{\otimes 2} \lambda \in \mathcal{P}(X^2)$ are continuous with respect to narrow convergence, hence so is the left hand side of (3.3). The assertion now follows from a).

4 Extremal N-representable k-plans on finite state spaces

Throughout this section, we assume $N \geq 2$ and restrict our attention to a finite state space X consisting of ℓ distinct points,

$$X = \{a_1, \ldots, a_\ell\}. \tag{4.1}$$

4.1 Extreme points

Our goal is to describe the geometry of the convex set of N-representable k-plans on the finite state space X, i.e., $\mathcal{P}_{N-\mathrm{rep}}(X^k)$. This set is a compact polyhedron in a finite-dimensional vector space and therefore coincides, by Minkowski's theorem (see e.g. [20]), with the convex hull of its extreme points. Therefore, classifying the extreme points is one way to characterize the geometry of the object. We recall that a point x in a convex set K is an extreme point if, whenever $x = \alpha x_1 + (1-\alpha)x_2$ for some x_1, $x_2 \in K$ and some $\alpha \in (0,1)$, we have that $x_1 = x_2 = x$. The set of extreme points of K will be denoted $\operatorname{ext} K$.

In [16] the extremal N-representable k-plans are determined in the case $k = 2$ and $k = N$. They correspond exactly to the symmetrized Dirac measures respectively their two-point marginals:

Theorem 4.1. [16] a) *A measure μ on X^N is an extreme point of $\mathcal{P}_{\mathrm{sym}}(X^N)$ if and only if it is of the form*

$$S_N \delta_{a_{i_1},\ldots,a_{i_N}} \qquad \textit{for some } 1 \le i_1 \le \ldots \le i_N \le \ell. \tag{4.2}$$

Moreover different index vectors $(i_1, ..., i_N)$ with $1 \le i_1 \le \ldots \le i_N \le \ell$ yield different extreme points.

b) A measure μ on X^2 is an extreme point of $\mathcal{P}_{N-\mathrm{rep}}(X^2)$ if and only if it is of the form

$$M_2 S_N \delta_{a_{i_1},\ldots,a_{i_N}} \qquad \textit{for some } 1 \le i_1 \le \ldots \le i_N \le \ell. \tag{4.3}$$

c) Moreover the marginal maps $M_2 : \operatorname{ext}(\mathcal{P}_{\mathrm{sym}}(X^N)) \to \operatorname{ext}(\mathcal{P}_{N-\mathrm{rep}}(X^2))$ and $M_1 : \operatorname{ext}(\mathcal{P}_{\mathrm{sym}}(X^N)) \to \mathcal{P}_{\frac{1}{N}}(X)$ are bijections.

Here a) and the fact that the set of measures in (4.3) contains the set $\operatorname{ext}(\mathcal{P}_{N-\mathrm{rep}}(X^2))$ of extremal N-representable two-plans is easy to see, but the reverse inclusion and the bijectivity of M_2 between extremal symmetric N-plans and extremal N-representable two-plans is nontrivial; geometrically it says that none of the corners of the high-dimensional polytope $\mathcal{P}_{\mathrm{sym}}(X^N)$ is mapped into the interior (or face interior or edge interior) of the low-dimensional polytope $\mathcal{P}_{N-\mathrm{rep}}(X^2)$ by the highly non-injective marginal map M_2. Using this nontrivial fact it is easy to extend Theorem 4.1 to an arbitrary choice of $k \in \{2, \ldots, N\}$.

Theorem 4.2. *A measure μ on X^k is an extreme point of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ if and only if it is of the form*

$$M_k S_N \delta_{a_{i_1},\ldots,a_{i_N}} \qquad \textit{for some } 1 \le i_1 \le \ldots \le i_N \le \ell. \tag{4.4}$$

Moreover the marginal map $M_k : \text{ext}\,(\mathcal{P}_{\text{sym}}(X^N)) \to \text{ext}\,(\mathcal{P}_{N-\text{rep}}(X^k))$ *is a bijection.*

Proof. We will abbreviate $\mathcal{P}_{\text{sym}}(X^N) = \mathcal{P}_{\text{sym}}$, $(i_1, \ldots, i_N) = i$, $\{(i_1, \ldots, i_N) : 1 \leq i_1 \leq \ldots \leq i_N \leq \ell\} = \mathcal{I}$. By the definition of N-representability, $\mathcal{P}_{N-\text{rep}}(X^k) = M_k \mathcal{P}_{\text{sym}}$. Using, in order of appearance, this fact, the linearity of M_k, and Theorem 4.1 a), we have

$$\text{ext}\,(\mathcal{P}_{N-\text{rep}}(X^k)) = \text{ext}\,(M_k \mathcal{P}_{\text{sym}}) \subseteq M_k \text{ext}\,(\mathcal{P}_{\text{sym}}) = \{M_k S_N \delta_{a_{i_1},\ldots,a_{i_N}} \,:\, i \in \mathcal{I}\}. \tag{4.5}$$

To establish the reverse inclusion it suffices to show that the number of elements of the set on the left is bigger or equal that of the set on the right. By the fact that $M_2 = M_2 M_k$ and the linearity of M_2,

$$\text{ext}\,(M_2 \mathcal{P}_{\text{sym}}) = \text{ext}\,(M_2 M_k \mathcal{P}_{\text{sym}}) \subseteq M_2 \text{ext}\,(M_k \mathcal{P}_{\text{sym}}) = M_2 \text{ext}\,(\mathcal{P}_{N-\text{rep}}(X^k))$$

and consequently $|\text{ext}\,(M_2 \mathcal{P}_{\text{sym}})| \leq |\text{ext}\,(\mathcal{P}_{N-\text{rep}}(X^k))|$, where $|\cdot|$ denotes the number of elements of a set. Combining this inequality with the bijectivity property of M_2 in Theorem 4.1 b) and Theorem 4.1 a) yields

$$\begin{aligned} |\text{ext}\,(\mathcal{P}_{N-\text{rep}}(X^k))| &\geq |\text{ext}\,(M_2 \mathcal{P}_{\text{sym}})| = |\text{ext}\,(\mathcal{P}_{\text{sym}})| = |\{S_N \delta_{a_{i_1},\ldots,a_{i_N}} \colon i \in \mathcal{I}\}| \\ &\geq |\{M_k S_N \delta_{a_{i_1},\ldots,a_{i_N}} \colon i \in \mathcal{I}\}|. \end{aligned}$$

□

In [16] it was established that for the set (4.1) consisting of ℓ distinct points, the cardinality of $\mathcal{P}_{\frac{1}{N}}(X)$ – and hence, by Theorem 4.1 b), the number of extreme points of $\mathcal{P}_{\text{sym}}(X^N)$ – equals $\binom{N+\ell-1}{\ell-1}$. Now the following corollary is an immediate consequence of Theorem 4.2.

Corollary 4.3. *For any* $k \in \{2, \ldots, N\}$, $\mathcal{P}_{N-\text{rep}}(X^k)$ *has* $\binom{N+\ell-1}{\ell-1}$ *extreme points.*

Combining the isomorphisms M_k and M_1 from Theorems 4.2 respectively 4.1 shows that the extreme points of $\mathcal{P}_{N-\text{rep}}(X^k)$, i.e. the k-plans of form (4.4), can be uniquely recovered from their one-point marginals $\frac{1}{N}(\delta_{a_{i_1}} + \ldots + \delta_{a_{i_N}}) \in \mathcal{P}_{\frac{1}{N}}(X)$. But the above abstract reasoning does not provide a convenient formula for the recovery map. This issue is dealt with in the next section.

4.2 A universal polynomial formula for extreme points in terms of their one-point marginals

Our aim now is to derive an explicit polynomial formula for the extremal measures (4.4) in terms of their one-point marginals. In order to do so we consider any extremal N-representable k-plan

$$\mu_k := M_k S_N \delta_{x_1,\ldots,x_N} \tag{4.6}$$

for $(x_1, \ldots, x_N) \in X^N$.

In [16] it was shown that, in the case of $k = 2$, μ_2 can be expressed explicitly as

$$\mu_2 = \frac{N}{N-1}\lambda^{\otimes 2} - \frac{1}{N-1}\,\mathrm{id}_{\#}^{\otimes 2}\,\lambda, \tag{4.7}$$

where

$$\lambda := M_1\mu_k = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i} \in \mathcal{P}_{\frac{1}{N}}(X) \tag{4.8}$$

is the one-point marginal of μ_k. (Recall the notation $\lambda^{\otimes\ell}$ and $\mathrm{id}_{\#}^{\otimes\ell}\,\lambda$ for the ℓ-fold tensor product of λ with itself respectively the push-forward of λ under the ℓ-fold cartesian product of the identity; see the end of Section 2.) A similar computation for $k = 3$ gives

$$\mu_3 = \frac{N^2}{(N-1)(N-2)}\left[\lambda^{\otimes 3} - \frac{3}{N}S_3\Big((\mathrm{id}_{\#}^{\otimes 2}\,\lambda)\otimes\lambda\Big) + \frac{2}{N^2}\,\mathrm{id}_{\#}^{\otimes 3}\,\lambda\right]. \tag{4.9}$$

(For a justification of (4.9) using our general results see the examples below Theorem 4.5.) In view of (4.7) and (4.9), it is natural to look for a similar *polynomial* of degree k in λ expression of μ_k, consisting of a mean field term $\lambda^{\otimes k}$ and corrections of order $\frac{1}{N^j}$ for $j = 1, \ldots, k-1$. As turns out, the jth order correction is related to the partitions of the number j.

Definition 4.4. *Let $\mathbb{N} = \{1, 2, 3, ...\}$ denote the set of positive integers. A partition of $j \in \mathbb{N}$ of length $n \in \mathbb{N}$ is a vector $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{N}^n$ such that $\sum_{i=1}^{n} p_i = j$, $p_1 \geq \ldots \geq p_n$. For any partition $\mathbf{p}$ we denote its length by $n(\mathbf{p})$.*

For example, the partitions of 4 are

$$\begin{aligned}
&1+1+1+1\\
&2+1+1\\
&2+2\\
&3+1\\
&4.
\end{aligned}$$

This corresponds in the above notation to $\mathbf{p} = (1,1,1,1) \in \mathbb{N}^4$, $\mathbf{p} = (2,1,1) \in \mathbb{N}^3$, $\mathbf{p} = (2,2) \in \mathbb{N}^2$, $\mathbf{p} = (3,1) \in \mathbb{N}^2$, and $\mathbf{p} = 4 \in \mathbb{N}$.

Theorem 4.5. *Let $N \geq 2$, $k \in \{2, \ldots, N\}$. Any extremal N-representable k-plan μ_k (see (4.6)) can be written in terms of its one-point marginal λ (see (4.8)) as*

$$\mu_k = \frac{N^{k-1}}{\prod_{i=1}^{k-1}(N-i)} \left[\lambda^{\otimes k} + \sum_{j=1}^{k-1} \frac{(-1)^j}{N^j} S_k P_j^{(k)}(\lambda) \right] =: F_{N,k}(\lambda) \tag{4.10}$$

where for $j = 1, ..., k-1$

$$P_j^{(k)}(\lambda) = \sum_{\substack{\mathbf{p}=(p_1,\ldots p_{n(\mathbf{p})}) partition \\ of\ j with\ j+n(\mathbf{p})\leq k}} d_{\mathbf{p}}^{(k)}\ \mathrm{id}_{\#}^{\otimes(p_1+1)}\lambda\ \otimes \ldots \otimes \mathrm{id}_{\#}^{\otimes(p_{n(\mathbf{p})}+1)}\lambda\ \otimes \lambda^{\otimes(k-j-n(\mathbf{p}))} \tag{4.11}$$

with positive coefficients $d_{\mathbf{p}}^{(k)}$ given by

$$d_{\mathbf{p}}^{(k)} = \frac{k!}{(k-j-n(\mathbf{p}))!} \prod_{i=1}^{n(\mathbf{p})} \frac{1}{p_i+1} \prod_{q \in Ran\,\mathbf{p}} \frac{1}{(|\mathbf{p}^{-1}(q)|)!} \quad . \tag{4.12}$$

Moreover the coefficients satisfy the sum rule

$$\sum_{\substack{\mathbf{p}\ partition\ of\ j \\ with\ j+n(\mathbf{p})\leq k}} d_{\mathbf{p}}^{(k)} = \sum_{1\leq i_1<\ldots<i_j\leq k-1} i_1 \cdot \ldots \cdot i_j \quad =: c_j^{(k)} \quad (j = 1, \ldots, k-1). \tag{4.13}$$

In particular, $||P_j^{(k)}(\lambda)||_{TV} = \int_{X^k} dP_j^{(k)}(\lambda) = c_j^{(k)}$.

In the last term in eq. (4.12), a partition $\mathbf{p}$ is viewed as a map from the set of its component indices to $\mathbb{N}$; the range $\mathrm{Ran}\,\mathbf{p}$ of this map is the set of values taken by the components, and $|\mathbf{p}^{-1}(q)|$ denotes the number of components with value q. For example, for $\mathbf{p} = (3,1,1)$ and $q = 1$, $|\mathbf{p}^{-1}(q)| = 2$. The factor $(|\mathbf{p}^{-1}(q)|)!$ in the denominator says that a partition with many repeat components contributes much less than a partition with few repeat components.

Some remarks are in order.

1) The first term in expression (4.10) for the extreme points is a mean field term and the remaining terms are correlation corrections. We emphasize that the $P_j^{(k)}$ are independent of N and hence the correlation corrections form a *finite series* in inverse powers of N.

2) The $P_j^{(k)}$ are polynomials of degree $k-j$ in λ.

3) Due to the presence of the signs $(-1)^j$, it is far from trivial that $F_{N,k}(\lambda)$ is a nonnegative measure; but it must be, e.g. because the left hand side of (4.10) equals (4.6). Nonnegativity relies on a subtle interplay between the explicit coefficients in Theorem 4.5 and the quantization condition $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, and does not hold for arbitrary $\lambda \in \mathcal{P}(X)$.

4) The coefficients $c_j^{(k)} = \sum i_1 \cdot \ldots \cdot i_j$ introduced in (4.13) which measure the total mass of the jth-order correction to independence are related to the well-known Stirling numbers, particularly the (absolute) Stirling numbers of the first kind. For given natural numbers $q, r \in \mathbb{N} \cup \{0\}$ with $r \leq q$ the corresponding absolute Stirling number of the first kind $s(q,r)$ gives the number of permutations of $\{1, \ldots, q\}$ that decompose into r cycles, with the convention that $s(q,r)$ is zero when exactly one of q and r is zero and that $s(0,0) = 1$. From well-known expressions for Stirling numbers one can see that the following holds

$$c_j^{(k)} = s(k, k-j),$$

that is to say the present coefficients $c_j^{(k)}$ equal the number of permutations of $\{1, \ldots, k\}$ that decompose into $k-j$ cycles. For more information about Stirling numbers we refer the interested reader to [6].

5) By combining Theorem 4.5, Theorem 4.2, and the isomorphism property of M_1 from Theorem 4.1, we immediately obtain:

Corollary 4.6. *A measure μ_k on X^k is an extreme point of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ if and only if it is of the form $\mu_k = F_{N,k}(\lambda)$ for some $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, with $F_{N,k}$ given by (4.10)–(4.12). Moreover $F_{N,k}(\lambda)$ has one-point marginal λ, and j-point marginal $F_{N,j}(\lambda)$ for $2 \leq j \leq k-1$.*

We now write out the universal polynomials $F_{N,k}$ explicitly for small k.

Example: k=2. The finite sum over j in (4.10) reduces to a single term for $j = 1$, $-\frac{1}{N} S_2 P_1^{(2)}(\lambda)$, and $P_1^{(2)}(\lambda)$ consists of a single term associated with the only partition $\mathbf{p} = 1$ of 1, $id_{\#}^{\otimes 2}\lambda$. Consequently

$$F_{N,2}(\lambda) = \frac{N}{N-1}\left[\lambda^{\otimes 2} - \frac{1}{N}\,\mathrm{id}_{\#}^{\otimes 2}\,\lambda\right].$$

This expression agrees with (4.7), and so Theorem 4.5 recovers [16] Theorem 2.1.

Example: k=3. The finite series in (4.10) runs from $j = 1$ to $j = 2$, and for these two values of j, the partitions contributing to the sum in (4.11) are the partitions $\mathbf{p}$ of j satisfying $j + n(\mathbf{p}) \leq 3$. These partitions and the associated coefficient $d_{\mathbf{p}}$ given by (4.12) are

j	Partitions of j with $j + n(\mathbf{p}) \leq 3$	our notation: $\mathbf{p} =$	coefficient $d_{\mathbf{p}} =$
1	1	1	3
2	2	2	2

and consequently

$$F_{N,3}(\lambda) = \frac{N^2}{(N-1)(N-2)}\left[\lambda^{\otimes 3} - \frac{3}{N} S_3\Big((\mathrm{id}_{\#}^{\otimes 2}\,\lambda) \otimes \lambda\Big) + \frac{2}{N^2}\,\mathrm{id}_{\#}^{\otimes 3}\,\lambda\right].$$

Example: k=4. By formulae (4.10)–(4.12), the partitions $\mathbf{p}$ of j contributing to the jth order correction are:

j	Partitions of j with $j+n(\mathbf{p})\leq 4$	our notation: $\mathbf{p}=$	coefficient $d_{\mathbf{p}}=$
1	1	1	6
2	2	2	8
	1+1	(1,1)	3
3	3	3	6

and consequently

$$F_{N,4}(\lambda)=\frac{N^3}{(N-1)(N-2)(N-3)}\left[\lambda^{\otimes 4}-\frac{6}{N}S_4\,\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\lambda^{\otimes 2}+\frac{8\,S_4\,\mathrm{id}_{\#}^{\otimes 3}\lambda\otimes\lambda+3\,S_4\,\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\mathrm{id}_{\#}^{\otimes 2}\lambda}{N^2}\right.$$
$$\left.-\frac{6}{N^3}\,\mathrm{id}_{\#}^{\otimes 4}\lambda\right].$$

Example: k=5. The contributing partitions are

j	Partitions of j with $j+n(\mathbf{p})\leq 5$	our notation: $\mathbf{p}=$	coefficient $d_{\mathbf{p}}=$
1	1	1	10
2	2	2	20
	1+1	(1,1)	15
3	3	3	30
	2+1	(2,1)	20
4	4	4	24

and so

$$F_{N,5}(\lambda)=\frac{N^4}{(N-1)(N-2)(N-3)(N-4)}\left[\lambda^{\otimes 5}-\frac{10}{N}S_5\,\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\lambda^{\otimes 3}+\frac{20\,S_5\,\mathrm{id}_{\#}^{\otimes 3}\lambda\otimes\lambda^{\otimes 2}+15\,S_5\,\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\lambda}{N^2}\right.$$
$$\left.-\frac{30\,S_5\,\mathrm{id}_{\#}^{\otimes 4}\lambda\otimes\lambda+20\,S_5\,\mathrm{id}_{\#}^{\otimes 3}\lambda\otimes\mathrm{id}_{\#}^{\otimes 2}\lambda}{N^3}+\frac{24}{N^4}\,\mathrm{id}_{\#}^{\otimes 5}\lambda\right].$$

Example: j=1 and j=k-1. In these cases only one partition of j satisfies $j+n(\mathbf{p})\leq k$ and formula (4.12) for the coefficient $d_{\mathbf{p}}$ becomes particularly simple:

j	Partitions of j with $j+n(\mathbf{p})\leq k$	our notation: $\mathbf{p}=$	coefficient $d_{\mathbf{p}}=$
1	1	1	$\frac{k(k-1)}{2}$
k-1	k-1	k-1	(k-1)!

It follows that the polynomials describing the first-order respectively order-(k-1) contribution to $F_{N,k}(\lambda)$ are

$$P_1^{(k)}(\lambda)=\frac{k(k-1)}{2}\,\mathrm{id}_{\#}^{\otimes 2}\lambda\otimes\lambda^{\otimes(k-2)},\qquad P_{k-1}^{(k)}(\lambda)=(k-1)!\,\mathrm{id}_{\#}^{\otimes k}\lambda.$$

We now discuss the error when truncating the finite series in (4.10). Retaining only the mean-field term gives

$$F_{N,k}(\lambda)=\frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)}\left(\lambda^{\otimes k}+\varepsilon_{N,k}(\lambda)\right)\ \text{with}\ \|\varepsilon_{N,k}(\lambda)\|_{\mathrm{TV}}\leq\frac{C_k}{N}\tag{4.14}$$

and keeping the first p correction terms ($p\in\{1,...,k-2\}$) we have

$$F_{N,k}(\lambda)=\frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)}\left[\lambda^{\otimes k}+\sum_{j=1}^{p}\frac{(-1)^j}{N^j}S_k\,P_j^{(k)}(\lambda)+\varepsilon_{N,k,p}(\lambda)\right]$$
$$\text{with}\ \|\varepsilon_{N,k,p}(\lambda)\|_{\mathrm{TV}}\leq\frac{C_k}{N^{p+1}},\tag{4.15}$$

with constants C_k independent of N and p. For example, to give explicit values,

$$C_k = \sum_{j=1}^{k-1} c_j^{(k)} \tag{4.16}$$

will do. Moreover the coefficients $d_{\mathbf{p}}^{(k)}$ and hence the C_k are independent of the size ℓ of the finite state space. Thus, for k fixed and any $N \geq 2$, retaining only the first p correlation terms captures the extreme points up to an error which decreases *exponentially* in p, the rate being *uniform* in the size of the finite state space and *improving* logarithmically with N.

Before proving Theorem 4.5 let us give a quick heuristic derivation of the formulae for the coefficients $c_j^{(k)} = \int dP_j^{(k)}(\lambda)$ which give the total mass of the jth order correction to independence for extremal N-representable k-plans. Expressions (4.7) and (4.9) suggest to try the ansatz

$$\mu_k = \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \left[\lambda^{\otimes k} + \sum_{j=1}^{k-1} \frac{(-1)^j}{N^j} c_j^{(k)} \nu_j \right] \tag{4.17}$$

with normalized measures ν_j (i.e. $\int d\nu_j = 1$) and a priori unknown but N-independent coefficients c_j. Consider for example $k = 4$. Integrating over X^k, using that μ_k and the ν_j are normalized, and multiplying both sides by the product $\prod_{j=1}^{k-1}(N-j)$ gives

$$N^3(1 - \frac{c_1}{N} + \frac{c_2}{N^2} - \frac{c_3}{N^3}) = (N-1)(N-2)(N-3). \tag{4.18}$$

Expanding the right hand side into powers of N gives

$$(N-1)(N-2)(N-3) = N^3 - (1+2+3)N^2 + (1\cdot 2 + 1\cdot 3 + 2\cdot 3)N - (1\cdot 2\cdot 3)$$

so equating coefficients yields $c_1 = \sum_{1\leq i\leq 3} i\,(=6)$, $c_2 = \sum_{1\leq i<j\leq 3} ij\,(=11)$, $c_3 = \sum_{1\leq i<j<k\leq 3} ijk\,(=6)$, i.e. the asserted formulae for the c_j. Extending this heuristic argument to general k is straightforward.

Of course this argument is not a proof because it rests on the (as yet unjustified) ansatz (4.17) with N-independent coefficients. This ansatz is a corollary of the more detailed result (4.10)–(4.13) to whose proof we now turn.

We begin by eliminating the high-dimensional space $\mathcal{P}_{\mathrm{sym}}(X^N)$ which appears in (4.6).

Lemma 4.7. *Any extremal N-representable k-plan μ_k given by* (4.6) *can be written as*

$$\mu_k = \frac{(N-k)!}{N!} \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k,\\ \textit{pairwise distinct}}} \delta_{x_{m_1}\dots x_{m_k}}. \tag{4.19}$$

Proof. Proceeding as in [16], more specifically the proof of Lemma 2.1, we rewrite μ_k by plugging in the definition of the symmetrization operator S_N and conditioning the sum over all permutations $\sigma : \{1,...,N\} \to \{1,...,N\}$ on the values on the first k integers:

$$\begin{aligned} N!\mu_k &= M_k \sum_{\sigma\in S_N} \delta_{x_{\sigma(1)}\dots x_{\sigma(N)}} \\ &= M_k \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k \\ \text{pairwise distinct}}} \sum_{\substack{\sigma\in S_N \\ \sigma(1)=m_1,\dots,\sigma(k)=m_k}} \delta_{x_{m_1}\dots x_{m_k} x_{\sigma(k+1)}\dots x_{\sigma(N)}} \\ &= (N-k)! \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k \\ \text{pairwise distinct}}} \delta_{x_{m_1}\dots x_{m_k}}. \end{aligned}$$

□

To establish Theorem 4.5 we will proceed by induction over k. The next lemma gives a deceptively simple recursion formula for extremal N-representable k-plans. Just like (4.19), it hides the inverse power series structure (4.10) and the combinatorial complexity of the coefficients (4.12) by expanding the symmetric plan μ_k in a non-symmetric basis of delta functions, leading to many terms with identical symmetrization.

Lemma 4.8. *Let $N \geq 2$, and consider the k-plans μ_k $(k = 2,...,N)$ defined by* (4.6) *for fixed $(x_1,...,x_N) \in X^N$. Then for $k = 1,...,N-1$ and $\lambda = M_1\mu_k$,*

$$\mu_{k+1} = \frac{N}{N-k}\mu_k \otimes \lambda - \frac{1}{N-k}\sum_{j=1}^{k} R_{j\#}\mu_k \tag{4.20}$$

where $R_j : X^k \to X^{k+1}$ is given by $R_j(z_1,\dots,z_k) := (z_1,\dots,z_k,z_j)$.

Proof. We observe that $(m_1,\dots,m_k,m_{k+1}) \in \{1,...,N\}^{k+1}$ has pairwise distinct components if and only if $(m_1,...,m_k)$ has pairwise distinct components

and $m_{k+1} \in \{1, \dots, N\} \setminus \{m_1, \dots, m_k\}$. So by inclusion-exclusion we get, using $(N-(k+1))! = (N-k)!/(N-k)$,

$$\begin{aligned}
\mu_{k+1} &= \frac{(N-(k+1))!}{N!} \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k \\ \text{pairwise distinct}}} \sum_{m_{k+1}\in\{1,\dots,N\}\setminus\{m_1,\dots,m_k\}} \delta_{x_{m_1}\dots x_{m_{k+1}}} \\
&= \frac{(N-k)!}{N!\,(N-k)} \Big(\sum_{i=1}^{N} \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k \\ \text{pairwise distinct}}} \delta_{x_{m_1}\dots x_{m_k} x_i} - \sum_{\substack{(m_1,\dots,m_k)\in\{1,\dots,N\}^k \\ \text{pairwise distinct}}} \sum_{j=1}^{k} \delta_{x_{m_1}\dots x_{m_k} x_{m_j}} \Big) \\
&= \frac{N}{N-k} \mu_k \otimes \lambda - \frac{1}{N-k} \sum_{j=1}^{k} R_{j\#} \mu_k .
\end{aligned}$$

□

Next we derive a non-recursive formula in terms of *set partitions* of $\{1, ..., k\}$. To state it we need to introduce some notation. Recall that a set partition of $\{1, ..., k\}$ is a set $\mathcal{P}$ of pairwise disjoint nonempty subsets of $\{1, ..., k\}$ (called blocks) whose union equals $\{1, ..., k\}$. The set of all such partitions will be denoted Part_k. For a partition $\mathcal{P} \in \mathrm{Part}_k$, we denote by $n(\mathcal{P})$ the cardinality of $\mathcal{P}$, so that $\mathcal{P} = \{P_1, \dots, P_{n(\mathcal{P})}\}$ for some $P_i \subseteq \{1, ..., k\}$, and introduce the combinatorial factor

$$\beta_{\mathcal{P}} := \prod_{i=1}^{n(\mathcal{P})} (|P_i| - 1)! = \prod_{P\in\mathcal{P}} (|P| - 1)! \,. \tag{4.21}$$

Next, each partition $\mathcal{P}$ induces a certain natural mapping $G_{\mathcal{P}} : \mathcal{P}(X) \to \mathcal{P}(X^k)$. Informally, this mapping pushes, for each block P of the partition $\mathcal{P}$, a tensor factor $\lambda \in \mathcal{P}(X)$ forward onto the diagonal of those cartesian factors $X_{i_1}, ..., X_{i_{|P|}}$ of the product space X^k whose indices belong to P. More precisely, if $\mathcal{P} = \{P_1, \dots, P_n\}$, define $G_{\mathcal{P}}(\lambda)$ by

$$G_{\mathcal{P}}(\lambda)(A_1 \times ... \times A_k) = \prod_{P\in\mathcal{P}} \Big(\mathrm{id}_{\#}^{\otimes |P|} \lambda \Big) \Big(\prod_{i\in P} A_i \Big) \quad \text{for any } A_1, ..., A_k \subseteq X. \tag{4.22}$$

For instance if $k = 4$ and $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$, $G_{\mathcal{P}}(\lambda) = (\mathrm{id}_{\#}^{\otimes 2} \lambda) \otimes (\mathrm{id}_{\#}^{\otimes 2} \lambda)$, whereas if $k = 5$ and $\mathcal{P} = \{\{1, 3\}, \{2, 4, 5\}\}$ then $G_{\mathcal{P}}(\lambda)$ is defined by

$$\int_{X^5} \varphi \;\, \mathrm{d}G_{\mathcal{P}}(\lambda) = \int_{X^2} \varphi(x, y, x, y, y) \, \mathrm{d}\lambda(x) \, \mathrm{d}\lambda(y), \quad \text{for all } \varphi \in C_b(X^5).$$

We then have the following representation formula.

Proposition 4.9. *Let μ_k be defined by* (4.6)*, and let λ be its one-point marginal (4.8). Then*

$$\mu_k = \frac{(N-k)!}{N!} \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k-n(\mathcal{P})} N^{n(\mathcal{P})} \beta_{\mathcal{P}} G_{\mathcal{P}}(\lambda). \tag{4.23}$$

Proof. For $k=1$ the assertion is obvious, and for $k=2$ it easily follows from (4.7). Let us assume that (4.23) holds for $k \le N-1$. By (4.20)

$$\mu_{k+1} = \tfrac{(N-(k+1))!}{N!}(A+B), \text{ with } A := \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k-n(\mathcal{P})} N^{n(\mathcal{P})+1} \beta_{\mathcal{P}} G_{\mathcal{P}}(\lambda) \otimes \lambda$$

$$\text{and } B := \sum_{j=1}^{k} \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k+1-n(\mathcal{P})} N^{n(\mathcal{P})} \beta_{\mathcal{P}} R_{j\#} G_{\mathcal{P}}(\lambda).$$

Now let us partition Part_{k+1} into the two subsets Part^a_{k+1} and its complement Part^b_{k+1} where Part^a_{k+1} consists of all partitions $\mathcal{P}'$ of $\{1,\dots,k+1\}$ for which the singleton $\{k+1\}$ belongs to $\mathcal{P}'$. Thus $\mathcal{P}' \in \mathrm{Part}^a_{k+1}$ if and only if it can be written as $\mathcal{P} \cup \{\{k+1\}\}$ with $\mathcal{P} \in \mathrm{Part}_k$. Note then that $n(\mathcal{P}') = n(\mathcal{P})+1$, $\beta_{\mathcal{P}} = \beta_{\mathcal{P}'}$ and $G_{\mathcal{P}'}(\lambda) = G_{\mathcal{P}}(\lambda) \otimes \lambda$. Therefore we have

$$\begin{aligned}
&\sum_{\mathcal{P}'\in\mathrm{Part}^a_{k+1}} (-1)^{k+1-n(\mathcal{P}')} N^{n(\mathcal{P}')} \beta_{\mathcal{P}'} G_{\mathcal{P}'}(\lambda) \\
&= \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k-n(\mathcal{P})} N^{n(\mathcal{P})+1} \beta_{\mathcal{P}} G_{\mathcal{P}}(\lambda) \otimes \lambda \\
&= A.
\end{aligned}$$

Partitions $\mathcal{P}'$ in Part^b_{k+1} are those for which $k+1$ does not form a singleton in $\mathcal{P}'$. This is the same as saying that the following map a^+_{k+1} from $\{(\mathcal{P},P) : \mathcal{P} \in \mathrm{Part}_k,\ P \in \mathcal{P}\}$ to Part^b_{k+1} is a bijection: $a^+_{k+1}(\{P_1,...,P_n\},\ P_i) := \mathcal{P}' := \{P_1,...,P_i \cup \{k+1\},...,P_n\}$ $(i=1,...,n)$, or – in label-free notation –

$$a^+_{k+1}(\mathcal{P},P) := \{P \cup \{k+1\}\} \cup \{Q \in \mathcal{P} \,:\, Q \neq P\}.$$

We chose the notation a^+_{k+1} to emphasize the analogy with creation operators in quantum theory: the map a^+_{k+1} "creates" an extra entry $k+1$ in some block of the partition. Note that if $\mathcal{P}' = a^+_{k+1}(\mathcal{P},P)$, then $n(\mathcal{P}) = n(\mathcal{P}')$, $\beta_{\mathcal{P}'} = \beta_{\mathcal{P}} \cdot |P|$, and $G_{\mathcal{P}'}(\lambda) = R_{j\#} G_{\mathcal{P}}(\lambda)$ for every $j \in P$, so that

$$\beta_{\mathcal{P}'} G_{\mathcal{P}'}(\lambda) = \beta_{\mathcal{P}} \sum_{j\in P} R_{j\#} G_{\mathcal{P}}(\lambda).$$

We thus have

$$\begin{aligned}
&\sum_{\mathcal{P}'\in\mathrm{Part}^b_{k+1}} (-1)^{k+1-n(\mathcal{P}')} N^{n(\mathcal{P}')} \beta_{\mathcal{P}'} G_{\mathcal{P}'}(\lambda) \\
&= \sum_{\mathcal{P}\in\mathrm{Part}_k} \sum_{P\in\mathcal{P}} (-1)^{k+1-n(\mathcal{P})} N^{n(\mathcal{P})} \beta_{\mathcal{P}} \sum_{j\in P} R_{j\,\#} G_{\mathcal{P}}(\lambda) \\
&= \sum_{j=1}^{k} \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k+1-n(\mathcal{P})} N^{n(\mathcal{P})} \beta_{\mathcal{P}} \Big(\sum_{P\in\mathcal{P}\,:\,j\in P} 1 \Big) R_{j\,\#} G_{\mathcal{P}}(\lambda) \\
&= \sum_{j=1}^{k} \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k+1-n(\mathcal{P})} N^{n(\mathcal{P})} \beta_{\mathcal{P}} R_{j\,\#} G_{\mathcal{P}}(\lambda) \\
&= B,
\end{aligned}$$

which gives the desired expression for $\mu_{k+1} = \frac{(N-(k+1))!}{N!}(A+B)$.

□

It remains to match the unwieldy-to-evaluate expression (4.23) with the more explicit expansion (4.10), (4.11), (4.12).

We begin by dealing with the fact that expression (4.23) contains many terms with identical symmetrization. Since μ_k is symmetric, applying the symmetrization operator S_k to both sides gives

$$\mu_k = \sum_{\mathcal{P}\in\mathrm{Part}_k} (-1)^{k-n(\mathcal{P})} \frac{N^{n(\mathcal{P})}}{N\cdot(N-1)\cdot\ldots\cdot(N-k+1)}\, \beta_{\mathcal{P}}\, S_k\, G_{\mathcal{P}}(\lambda). \tag{4.24}$$

If $\mathcal{P}\in\mathrm{Part}_k$, then according to (4.22) $G_{\mathcal{P}}(\lambda)$ pushes $n(\mathcal{P})$ factors λ onto the k cartesian factors of the product space X^k. The different $S_k G_{\mathcal{P}}(\lambda)$'s which can arise from such a set partition are in bijective correspondence to the partitions $\mathbf{p}'$ of the number k, via

$$\mathbf{p}' = (p'_1, \ldots, p'_m) \longmapsto S_k\, \mathrm{id}_{\#}^{\otimes p'_1}\lambda \otimes \ldots \otimes \mathrm{id}_{\#}^{\otimes p'_m}\lambda. \tag{4.25}$$

The set partitions $\mathcal{P}\in\mathrm{Part}_k$ satisfying $S_k G_{\mathcal{P}}(\lambda) =$ r.h.s. of (4.25) for some given partition $\mathbf{p}' = (p'_1, \ldots, p'_m)$ of k are precisely those consisting of m sets $P_1, \ldots, P_m$ with cardinalities $|P_i| = p'_i$ for all i. Let us denote their totality by $\mathrm{Part}_k(\mathbf{p}')$. A canonical set partition in $\mathrm{Part}_k(\mathbf{p}')$ is

$$\begin{aligned}
&\mathcal{P}(\mathbf{p}') = \Big\{ P_1, \ldots, P_m \Big\} \text{ with} \\
&P_1 = \{1, \ldots, p'_1\},\ P_2 = \{p'_1+1, \ldots, p'_1+p'_2\}, \ldots, \\
&P_m = \{p'_1 + \ldots + p'_{m-1}, \ldots, \underbrace{p'_1 + \ldots + p'_m}_{=k}\}.
\end{aligned}$$

For this set partition, as well as any other $\mathcal{P} \in \mathrm{Part}_k(\mathbf{p}')$,

$$n(\mathcal{P}) = m, \quad \beta_{\mathcal{P}} = (p'_1 - 1)! \cdot ... \cdot (p'_m - 1)!. \tag{4.26}$$

To reduce (4.23) to a sum over partitions $\mathbf{p}'$ of the number k, it remains to determine the number $\mathcal{N}(\mathbf{p}')$ of set partitions belonging to $\mathrm{Part}_k(\mathbf{p}')$. Let us fix any partition $\mathbf{p}' = (p'_1, ..., p'_m)$ of k. First of all we note that

$$\mathcal{N}(\mathbf{p}') = \frac{1}{\prod\limits_{q \in \mathrm{Ran}\, \mathbf{p}'} (|\mathbf{p}'^{-1}(q)|)!} \mathcal{N}' \tag{4.27}$$

where $\mathcal{N}'$ is the number of set partitions corresponding to $\mathbf{p}'$ endowed with an ordering of the blocks such that larger blocks come before smaller ones, $\mathcal{N}' = |\{(P'_1, ..., P'_m) \,:\, |P'_i| = p'_i \text{ for all } i\}|$. Here the combinatorial factor in the denominator accounts for the fact that any group of b equal-sized blocks in a set partition admits $b!$ orderings. But the number $\mathcal{N}'$ is straightforward to compute: choosing P'_1 means choosing p'_1 numbers out of k, so there are $\binom{k}{p'_1}$ choices; given P'_1, choosing P'_2 means choosing p'_2 numbers out of the remaining $k - p'_1$ numbers, yielding $\binom{k-p'_1}{p'_2}$ choices; and so on. It follows that

$$\mathcal{N}' = \binom{k}{p'_1} \cdot \binom{k - p'_1}{p'_2} \cdot ... \cdot \binom{k - (p'_1 + ... + p'_{m-1})}{p'_m} = \frac{k!}{p'_1! \cdot ... \cdot p'_m!}. \tag{4.28}$$

Combining (4.24), the bijectivity of the map (4.25), (4.26), (4.27), and (4.28) yields:

Proposition 4.10. *Let μ_k be any extremal N-representable k-plan (see* (4.6)*), and let λ be its one-point marginal (4.8). Then*

$$\mu_k = \sum_{p' \text{ partition of } k} c_{\mathbf{p}'} S_k \, \mathrm{id}_{\#}^{\otimes p'_1} \lambda \otimes ... \otimes \mathrm{id}_{\#}^{\otimes p'_{n(\mathbf{p}')}} \lambda \text{ with}$$

$$c_{\mathbf{p}'} = (-1)^{k - n(\mathbf{p}')} \frac{k!}{N(N-1)...(N-k+1)} N^{n(\mathbf{p}')} \frac{1}{p'_1 \cdot ... \cdot p'_{n(\mathbf{p}')}} \cdot \frac{1}{\prod\limits_{q \in Ran\, \mathbf{p}'} (|\mathbf{p}'^{-1}(q)|)!}. \tag{4.29}$$

The expression for μ_k in (4.29) may be taken as an alternative definition of the polynomial $F_{N,k}(\lambda)$ introduced in Theorem 4.5.

4.3 End of the proof of Theorem 4.5

The last step in the proof of Theorem 4.5 is to match the above expression with the series given in the theorem. We would like to decompose (4.29) into

terms of order $\frac{1}{N^j}$ (times the overall order 1 prefactor in (4.10)). To this end, we re-write the N-dependent factors in (4.29) as

$$\frac{1}{N\cdot(N-1)\cdot\ldots\cdot(N-k+1)}N^{n(\mathbf{p}')}=\frac{N^{k-1}}{(N-1)\cdot\ldots\cdot(N-k+1)}\,\frac{1}{N^{k-n(\mathbf{p}')}}.$$

This together with the fact that the only partition $\mathbf{p}'$ of k with $k-n(\mathbf{p}')=0$ is $\mathbf{p}'=(1,\ldots,1)$, in which case the r.h.s. of (4.25) is $\lambda^{\otimes k}$ and $\frac{1}{p_1'\cdot\ldots\cdot p_m'}=1$, shows that

$$\mu_k=\frac{N^{k-1}}{(N-1)\ldots(N-k+1)}\Big(\lambda^{\otimes k}+\sum_{j=1}^{k-1}\frac{(-1)^j}{N^j}\sum_{\substack{\mathbf{p}'\text{ partition of }k\\ \text{with }n(\mathbf{p}')=k-j}}\frac{k!}{p_1'\cdot\ldots\cdot p_{n(\mathbf{p}')}'}\cdot \frac{1}{\prod_{q\in\operatorname{Ran}\mathbf{p}'}(|\mathbf{p}'^{-1}(q)|)!}S_k\,\mathrm{id}_{\#}^{\otimes p_1'}\lambda\otimes\ldots\otimes\mathrm{id}_{\#}^{\otimes p_{n(\mathbf{p}')}'}\lambda\Big). \tag{4.30}$$

The partitions $\mathbf{p}'$ of k of length $n(\mathbf{p}')=k-j$ are in bijective correspondence to the partitions $\mathbf{p}$ of the number j with length $n(\mathbf{p})\leq k-j$, via

$$\begin{aligned}&\mathbf{p}=(p_1,\ldots,p_n)\text{ partition of }j\text{ of length }n\leq k-j\\ &\longmapsto \mathbf{p}'=(p_1',\ldots,p_{k-j}')=\begin{cases}(p_1+1,\ldots,p_n+1) & \text{if } n=k-j\\ (p_1+1,\ldots,p_n+1,\underbrace{1,\ldots,1}_{k-j-n\text{ times}}) & \text{if } n<k-j.\end{cases}\end{aligned} \tag{4.31}$$

Moreover, for any two partitions $\mathbf{p}$, $\mathbf{p}'$ related by (4.31), we have

$$S_k\,\mathrm{id}_{\#}^{\otimes p_1'}\lambda\otimes\ldots\otimes\mathrm{id}_{\#}^{\otimes p_m'}\lambda=S_k\,\mathrm{id}_{\#}^{\otimes p_1+1}\lambda\otimes\ldots\otimes\mathrm{id}_{\#}^{\otimes p_n+1}\lambda\otimes\lambda^{\otimes k-(j+n)} \tag{4.32}$$

and

$$p_1'\cdot\ldots\cdot p_{n(\mathbf{p}')}'=(p_1+1)\cdot\ldots\cdot(p_n+1), \tag{4.33}$$

$$\prod_{q\in\operatorname{Ran}\mathbf{p}'}(|\mathbf{p}'^{-1}(q)|)!=\prod_{q\in\operatorname{Ran}\mathbf{p}}(|\mathbf{p}^{-1}(q)|)!\cdot(k-(j+n))!, \tag{4.34}$$

with the last factor above accounting for the $k-(j+n)$ components with value 1 occurring in $\mathbf{p}'$. Combining eq. (4.30), the bijectivity of (4.31), and identities (4.32)–(4.34) yields the desired expression for μ_k in (4.10)-(4.11)-(4.12). Let us finally prove formula (4.13). Observe that the detailed representation

(4.10) implies (4.17), where each ν_j is a probability measure and the coefficients $c_j^{(k)}$ do not depend on N. Taking the total mass of each side of (4.17) we get that for every $N \geq k$, one has

$$\prod_{j=1}^{k-1}(N-j) = N^{k-1} + \sum_{j=1}^{k-1}(-1)^j c_j^{(k)} N^{k-j-1}$$

so that $(-1)^j c_j^{(k)}$ is the coefficient of X^{k-j-1} in the polynomial $\prod_{j=1}^{k-1}(X-j)$ which proves (4.13) thanks to Vieta's formulas.

5 Extreme points and integral representation of N-representable k-plans on continuous state spaces

Now we return to the case of a general state space, i.e. we just assume that X is a Polish space. Importantly, the results in this section cover the prototypical continuous state space $X = \mathbb{R}^d$.

Recall the set $\mathcal{P}_{\frac{1}{N}}(X)$ of $\frac{1}{N}$-quantized probability measures on X introduced in (3.1), (3.2).

Given $\lambda \in \mathcal{P}(X)$ we define $F_{N,k}(\lambda) \in \mathcal{P}(X^k)$ by formulae (4.10), (4.11), (4.12); we note that the expressions in these formulae make sense for general Polish spaces X and general (not necessarily $\frac{1}{N}$-quantized) $\lambda \in \mathcal{P}(X)$.

5.1 De Finetti style representation

We now state a de Finetti style representation result for N-representable k-plans.

Theorem 5.1. *Let $N \geq k \geq 2$. A measure $\mu_k \in \mathcal{P}(X^k)$ is N-representable if and only if there exists $\alpha \in \mathcal{P}(\mathcal{P}(X))$ such that $\alpha(\mathcal{P}_{\frac{1}{N}}(X)) = 1$ and*

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda)\, d\alpha(\lambda) \tag{5.1}$$

where $F_{N,k}$ is defined by (4.10)–(4.12). Moreover, if $k = N$, the measure α in (5.1) is unique.

The meaning of (5.1) is that for every test function $\varphi \in C_b(X^k)$,

$$\int_{X^k} \varphi\, d\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} \Phi(\lambda) d\alpha(\lambda),$$

where

$$\Phi(\lambda) := \int_{X^k} \varphi(x_1, ..., x_k) d\Big(F_{N,k}(\lambda)\Big)(x_1, ..., x_k).$$

Note that $\Phi : \lambda \mapsto \Phi(\lambda)$ is a continuous function on $\mathcal{P}(X)$ endowed with the narrow topology, by the narrow-to-narrow continuity of $F_{N,k}$. Moreover the sup norm of Φ is bounded by the sup norm of φ times the TV norm of $F_{N,k}(\lambda)$, the finiteness of the latter being clear from the definition. It follows that $\Phi \in C_b(\mathcal{P}(X))$, whence $\int_{\mathcal{P}_{\frac{1}{N}}(X)} \Phi(\lambda) d\alpha(\lambda)$ is well defined for any $\alpha \in \mathcal{P}(\mathcal{P}(X))$.

Theorem 5.1 generalizes the celebrated de Finetti-Hewitt-Savage theorem from infinitely representable to finitely representable measures, or - in probabilistic language - from infinitely to finitely extendible sequences of random variables. Formally, in the limit $N \to \infty$, the domain of integration $\mathcal{P}_{\frac{1}{N}}(X)$ in (5.1) approaches all of $\mathcal{P}(X)$ and the integrand tends to the independent measure $\lambda^{\otimes k}$, recovering de Finetti (see Section 6 for a rigorous account). Thus in the finitely representable case, the role of the independent measures $\lambda^{\otimes k}$ in de Finetti is taken by the universally correlated measures $F_{N,k}(\lambda)$, which contain corrections of order N^{-j} for $j = 1, ..., k$.

The additional assumption $k = N$ for uniqueness cannot be omitted, see the next section for a simple counterexample.

Proof. Recall that in a Polish space, finitely supported probability measures are dense with respect to narrow convergence. We know from Theorem 4.5 that when $\lambda := N^{-1} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}_{\frac{1}{N}}(X)$ then $F_{N,k}(\lambda) = M_k S_N \delta_{x_1 \dots x_N}$ and so $F_{N,k}(\lambda) \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$. By convexity of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, any measure of the form (5.1) with a finitely supported probability measure α on $\mathcal{P}_{\frac{1}{N}}(X)$ also belongs to $\mathcal{P}_{N-\mathrm{rep}}(X^k)$. Now if $\alpha \in \mathcal{P}(\mathcal{P}_{\frac{1}{N}}(X))$ is arbitrary and μ_k is given by (5.1), we approximate α by a sequence of finitely supported measures α^n. We now use that the map $\lambda \mapsto F_{N,k}(\lambda)$ is continuous under narrow convergence. This follows immediately from the definition (4.10)–(4.12) and the continuity of the maps $\lambda \mapsto \lambda^{\otimes j}$ and $\lambda \mapsto \mathrm{id}_{\#}^{\otimes j} \lambda$. Hence $\mu_k^n := \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda) \mathrm{d}\alpha^n(\lambda)$ converges narrowly to μ_k which therefore belongs to $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, since the latter is closed under narrow convergence (see Lemma 2.3).

Conversely, given $\gamma \in \mathcal{P}(X^N)$ and $\mu_k := M_k S_N \gamma \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$, let γ^n be a sequence of finitely supported probability measures which narrowly converges to γ, so that $\mu_k^n := M_k S_N \gamma^n$ converges to μ_k. Let us write $\gamma^n := \sum_{j \in J_n} \gamma_j^n \delta_{\mathbf{x}_j^n}$ where J_n is finite and $\mathbf{x}_j^n = (x_{1,j}^n, \dots x_{N,j}^n) \in X^N$. Using Theorem

4.5 again, we know that μ_k^n can be written as

$$\mu_k^n := \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda)\, \mathrm{d}\alpha^n(\lambda) \tag{5.2}$$

with

$$\alpha^n := \sum_{j\in J_n} \gamma_j^n \delta_{\Lambda(\mathbf{x}_j^n)} = \Lambda_\# \gamma^n \tag{5.3}$$

where the map $\Lambda : X^N \mapsto \mathcal{P}_{\frac{1}{N}}(X)$ is defined for all $\mathbf{x} := (x_1, \ldots, x_N) \in X^N$ by

$$\Lambda(\mathbf{x}) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}.$$

Note that Λ is Lipschitz with respect to W_1; in particular it is continuous from X^N to $\mathcal{P}(X)$ endowed with the narrow topology. Since γ^n converges narrowly to γ, α^n converges narrowly to $\alpha := \Lambda_\# \gamma$. Since $\mathcal{P}_{\frac{1}{N}}(X)$ is closed (see Lemma 3.1), it follows from the Portmanteau theorem (see, e.g., [2]) and the fact that $\alpha^n(\mathcal{P}_{\frac{1}{N}}(X)) = 1$ for every n that $\alpha(\mathcal{P}_{\frac{1}{N}}(X)) = 1$. (Here we use the following part of the Portmanteau theorem: if $C \subseteq \mathcal{P}(X)$ is closed and $\alpha^n \in \mathcal{P}(\mathcal{P}(X))$ converges narrowly to α, then $\alpha(C) \geq \limsup_{n\to\infty} \alpha^n(C)$.) Finally, thanks to the narrow continuity of $F_{N,k}$ we deduce the desired representation by integrating (5.2) against a test function and passing to the limit $n \to \infty$:

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda) \mathrm{d}\alpha(\lambda).$$

Finally, assume $k = N$ and that $\mu \in \mathcal{P}_{\mathrm{sym}}(X^N)$ can be written as

$$\mu = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,N}(\lambda) \mathrm{d}\alpha(\lambda). \tag{5.4}$$

Let $\psi : \mathcal{P}(X) \to \mathbb{R}$ be bounded and continuous for the narrow topology so that $\psi \circ \Lambda \in C_b(X^N)$. Let us now observe that if $\lambda = \Lambda(\mathbf{x}) \in \mathcal{P}_{\frac{1}{N}}(X)$ for some $\mathbf{x} = (x_1, \ldots, x_N) \in X^N$, then $S_N \delta_{\mathbf{x}} = F_{N,N}(\lambda)$ and since $\Lambda(\mathbf{x}) = \Lambda(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) = \lambda$, for every $\sigma \in \mathcal{S}_N$, we have

$$\int_{X^N} \psi \circ \Lambda \, \mathrm{d}F_{N,N}(\lambda) = \psi(\lambda).$$

Taking $\psi \circ \Lambda$ as a test function in (5.4) (recall that Λ and hence $\psi \circ \Lambda$ is continuous) yields

$$\int_{X^N} \psi \circ \Lambda \, \mathrm{d}\mu = \int_{\mathcal{P}_{\frac{1}{N}}(X)} \psi(\lambda) \, \mathrm{d}\alpha(\lambda)$$

i.e. $\alpha = \Lambda_{\#}\mu$, showing in particular the uniqueness of α.

□

From Theorem 5.1 and the law of large numbers, we easily deduce how to recover the prior α from sampling, as emphasized in our introduction.

Corollary 5.2. *Let $Z = (Z_1, \dots, Z_N)$ be a finitely exchangeable sequence of random variables with values in X, let $\mu \in \mathcal{P}_{\mathrm{sym}}(X^N)$ be the law of Z and let $\alpha \in \mathcal{P}(\mathcal{P}(X))$ be such that $\alpha(\mathcal{P}_{\frac{1}{N}}(X)) = 1$ and*

$$\mu = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,N}(\lambda)\, d\alpha(\lambda). \tag{5.5}$$

Let $(Z^{(\nu)})_{\nu\in\mathbb{N}}$ be i.i.d drawn according to μ, and consider the $\mathcal{P}(X)$-valued sequence

$$\Lambda(Z^{(\nu)}) := \frac{1}{N}\sum_{i=1}^{N} \delta_{Z_i^{(\nu)}}.$$

Then, almost surely, the empirical measure $\frac{1}{n}\sum_{\nu=1}^{n} \delta_{\Lambda(Z^{(\nu)})}$ converges narrowly to α as $n \to \infty$.

Proof. We have seen in the proof of Theorem 5.1 that $\alpha = \Lambda_{\#}\mu$ so that $\Lambda(Z)$ has law α. Hence $(\Lambda(Z^{(\nu)}))_\nu$ are i.i.d $\mathcal{P}(X)$-valued drawn according to α. Since $\mathcal{P}(X)$ endowed with the topology of narrow convergence is a Polish space, the claim follows from the strong law of large numbers for empirical measures on Polish spaces. □

5.2 Independent sequences as a convex mixture of extremal exchangeable sequences

Let us compare, in a simple example, the traditional "basis" for representing exchangeable sequences, independent sequences, with the new one advocated in this paper, extremal exchangeable sequences. This example illustrates that extremal exchangeables may not be a convex mixture of independents, but that independents are always a unique convex mixture of extremal exchangeables (Theorem 5.1).

Example. Let X be the finite state space consisting of the three colors red, green, and blue, that is to say

$$X = \{\mathtt{r},\, \mathtt{g},\, \mathtt{b}\},$$

and let $N = k = 3$. Consider the probability measure

$$\lambda_* = \tfrac{2}{3}\delta_{\mathtt{r}} + \tfrac{1}{3}\delta_{\mathtt{g}},$$

which corresponds to an `rrg` urn (i.e. an urn containing two red balls and one green ball). Now consider a sequence of three independent random variables with law λ_*, that is to say

$$(Z_1, Z_2, Z_3) \sim \lambda_*^{\otimes 3}.$$

This sequence corresponds to three independent draws with replacement from an `rrg` *urn.* Our goal is to find the unique representation of this joint law as a convex mixture of extreme exchangeables.

Denote the above λ_* by `rrg`, and similiarly $\delta_{\mathtt{r}}$ by `rrr` etc. By Theorem 5.1 and the fact that λ_* contains only red and green balls, the joint law $\lambda_*^{\otimes 3}$ must be representable as a convex combination of those $F_{N,3}(\lambda)$ where λ contains only red and green balls, that is, $\lambda =$`rrr`, `rrg`, `rgg`, `ggg`. To determine the convex combination we need to compute the $F_{N,3}(\lambda)$, which is a straightforward task given our explicit formula from section 4.2,

$$F_{N,3}(\lambda) = \frac{N^2}{(N-1)(N-2)} \left[\lambda^{\otimes 3} - \frac{3}{N} S_3\Big((\mathrm{id}_{\#}^{\otimes 2}\, \lambda) \otimes \lambda \Big) + \frac{2}{N^2}\, \mathrm{id}_{\#}^{\otimes 3}\, \lambda \right].$$

The result is given in the following table, where we identify probability measures $\lambda = \lambda_1\delta_{\mathtt{r}} + \lambda_2\delta_{\mathtt{g}} + \lambda_3\delta_{\mathtt{b}}$ in $\mathcal{P}(X)$ with their coefficient vectors in $\mathbb{R}^3$, and – analogously – probability measures in $\mathcal{P}(X^3)$ with their coefficient tensors in $\mathbb{R}^{3\times 3\times 3}$.

	$\big(F_{N,3}(\lambda)\big)_{\cdot\cdot 1}$	$\big(F_{N,3}(\lambda)\big)_{\cdot\cdot 2}$	$\big(F_{N,3}(\lambda)\big)_{\cdot\cdot 3}$
$\lambda = \mathtt{rrr}$	$\begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$
$\lambda = \mathtt{rrg}$	$\begin{pmatrix} 0 & \frac{1}{3} & \\ \frac{1}{3} & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{3} & & \\ & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$
$\lambda = \mathtt{rgg}$	$\begin{pmatrix} 0 & & \\ & \frac{1}{3} & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & \frac{1}{3} & \\ \frac{1}{3} & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$
$\lambda = \mathtt{ggg}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 1 & \\ & & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$

On the other hand, the joint measure we seek to represent is

$$(\lambda_*^{\otimes 3})_{\cdot\cdot 1} = \begin{pmatrix} \frac{8}{27} & \frac{4}{27} & \\ \frac{4}{27} & \frac{2}{27} & \\ & & 0 \end{pmatrix}, \quad (\lambda_*^{\otimes 3})_{\cdot\cdot 2} = \begin{pmatrix} \frac{4}{27} & \frac{2}{27} & \\ \frac{2}{27} & \frac{1}{27} & \\ & & 0 \end{pmatrix}, \quad (\lambda_*^{\otimes 3})_{\cdot\cdot 3} = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}.$$

From the above explicit expressions one sees that

$$\lambda_*^{\otimes 3} = \tfrac{8}{27} F_{N,3}(\mathtt{rrr}) + \tfrac{4}{9} F_{N,3}(\mathtt{rrg}) + \tfrac{2}{9} F_{N,3}(\mathtt{rgg}) + \tfrac{1}{27} F_{N,3}(\mathtt{ggg}). \tag{5.6}$$

Thus the probability measure α in $\mathcal{P}(\mathcal{P}(X))$ in (5.1) is in our case given by $\alpha = \frac{8}{27}\delta_{\mathtt{rrr}} + \frac{4}{9}\delta_{\mathtt{rrg}} + \frac{2}{9}\delta_{\mathtt{rgg}} + \frac{1}{27}\delta_{\mathtt{ggg}}$.

The representation (5.6) has the following probabilistic meaning. *One can simulate drawing three balls with replacement from an* `rrg` *urn by*

- *first picking one of the four urns* `rrr`*,* `rrg`*,* `rgg`*,* `ggg` *with probabilities* $\frac{8}{27}$*,* $\frac{4}{9}$*,* $\frac{2}{9}$*,* $\frac{1}{27}$ *(i.e. probability ratios* $8 : 12 : 6 : 1$*),*
- *then drawing three balls without replacement from the picked urn.*

Moreover thanks to the uniqueness result in Theorem 5.1, the above choice of urns and probabilities provides the unique way of simulating the given draws with replacement by draws *without* replacement.

We emphasize that the converse (simulating draws without replacement by draws *with* replacement) is not possible, due to the well known fact that finite exchangeable laws may not be representable as convex mixtures of independents. For instance, $F_{N,3}(\mathtt{rrg})$ (the joint law for three draws without replacement from an `rrg` urn), being extremal thanks to Theorem 4.5, is not a convex combination of any other joint laws, and in particular not of any independent joint laws.

Let us also provide a simple example which shows that uniqueness of the measure α in (5.1) can fail when $k < N$.

Example. Let X, N, and λ_* be as above, but $k = 2$. The measure $\lambda_*^{\otimes 2}$ (corresponding to two independent draws from a `rrg` urn) cannot just be represented by the right hand side of (5.6) with $F_{N,3}$ replaced by $F_{N,2}$, but alternatively by $\frac{2}{9} F_{N,2}(\mathtt{rrr}) + \frac{2}{3} F_{N,2}(\mathtt{rrg}) + \frac{1}{9} F_{N,2}(\mathtt{ggg})$, as the reader can easily check.

5.3 Extremal N-representable k-plans

The integral representation given by (5.1) in Theorem 5.1 will allow us to identify the set of extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ as

$$\begin{aligned} E_{N,k} &:= \{F_{N,k}(\lambda) \,:\, \lambda \in \mathcal{P}_{\frac{1}{N}}(X)\} \\ &= \{M_k S_N \delta_{x_1,\ldots,x_N} \,:\, (x_1,\ldots,x_N) \in X^N\} \end{aligned} \tag{5.7}$$

and to show in addition that all these points are exposed. Recall that if C is a convex subset of $\mathcal{P}(X^k)$ then $\mu \in C$ is an exposed point of C if there exists $\varphi \in C_b(X^k)$ such that $\int_{X^k} \varphi \, \mathrm{d}\mu < \int_{X^k} \varphi \, \mathrm{d}\nu$ for every $\nu \in C \setminus \{\mu\}$. It is obvious that exposed points are extreme points but the converse need not be true in general. For the set of N-representable k-plans, we have the following

Theorem 5.3. *Let $N \geq k \geq 2$.*

a) The set of extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is given by the set $E_{N,k}$ defined in (5.7).

b) Every such extreme point is also exposed.

Proof. a) Let μ be an extreme point of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$. Let us write μ as in (5.1) and prove that α is a Dirac mass. If this was not the case, we could find $\varphi \in C_b(X)$ and $t \in \mathbb{R}$ such that $A_1 := \{\lambda \in \mathcal{P}(X) \ : \ \int_X \varphi \, \mathrm{d}\lambda > t\}$ and $A_2 := \{\lambda \in \mathcal{P}(X) \ : \ \int_X \varphi \, \mathrm{d}\lambda \leq t\}$ satisfy $\alpha(A_1) > 0$ and $\alpha(A_2) > 0$. Then decomposing α as $\alpha(A_1)\alpha_1 + \alpha(A_2)\alpha_2$ with α_i a probability measure giving full mass to $\mathcal{P}_{\frac{1}{N}}(X) \cap A_i$, the extremality of μ would imply

$$\int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda) \, \mathrm{d}\alpha_1(\lambda) = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda) \, \mathrm{d}\alpha_2(\lambda).$$

But recalling that for $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$ we have $M_1 F_{N,k}(\lambda) = \lambda$, this would also give

$$\int_{\mathcal{P}_{\frac{1}{N}}(X)} \lambda \, \mathrm{d}\alpha_1(\lambda) = \int_{\mathcal{P}_{\frac{1}{N}}(X)} \lambda \, \mathrm{d}\alpha_2(\lambda),$$

and integrating φ against this identity would lead to a contradiction. Extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ therefore belong to $E_{N,k}$

The reverse inclusion follows from b). Alternatively, it follows from Theorem 4.2, for if $\mu = F_{N,k}(\lambda)$ for some $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, and μ is a strict convex combination of two measures $\mu^1 \neq \mu^2$ in $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, then the support of the μ^i must be contained the support of μ and thus – a fortiori – in $(\operatorname{supp} \lambda)^{\mathrm{k}}$, contradicting the extremality of $F_{N,k}(\lambda)$ among N-representable k-point measures supported on the finite state space $(\operatorname{supp} \lambda)^k$.

b) We need to show that if $\lambda := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}_{\frac{1}{N}}(X)$, then $\mu := F_{N,k}(\lambda)$ is an exposed point of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$. Again thanks to the integral representation (5.1), this amounts to finding $\varphi \in C_b(X^k)$ such that $\int_{X^k} \varphi \mathrm{d}\mu < \int_{X^k} \varphi \mathrm{d}\nu$ for every $\nu \in E_{N,k} \setminus \{\mu\}$.

First, let us rewrite $\lambda := \sum_{j=1}^{\ell} \lambda_j \delta_{y_j}$ with $\lambda_j > 0$ and $y_1, \ldots, y_\ell$ pairwise distinct. Next, we define

$$\varphi_0(z_1, \ldots, z_k) := \sum_{i=1}^{k} \min_{j=1,\ldots,\ell} d(z_i, y_j), \ \forall (z_1, \ldots, z_k) \in X^k.$$

Let u be a function in $C_b(X,\mathbb{R}^\ell)$ such that $u(y_1),\dots,u(y_\ell)$ are linearly independent. For every $\theta\in\mathcal{P}(X)$, let us now define

$$\begin{aligned}J(\theta):&=\frac{1}{2}\int_{X\times X}|u(x)-u(y)|^2\mathrm{d}\theta(x)\mathrm{d}\theta(y)\\&=\int_X|u(x)|^2\mathrm{d}\theta(x)-\Big|\int_X u(x)\mathrm{d}\theta(x)\Big|^2\end{aligned}$$

Obviously J is a concave quadratic functional and more precisely, defining

$$f_\lambda(x):=|u(x)|^2-2\Big(\int_X u(y)\mathrm{d}\lambda(y)\Big)\cdot u(x),\ \forall x\in X, \tag{5.8}$$

for every $\theta\in\mathcal{P}(X)$ we have

$$J(\theta)=J(\lambda)+\int_X f_\lambda(x)\mathrm{d}(\theta-\lambda)(x)-\Big|\int_X u(x)\mathrm{d}(\theta-\lambda)(x)\Big|^2.$$

In particular, this implies that

$$J(\theta)\le J(\lambda)+\int_X f_\lambda(x)\mathrm{d}(\theta-\lambda)(x) \tag{5.9}$$

and this inequality is strict unless

$$\int_X u\mathrm{d}\theta=\int_X u\mathrm{d}\lambda=\sum_{j=1}^{l}\lambda_j u(y_j). \tag{5.10}$$

Now if $\theta\in\mathcal{P}_{\frac{1}{N}}(X)$ and $\nu:=F_{N,k}(\theta)$, we know (see (4.7)) that

$$M_2\nu=\frac{N}{N-1}\theta\otimes\theta-\frac{1}{N-1}\,\mathrm{id}_{\#}^{\otimes 2}\,\theta,\ M_1\nu=\theta$$

since $(x,y)\in X^2\mapsto|u(x)-u(y)|^2$ vanishes on the diagonal and thanks to (5.9), we thus get

$$\begin{aligned}\frac{N}{N-1}J(\theta)&=\frac{1}{2}\int_{X^k}|u(z_1)-u(z_2)|^2\mathrm{d}\nu(z_1,\dots,z_k)\\&\le\frac{1}{2}\int_{X^k}|u(z_1)-u(z_2)|^2\mathrm{d}\mu(z_1,\dots,z_k)\\&+\frac{N}{N-1}\int_{X^k}f_\lambda(z_1)\mathrm{d}(\nu-\mu)(z_1,\dots,z_k)\end{aligned}$$

and the last inequality is strict unless (5.10) holds. Defining

$$\varphi_1(z_1,\dots,z_k):=-\frac{1}{2}|u(z_1)-u(z_2)|^2+\frac{N}{N-1}f_\lambda(z_1),\ \forall(z_1,\dots,z_k)\in X^k$$

we deduce that $\int_{X^k} \varphi_1 \mathrm{d}\mu \leq \int_{X^k} \varphi_1 \mathrm{d}\nu$, $\forall \nu \in E_{N,k}$. Since $\int_{X^k} \varphi_0 \mathrm{d}\mu = 0$ and $\varphi_0 \geq 0$, setting $\varphi := \varphi_0 + \varphi_1$ we also have

$$\int_{X^k} \varphi \mathrm{d}\mu \leq \int_{X^k} \varphi \mathrm{d}\nu, \ \forall \nu \in E_{N,k}. \tag{5.11}$$

It remains to show that if $\theta \in \mathcal{P}_{\frac{1}{N}}(X)$ is such that (5.11) is an equality at $\nu = F_{N,k}(\theta)$ then necessarily $\theta = \lambda$. But in this equality case we must have $\int_{X^k} \varphi_0 \mathrm{d}\nu = 0$ so that $\theta = M_1 \nu$ is supported by the finite set $\{y_1, \dots y_\ell\}$ (i.e. $\theta = \sum_{j=1}^{\ell} \theta_j \delta_{y_j}$), but we must also have an equality in (5.9). The latter implies that $\sum_{j=1}^{\ell} \theta_j u(y_j) = \sum_{j=1}^{\ell} \lambda_j u(y_j)$, i.e. $\lambda_j = \theta_j$ for $j = 1, \dots, \ell$, since we have chosen $u(y_1), \dots, u(y_\ell)$ linearly independent. □

Let us point out the following consequence of Theorem 5.3 a).

Corollary 5.4. *The set of extreme points of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is closed under narrow convergence.*

Proof. We have to show that $E_{N,k}$ is closed under narrow convergence. Take a sequence $\mu^n = F_{N,k}(\lambda^n) \in E_{N,k}$ converging narrowly to some $\mu \in \mathcal{P}(X^k)$. Since the one-point marginals $M_1 \mu^n = \lambda^n$ belong to $\mathcal{P}_{\frac{1}{N}}(X)$ and the marginal map M_1 is continuous under narrow convergence, $\lambda^n \to \lambda$ for some $\lambda \in \mathcal{P}(X)$. By the closedness of $\mathcal{P}_{\frac{1}{N}}(X)$ (see Lemma 3.1) we have $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$ and the continuity of $F_{N,k}$ yields $F_{N,k}(\lambda^n) \to F_{N,k}(\lambda)$, so $\mu = F_{N,k}(\lambda) \in E_{N,k}$. □

In some applications, one is interested in the following subset of $\mathcal{P}_{N-rep}(X^k)$ consisting of measures with no mass on the diagonal:

$$\mathcal{P}^{\mathrm{offdiag}}_{N-rep}(X^k) := \{\mu \in \mathcal{P}_{N-rep}(X^k) \ : \ \mu(diag_k(X)) = 0\} \tag{5.12}$$

where

$$diag_k(X) := \{(x, ..., x) \in X^k \ : \ x \in X\}. \tag{5.13}$$

The motivation for considering this set comes from singular k-body interactions. These interactions lead to costs of the form

$$c(x_1, ..., x_N) = \sum_{1 \leq i_1 < ... < i_k \leq N} \Phi(x_{i_1}, ..., x_{i_k})$$

for some measurable nonnegative Φ with $\Phi = +\infty$ on $diag_k(X)$, then the total cost $C[\gamma] = \int_{X^N} c \, d\gamma$ is infinite if $M_k \gamma \notin \mathcal{P}^{\mathrm{offdiag}}_{N-rep}(X^k)$. A prototypical example is the Coulomb cost, see Section 8, which led Khoo and Ying [26] to introduce the set (5.12)–(5.13) in the case $k = 2$ and X a finite state space. Recall the quantization constraint for measures $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$ that $\lambda(\{x\}) \in \{0, 1/N, ..., 1\}$.

Theorem 5.5. *a) The set of extreme points of* $\mathcal{P}^{\text{offdiag}}_{N-rep}(X^k)$ *is given by*

$$E^{\text{offdiag}}_{N,k} := \Big\{ F_{N,k}(\lambda) \,:\, \lambda \in \mathcal{P}_{\frac{1}{N}}(X) \,:\, \lambda(\{x\}) \in \{0, \tfrac{1}{N}, ..., \tfrac{k-1}{N}\} \Big\}. \tag{5.14}$$

b) Every such extreme point is also exposed.

Geometrically this means that intersecting the set of N-representable k-plans with the subspace of measures which vanish on the diagonal does not generate new extreme points.

Theorem 5.5 generalizes a recent result of [26] from $k = 2$ and finite state spaces X to general k and X.

Proof. First we show that any extreme point μ is contained in the set (5.14). By the representation (5.1), $\mu = \int_A F_{N,k}(\lambda)\, d\alpha(\lambda)$ where $A = \{\lambda \in \mathcal{P}_{\frac{1}{N}}(X) : F_{N,k}(\lambda)(diag_k(X)) = 0\}$. The key point is that the polynomial $F_{N,k}(\lambda)$ factorizes on the diagonal:

$$\begin{aligned} F_{N,k}(\lambda)(\{(x,...,x)\}) &= \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \Big[\lambda(\{x\})^k + \sum_{j=1}^{k-1} \frac{(-1)^j}{N^j} \lambda(\{x\})^{k-j} c_j^{(k)} \Big] \\ &= \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \lambda(\{x\}) \cdot \Big(\lambda(\{x\}) - \tfrac{1}{N} \Big) \cdot ... \cdot \Big(\lambda(\{x\}) - \tfrac{k-1}{N} \Big) \end{aligned} \tag{5.15}$$

The second equality follows because the coefficients $c_j^{(k)}$ are N-independent and satisfy

$$N^k \Big(1 - \frac{c_1^{(k)}}{N} + \frac{c_2^{(k)}}{N^2} - + ... + (-1)^{k-1} \frac{c_{k-1}^{(k)}}{N^{k-1}} \Big) = \prod_{j=0}^{k-1} (N-j)$$

for all nonnegative integers N, see eq. (4.18). Since λ is $\frac{1}{N}$-quantized, $\lambda(x)N$ is a nonnegative integer and can be substituted for N, yielding the asserted expression. The factorization identity (5.15) shows that $A = \{\lambda \in \mathcal{P}_{\frac{1}{N}}(X) : \lambda(x) \in \{0, \frac{1}{N}, ..., \frac{k-1}{N}\}\}$. The rest of the proof of a) is analogous to that of Theorem 5.3 a) and the exposedness of the elements in $E^{\text{offdiag}}_{N,k}$ is immediate from Theorem 5.3 b).

□

5.4 Exponentially convergent approximation of extreme points

The error estimates (4.14), (4.15) when approximating $F_{N,k}(\lambda)$ by a truncated series can be extended in a straightforward manner to Polish spaces.

Theorem 5.6. *For any extreme point μ_k of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, denoting its one-point marginal $M_1\mu_k \in \mathcal{P}_{\frac{1}{N}}(X)$ by λ we have*

$$\mu_k = \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \left[\lambda^{\otimes k} + \varepsilon_{N,k}(\lambda)\right] \quad \textit{with } \|\varepsilon_{N,k}(\lambda)\|_{\mathrm{TV}} \leq \frac{C_k}{N} \tag{5.16}$$

and

$$\begin{aligned} \mu_k =& \frac{N^{k-1}}{\prod_{j=1}^{k-1}(N-j)} \left[\lambda^{\otimes k} + \sum_{j=1}^{p} \frac{(-1)^j}{N^j} S_k\, P_j^{(k)}(\lambda) + \varepsilon_{N,k,p}(\lambda)\right] \\ & \textit{with } \|\varepsilon_{N,k,p}(\lambda)\|_{\mathrm{TV}} \leq \frac{C_k}{N^{p+1}}, \end{aligned} \tag{5.17}$$

with constants C_k independent of N and p. The explicit constants (4.16) *will do. Conversely, for every $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$ there exists an extreme point μ_k of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ such that* (5.16), (5.17) *hold.*

Proof. By Theorem 5.3, an extreme point μ_k equals $F_{N,k}(\lambda)$, and in particular is supported on the finite set $(\operatorname{supp}\lambda)^k$. Estimates (5.16), (5.17) now follow from (4.14)–(4.15) and the fact that

$$||\varepsilon_{N,k}(\lambda)||_{TV(X^k)} = ||\varepsilon_{N,k}(\lambda)||_{TV((\operatorname{supp}\lambda)^k)}$$

(and the analogous identity for $\varepsilon_{N,k,p}$). □

Of course, via the integral representation (5.1) this translates into an analogous exponentially convergent approximation result for arbitrary (non-extremal) N-representable measures, see the next section.

6 Recovering the de Finetti-Hewitt-Savage theorem

The celebrated Hewitt-Savage theorem [19] (going back to de Finetti in the special case of the state space $X = \{0, 1\}$) says the following.

Theorem 6.1. (Hewitt-Savage [19]) *If $(Z_n)_n$ is an infinite exchangeable sequence of random variables with values in the Polish space X, then there exists $\alpha \in \mathcal{P}(\mathcal{P}(X))$ such that for every k, the law μ_k of $(Z_1, \dots, Z_k)$ is given by*

$$\mu_k = \int_{\mathcal{P}(X)} \lambda^{\otimes k} \ \ d\alpha(\lambda). \tag{6.1}$$

Let us show how this theorem follows from Theorem 5.1 which therefore may be viewed as a non-asymptotic form of the Hewitt-Savage theorem. If μ_k is the law of the first k components of an infinite exchangeable sequence, it is N-representable for every N, so that there exists $\alpha^{N,k} \in \mathcal{P}(\mathcal{P}_{\frac{1}{N}}(X))$ such that

$$\mu_k = \int_{\mathcal{P}_{\frac{1}{N}}(X)} F_{N,k}(\lambda) \mathrm{d}\alpha^{N,k}(\lambda).$$

In particular, for the law μ_1 of Z_1 we have

$$\mu_1 := \int_{\mathcal{P}(X)} \lambda \ \mathrm{d}\alpha^{N,k}(\lambda).$$

Since X is Polish, μ_1 is tight so there exists $\Psi : X \to [0, +\infty]$ which is lower semicontinuous and coercive (i.e. with compact sublevel sets) such that

$$+\infty > \int_X \Psi(x) \mathrm{d}\mu_1(x) = \int_{\mathcal{P}(X)} \Big(\int_X \Psi(x) \mathrm{d}\lambda(x) \Big) \mathrm{d}\alpha^{N,k}(\lambda).$$

But since Ψ is coercive, $\lambda \in \mathcal{P}(X) \mapsto \int_X \Psi(x) \mathrm{d}\lambda(x)$ is coercive on $\mathcal{P}(X)$. It thus follows from Prokhorov's theorem that $\alpha^{N,k}$ is tight. Hence, taking a subsequence if necessary, $\alpha^{N,k}$ converges narrowly to some α^k as $N \to \infty$, and since $F_{N,k}(\lambda)$ converges uniformly to $\lambda^{\otimes k}$, thanks to (4.14) we get letting $N \to \infty$:

$$\mu_k = \int_{\mathcal{P}(X)} \lambda^{\otimes k} \ \mathrm{d}\alpha^k(\lambda).$$

It remains to show that α^k may be chosen independently of k. For this we simply observe that for $\ell \geq k$, $\mu_k = M_k(\mu_\ell)$. Since $M_k(\lambda^{\otimes \ell}) = \lambda^{\otimes k}$, thanks to the linearity and narrow continuity of M_k we thus get

$$\mu_k = \int_{\mathcal{P}(X)} \lambda^{\otimes k} \ \mathrm{d}\alpha^\ell(\lambda), \ \ \text{for all } \ell \geq k.$$

We then use the same tightness argument as above to see that for some sequence $\ell_n \to \infty$, α^{ℓ_n} converges narrowly to some α as $n \to +\infty$ and the Hewitt-Savage representation (6.1) holds for this measure α.

Diaconis and Freedman (in [11]) obtained TV bounds between elements of $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ and mixtures of independent measures $\lambda^{\otimes k}$. More precisely, Diaconis and Freedman showed that if $\mu_k \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$ and $k \leq N$, there exists $\alpha \in \mathcal{P}(\mathcal{P}(X))$ such that, for some constant B_k (for results on their k-dependence see [11])

$$\|\mu_k - \int_{\mathcal{P}(X)} \lambda^{\otimes k} \mathrm{d}\alpha\|_{\mathrm{TV}} \leq \frac{B_k}{N}. \tag{6.2}$$

These $O(\frac{1}{N})$ bounds are recovered by using the integral representation of $\mu_k \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$ of Theorem 5.1 and applying (5.16).

A better approximation of μ_k is obtained by adding the universally correlated correction terms from (5.17) to the independent measures $\lambda^{\otimes k}$ in (6.2) and applying (5.17). This yields:

Corollary 6.2. *For any measure μ_k in $\mathcal{P}_{N-\mathrm{rep}}(X^k)$, and with α as in Theorem 5.1, we have for $p = 1, ..., k-2$ and some constant B_k*

$$\|\mu_k - \int_{\mathcal{P}(X)} \tfrac{N^{k-1}}{\Pi_{j=1}^{k-1}(N-j)} \big[\lambda^{\otimes k} + \sum_{j=1}^{p} \frac{(-1)^j}{N^j} S_k P_j^{(k)}(\lambda)\big]\, d\alpha\|_{\mathrm{TV}} \leq \frac{B_k}{N^{p+1}}\,. \tag{6.3}$$

7 Connection with the Ewens sampling formula

When looking at the coefficients arising in the classification (4.10)–(4.12) of extremal N-representable measures, we noticed some resemblance to those in the celebrated *Ewens distribution* or *Ewens sampling formula* from genetics, which – mathematically speaking – is a probability distribution on integer partitions. On the other hand, the sign factor $(-1)^j$ in (4.10), or $(-1)^{k-n(\mathbf{p}')}$ in the alternative expression (4.29), means that some coefficients are negative, so the coefficients cannot be a probability distribution on integer partitions. *What is happening here?*

There is a precise connection which seems to us quite remarkable and which we now describe.

The Ewens distribution is usually stated in terms of the following parametrization of partitions. If $\mathbf{p}' = (p'_1, ..., p'_m)$ is a partition of the integer k, that is to say $p'_1 + ... + p'_m = k$ and $p'_1 \geq ... \geq p'_m \geq 1$, one denotes by m_q the number of components equal to q (i.e., the number of times the term q appears in the sum), $m_q := |\mathbf{p}'^{-1}(q)|$, and introduces the vector $(m_1, ..., m_k)$ (called allelic partition). For instance, the partition $\mathbf{p}' = (2, 1, 1, 1)$ of 5 corresponds to the allelic partition $(3, 1, 0, 0, 0)$, because 1 appears 3 times, 2 appears once, and

3, 4, 5 appear zero times. The Ewens distribution on the allelic partitions of the integer $k \in \mathbb{N}$ is now given by

$$\mathbb{P}(m_1, ..., m_k; \theta) = \frac{k!}{\theta(\theta+1)...(\theta+k-1)} \prod_{q=1}^{k} \frac{\theta^{m_q}}{q^{m_q} m_q!}, \tag{7.1}$$

where $\theta > 0$ is a parameter. In genetics, this distribution models the genetic diversity of a population (at statistical equilibrium and under neutral selection), and θ is the *mutation rate*. More precisely, for a random sample of k genes taken from a population at a particular locus, eq. (7.1) gives the probability that m_1 alleles (variant forms of the gene) appear exactly once, m_2 alleles appear exactly twice, and so on. The Ewens distribution has found widespread use in biology, statistics, probability, and even number theory [13], see [14, 24] for original papers and [9] for a recent review.

In terms of the standard representation of the partitions of k, expression (7.1) becomes

$$\frac{k!}{\theta(\theta+1)...(\theta+k-1)} \theta^{n(\mathbf{p}')} \prod_{i=1}^{n(\mathbf{p}')} \frac{1}{p'_i} \frac{1}{\prod_{q \in \operatorname{Ran} \mathbf{p}'} (|\mathbf{p}'^{-1}(q)|)!} =: \mathtt{ew}(\mathbf{p}'; \theta). \tag{7.2}$$

Definition 7.1. *Given any partition $\mathbf{p}'$ of the integer k, we define the complex Ewens function $\theta \mapsto \mathtt{ew}(\mathbf{p}'; \theta)$ by (7.2), where $\theta \in \mathbb{C}$ is a complex parameter and the arithmetic operations are the usual ones in $\mathbb{C}$.*

For a picture when $k = 3$ see Figure 2.

The complex Ewens function has the following properties:

1. $\mathtt{ew}(\mathbf{p}'; \cdot)$ has poles exactly at the negative integers $\theta = -1, ..., -(k-1)$, and is holomorphic on $\mathbb{C} \backslash \{-1, ..., -(k-1)\}$. In particular, $\mathtt{ew}(\mathbf{p}'; \cdot)$ is meromorphic on $\mathbb{C}$.

2. $\mathtt{ew}(\mathbf{p}'; \cdot)$ is nonzero for the trivial partition $\mathbf{p}' = k$, and zero only at $\theta = 0$ for all other partitions of k.

3. $\sum_{\mathbf{p}' \text{partition of } k} \mathtt{ew}(\mathbf{p}'; \cdot)$ is identically equal to 1.

4. $\{\mathtt{ew}(\mathbf{p}'; \theta)\}_{\mathbf{p}' \text{partition of } k}$ is a probability measure (i.e., in addition real and nonnegative) when θ belongs to the nonnegative real axis, and a signed measure (i.e., in addition real but not nonnegative) when θ belongs to the negative real axis excluding the poles.

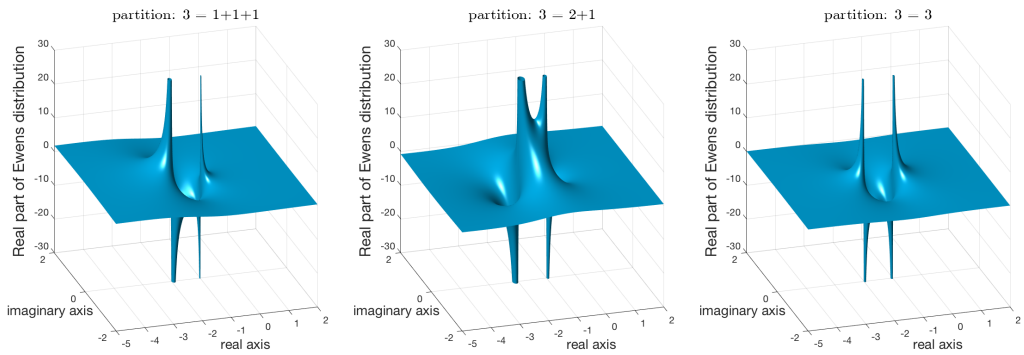

Figure 2: The complex Ewens distribution $(\mathbf{p}', \theta) \mapsto \mathtt{ew}(\mathbf{p}'; \theta)$ for $k = 3$. Since there are three partitions $\mathbf{p}'$ of 3, it is a collection of three meromorphic functions of the complex variable θ. The poles of each function are located at $\theta = -1$ and $\theta = -2$. Restricting θ to the positive real axis yields the classical Ewens distribution, which is a probability measure for each θ, and has long been known to be relevant in genetics. Its analytic continuation to the negative real axis, which to our knowledge has not been introduced before, is a signed measure for each θ, and turns out to describe extremal N-representable measures.

The first two properties are immediate from the definition. The third follows from the fact that the sum is identically equal to 1 for positive real θ and the identity theorem from complex analysis. Property 4 is an obvious consequence of Property 3 and the definition.

The connection to extremal N-representable measures is the following.

Theorem 7.2. *The coefficients $c_{\mathbf{p}'}$ in the expansion* (4.29) *of extremal N-representable k-point measures are given by $\mathtt{ew}(\mathbf{p}'; \theta)|_{\theta=-N}$; that is, they are given by the Ewens distribution evaluated at a probabilistically meaningless but by analytic continuation admissible negative parameter value.*

Properties 3 and 4 of the complex Ewens distribution thus recover the phenomenon discussed at the beginning of the section that the coefficients $c_{\mathbf{p}'}$ in (4.29) are only a signed measure of total mass 1 but not a probability measure. (As already noted below Theorem 4.5, due to this phenomenon it is highly nontrivial that the induced measures (4.23), $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, are nevertheless probability measures.)

Some sort of correspondence between N-representable k-point measures and the Ewens distribution at $\theta = -N$ also exists at the poles. By Property 1, the complex Ewens distribution stops making mathematical sense at the

points $-N$ when $N \in \mathbb{N}$, $N < k$; but these are precisely the N's for which the notion of N-representability (see Definition (2.1)) stops making sense as a k-plan cannot be N-representable when $N < k$. Thus the two poles seen in Figure 2 at -1 and -2 reflect the fact that it makes no sense for a 3-point probability measure to be 1- or 2-representable.

8 Applications to optimal transport

Our interest in the structure of the set $\mathcal{P}_{N-\mathrm{rep}}(X^k)$ is motivated by symmetric multi-marginal optimal transport problems. More precisely, given integers $2 \leq k \leq N$ and $\Phi \in C_b(X^k)$ *symmetric* (i.e., $\Phi(x_{\sigma(1)}, \ldots, x_{\sigma(k)}) = \Phi(x_1, \ldots, x_k)$ for every $(x_1, \ldots, x_k) \in X^k$ and every permutation $\sigma \in \mathcal{S}_k$), we consider the symmetric cost c_Φ defined on X^N by

$$c_\Phi(x_1, \cdots, x_N) := \frac{1}{\binom{N}{k}} \sum_{1 \leq i_1 < i_2 \ldots < i_k \leq N} \Phi(x_{i_1}, \ldots, x_{i_k}). \tag{8.1}$$

In this setting N is the total number of particles in the system, X^N is the space of all N-particle configurations, and Φ is a k-body interaction potential. In practice N is much larger than k. Given $\rho \in \mathcal{P}(X)$ we are interested in the multi-marginal optimal transport problem

$$C_{N,k}(\rho) := \inf \Big\{ \int_{X^N} c_\Phi(x_1, \ldots, x_N) \mathrm{d}\gamma(x_1, \ldots, x_N) \ : \ \gamma \in \mathcal{P}_{\mathrm{sym}}(X^N), \ M_1\gamma = \rho \Big\}. \tag{8.2}$$

An important example which has received much recent interest is $k = 2$, $X = \mathbb{R}^3$, $\Phi(x_1, x_2) = |x_1 - x_2|^{-1}$ (optimal transport with Coulomb cost [7, 5]), in which case (8.2) arises as the strictly correlated limit of Hohenberg-Kohn density functional theory [7]. Thanks to (8.1), the minimization can be reformulated in terms of the k-marginal $\mu_k = M_k(\gamma) \in \mathcal{P}_{N-\mathrm{rep}}(X^k)$:

$$C_{N,k}(\rho) := \inf \Big\{ \int_{X^k} \Phi \mathrm{d}\mu_k \ : \ \mu_k \in \mathcal{P}_{N-\mathrm{rep}}(X^k), \ M_1(\mu_k) = \rho \Big\}. \tag{8.3}$$

8.1 A $\frac{1}{N}$-quantized polynomial convexification problem

The de Finetti style representation from Theorem 5.1, together with the fact that $M_1(F_{N,k}(\lambda)) = \lambda$ for every $\lambda \in \mathcal{P}_{\frac{1}{N}}(X)$, enables us to write $C_{N,k}(\rho)$ as follows.

Theorem 8.1. (Reformulation of symmetric multi-marginal optimal transport) *The functional $C_{N,k}$ satisfies*

$$C_{N,k}(\rho) = \inf\Big\{ \int_{\mathcal{P}_{\frac{1}{N}}(X)} \Big(\int_{X^k} \Phi \; dF_{N,k}(\lambda) \Big) d\alpha(\lambda) \; : \; \alpha \in \mathcal{P}(\mathcal{P}_{\frac{1}{N}}(X)), \; (8.5) \Big\}, \tag{8.4}$$

with prescribed marginal constraint

$$\int_{\mathcal{P}_{\frac{1}{N}}(X)} \lambda \; d\alpha(\lambda) = \rho. \tag{8.5}$$

In view of formulae (4.10)–(4.12), one can observe that $\lambda \in \mathcal{P}_{\frac{1}{N}(X)} \mapsto \int_{X^k} \Phi \; \mathrm{d}F_{N,k}(\lambda)$ is a polynomial of degree k expression in the weights of the discrete measure λ, for instance

$$\int_{X^2} \Phi \; \mathrm{d}F_{N,2}(\lambda) = \frac{N}{N-1} \int_{X^2} \Phi(x,y) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) - \frac{1}{N-1} \int_X \Phi(x,x) \mathrm{d}\lambda(x)$$

and

$$\begin{aligned} \int_{X^3} \Phi \; \mathrm{d}F_{N,3}(\lambda) = & \frac{N^2}{(N-1)(N-2)} \int_{X^3} \Phi(x,y,z) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) \mathrm{d}\lambda(z) \\ & - \frac{3N}{(N-1)(N-2)} \int_{X^2} \Phi(x,x,y) \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) \\ & + \frac{2}{(N-1)(N-2)} \int_X \Phi(x,x,x) \mathrm{d}\lambda(x). \end{aligned}$$

Defining the polynomial $P_{N,k}$ for every single marginal (not necessarily $\frac{1}{N}$-quantized) $\lambda \in \mathcal{P}(X)$ by

$$P_{N,k}(\lambda) := \int_{X^k} \Phi \; \mathrm{d}F_{N,k}(\lambda), \quad \text{for all } \lambda \in \mathcal{P}(X), \tag{8.6}$$

we see that minimization problem (8.4)-(8.5) is a $\frac{1}{N}$-quantized constrained version of the convexification of the polynomial $P_{N,k}$:

$$P^{**}_{N,k}(\rho) := \inf_{\alpha \in \mathcal{P}(\mathcal{P}(X))} \Big\{ \int_{\mathcal{P}(X)} P_{N,k}(\lambda) \mathrm{d}\alpha(\lambda) \; : \; \int_{\mathcal{P}(X)} \lambda \mathrm{d}\alpha(\lambda) = \rho \Big\}. \tag{8.7}$$

In particular note that

$$C_{N,k}(\rho) \geq P^{**}_{N,k}(\rho), \; \forall \rho \in \mathcal{P}(X). \tag{8.8}$$

To understand the advantage of the formulation (8.4)-(8.5) compared to the initial multi-marginal problem (8.2), it is worth considering in detail the finite case where $X = X_\ell$ is finite with ℓ elements. In this case both (8.2) and (8.4)-(8.5) are linear programs which have $\binom{N+\ell-1}{\ell-1}$ variables and ℓ constraints. But the special structure of (8.4)-(8.5) makes it much more appealing from a computational viewpoint. Indeed, the computation of the value function $C_{N,k}$ given by (8.4)-(8.5) amounts to finding the convex envelope of the restriction of the polynomial $P_{N,k}$ to the finite set $\mathcal{P}_{\frac{1}{N}}(X)$. In the simplest case where $\ell = 2$, this amounts to computing the convex hull of $N+1$ points in the plane located in the graph of $P_{N,k}$. This convex envelope can be computed exactly in linear in N time thanks to the Graham scan algorithm [18] for instance.

8.2 Convergence as $N \to \infty$

Now we consider the asymptotic behavior of $C_{N,k}$ as $N \to \infty$. As first emphasized in [8] in the case of the Coulomb cost (or more general pairwise interactions with a potential having a positive Fourier transform), the Hewitt-Savage theorem enables one to drastically simplify the analysis of problems like (8.2) as $N \to \infty$. Since $\rho^{\otimes k}$ is admissible in the optimal transport problem (8.3) we have

$$C_{N,k}(\rho) \leq P_k(\rho) := \int_{X^k} \Phi \mathrm{d}\rho^{\otimes k}, \ \forall \rho \in \mathcal{P}(X)$$

but since $C_{N,k}$ is obviously convex this also gives

$$C_{N,k}(\rho) \leq P_k^{**}(\rho), \quad \text{for all } \rho \in \mathcal{P}(X). \tag{8.9}$$

Recalling (5.16), we observe that for some constant C_k we have for every $\lambda \in \mathcal{P}(X)$

$$P_{N,k}(\lambda) \geq \frac{(N-k)!N^k}{N!}\Big(P_k(\lambda) - \frac{C_k\|\Phi\|_\infty}{N}.\Big)$$

Taking convex envelopes and using (8.8) we thus get, for every $\rho \in \mathcal{P}(X)$:

$$\frac{(N-k)!N^k}{N!}\Big(P_k^{**}(\rho) - \frac{C_k\|\Phi\|_\infty}{N}\Big) \leq C_{N,k}(\rho) \leq P_k^{**}(\rho). \tag{8.10}$$

So $C_{N,k}$ converges uniformly on $\mathcal{P}(X)$ to P_k^{**} as $N \to +\infty$. We also have a Γ-convergence result:

Theorem 8.2. *As $N \to \infty$, the sequence of functionals $C_{N,k}$ defined in* (8.2) *Γ-converges with respect to the narrow topology of $\mathcal{P}(X)$ to P_k^{**}.*

Proof. The Γ-limsup inequality obviously follows from (8.9). As for the Γ-liminf inequality, it follows directly from (8.10) and the lower semicontinuity of P_k^{**} with respect to the narrow topology. □

To conclude, we see that the asymptotic problem obtained by letting $N \to \infty$ in (8.3) amounts to computing the convex envelope of the polynomial of degree k interaction functional P_k. This might be a challenging computational task in general but we believe that the theory developed by Lasserre for polynomial optimization (see in particular [27]) might be useful in certain situations.

Acknowledgments: G.C. gratefully acknowledges the hospitality of TU Munich, where this work was begun while he held a John Von Neumann visiting professorship.

References

[1] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[2] Patrick Billingsley. *Convergence of Probability Measures.* Wiley-Interscience, 1999.

[3] Sergey G. Bobkov. Generalized symmetric polynomials and an approximate de Finetti representation. *J. Theoret. Probab.*, 18(2):399–412, 2005.

[4] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[5] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Phys. Rev. A*, 85:062502, June 2012.

[6] Louis Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions.* Springer Netherlands, Dordrecht, Holland, 1974.

[7] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.

[8] Codina Cotar, Gero Friesecke, and Brendan Pass. Infinite-body optimal transport with Coulomb cost. *Calc. Var. Partial Differential Equations*, 54(1):717–742, 2015.

[9] Harry Crane. The ubiquitous Ewens sampling formula. *Statist. Sci.*, 31(1):1–19, 02 2016.

[10] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential. B*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982. Theory of martingales, Translated from the French by J. P. Wilson.

[11] P. Diaconis and D. Freedman. Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980.

[12] Persi Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36(2):271–281, 1977. Foundations of probability and statistics, II.

[13] Peter Donnelly and Geoffrey Grimmett. On the Asymptotic Distribution of Large Prime Factors. *Journal of the London Mathematical Society*, s2-47(3):395–404, 06 1993.

[14] W.J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1):87 – 112, 1972.

[15] Gero Friesecke, Christian B. Mendl, Brendan Pass, Codina Cotar, and Claudia Klüppelberg. N-density representability and the optimal transport limit of the Hohenberg-Kohn functional. *The Journal of Chemical Physics*, 139(16):164109, 2013.

[16] Gero Friesecke and Daniela Vögler. Breaking the Curse of Dimension in Multi-Marginal Kantorovich Optimal Transport on Finite State Spaces. *SIAM Journal on Mathematical Analysis*, 50(4):3996–4019, 2018.

[17] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[18] Ronald Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters*, 1:132–133, 1972.

[19] Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[20] Lars Hörmander. *Notions of Convexity.* Birkhäuser, Boston, Massachusetts, 1994.

[21] Svante Janson, Takis Konstantopoulos, and Linglong Yuan. On a representation theorem for finitely exchangeable random vectors. *J. Math. Anal. Appl.*, 442(2):703–714, 2016.

[22] E. Jaynes. Some applications and extensions of the de Finetti representation theorem. *in: Bayesian inverence and decision techniques: Essays in honor of Bruno de Finetti*, pages 31–42, 1986.

[23] Olav Kallenberg. *Probabilistic symmetries and invariance principles.* Probability and its Applications (New York). Springer, New York, 2005.

[24] S. Karlin and J. McGregor. Addendum to a paper of W. Ewens. *Theoretical Population Biology*, 3(1):113 – 116, 1972.

[25] G. Jay Kerns and Gábor J. Székely. De Finetti's theorem for abstract finite exchangeable sequences. *J. Theoret. Probab.*, 19(3):589–608, 2006.

[26] Yuehaw Khoo and Lexing Ying. Convex Relaxation Approaches for Strictly Correlated Density Functional Theory. *SIAM Journal on Scientific Computing*, 41(4):B773–B795, 2019.

[27] R. Laraki and J. B. Lasserre. Computing uniform convex approximations for convex envelopes and convex hulls. *J. Convex Anal.*, 15(3):635–654, 2008.