COMMUN. MATH. SCI.
Vol. 15, No. 1, pp. 77–96

EXTREME POINTS OF A BALL ABOUT A MEASURE WITH FINITE SUPPORT*

HOUMAN OWHADI† AND CLINT SCOVEL‡

Abstract. We show that, for the space of Borel probability measures on a Borel subset of a Polish metric space, the extreme points of the Prokhorov, Monge–Wasserstein and Kantorovich metric balls about a measure whose support has at most n points, consist of measures whose supports have at most $n+2$ points. Moreover, we use the Strassen and Kantorovich–Rubinstein duality theorems to develop representations of supersets of the extreme points based on linear programming, and then develop these representations towards the goal of their efficient computation.

Key words. Extreme points, Prokhorov, Kantorovich, Monge–Wasserstein, Strassen, Kantorovich–Rubinstein, optimization, ambiguity.

AMS subject classifications. 60D05, 52A05.

1. Introduction

In a recent work by Wozabal [20], a framework for optimization under ambiguity is developed –including a discussion of the history of the subject and the current literature. See also Dupačová [9] and the recent work by Esfahani and Kuhn [10], which expands Wozabal's approach to develop convex reductions for an important class of objective functions. We quote from the abstract: "Though the true distribution is unknown, existence of a reference measure P enables the construction of non-parametric ambiguity sets as Kantorovich balls around P. The original stochastic optimization problems are robustified by a worst case approach with respect to these ambiguity sets." Fundamental to the development of this framework, Wozabal [20, Cor. 1] asserts that, when the domain is a compact metric space, the extreme points of a Kantorovich ball about a measure whose support has at most n points consist of measures whose supports have at most $n+3$ points. The purpose of this paper is to extend and sharpen this result; extending the domain from a compact metric space to a Borel subset of a Polish metric space, and improving the bound on the number of Dirac masses from $n+3$ to $n+2$. In addition, we provide similar results for the Prokhorov metric and for the Monge–Wasserstein distances. This increase in generality from a compact metric space to a Borel subset of a Polish space has two nontrivial components. The first is that it replaces compactness with separability. That is, since a compact metric space is complete, it amounts to a generalization from compact complete metric spaces to separable complete metric spaces. The second is that it replaces completeness with measurability. That is, it eliminates the completeness requirement and substitutes it with the requirement that it be a Borel subset of separable complete metric space. For example, these results now apply to the case of probability measures on the (noncompact) open interval (0,1).

To outline how they are obtained, recall Rogosinski's lemma [13] that, on an arbitrary measurable space, the n moments corresponding to the expected values of n integrable functions with respect to a probability measure can be achieved by a convex sum of $n+1$ Dirac masses. Moreover, recall that an exposed point of a convex set in a locally convex space is a point which is the unique maximizer of some continuous affine

*Received: July 20, 2015; accepted (in revised form): March 10, 2016. Communicated by Shi Jin.
†California Institute of Technology 1200 E. California Blvd., Pasadena, CA 91125, USA (owhadi@caltech.edu).
‡Corresponding author, 1200 E. California Blvd., Pasadena, CA 91125, USA (clintscovel@gmail.com).

function, and Straszewicz theorem [15], that the exposed points of a finite dimensional compact convex set is dense in its extreme points. Wozabal uses the Kantorovich–Rubinstein theorem combined with Rogosinski's lemma [13] to characterize the exposed points of the Kantorovich ball about a measure whose support has at most n points to be a measure with support at most $n+3$ points. The fact that one obtains $n+3$ Dirac masses comes from the fact that Kantorovich–Rubinstein theorem introduces one function, the notion of an exposed point another, and the central measure having support of size n introduces n more functions, leading to a total of $n+2$ continuous functions on the set of probability measures on $X \times X$, so that Rogosinski's lemma implies that the exposed points are convex sums of $(n+2)+1=n+3$ Dirac masses. Then, Choquet's [5, Sec. 17, pg. 99] extension of Straszewicz' theorem [15] to compact metrizable subsets of locally convex space along with the fact that the set of probability measures equipped with the weak topology is compact and metrizable when the domain is, is used to show that these exposed points are dense in the extreme points. A limiting argument showing that the weak limit of a convex sum of $n+3$ Dirac masses is a convex sum of $n+3$ Dirac masses establishes the assertion.

In our approach, we use Dudley's [8, Thm. 11.8.2] version of the Kantorovich–Rubinstein theorem for tight measures on separable metric spaces, and characterize the extreme points of the space of measures corresponding to the Kantorovich–Rubinstein duality using results of Winkler [18, 19], previously applied in [12] to the reduction of optimization problems on *non-compact* spaces of *tight* probability measures arising in uncertainty quantification. Since, by Suslin's theorem, a Borel subset of a Polish space is Suslin and since all probability measures on Suslin spaces are tight, these results allow the extension of many results regarding the extreme points of sets of probability measures from compact metric domains and continuous moment functions to Borel subsets of Polish metric spaces and measurable moment functions. Then a fundamental result that is implicit in the results of Winkler [18, 19] is proven in Theorem 2.2; that a weakly closed convex set of probability measures on a Borel subset of a Polish metric space has an extreme point. This result combined with Lemma A.2, giving sufficient conditions that the affine image of the extreme points of a set cover the extreme points of the affine image of that set, shows that the image of these extreme points in the dual cover the extreme points of the Kantorovich ball. This latter approach has the advantage that it does not pass through the intermediate stage of exposed points, so does not add an additional function, and does not require a generalization of Straszewicz' theorem [15] to non-compact sets, although it does suggest that such a generalization may exist for weakly closed convex sets of tight measures.

To establish our main result, Theorem 2.1, we develop a more general and expressive result in Theorem 2.3, which not only produces a similar result for the Monge–Wasserstein metric, but its Corollary 3.1 shows how the duality results of Kantorovich–Rubinstein and Strassen, combined with the results of Winkler [19] on the extreme points of moment constraints, facilitate a Monge–Wasserstein linear programming representation of supersets of the extreme points which can be used for convex maximization over the Kantorovich or Prokhorov ball about a measure whose support has at most n points. A stronger application of Winkler [19, Thm. 2.1] is then used to more fully develop these representations in Section 3 towards the goal of their efficient computation. Finally, in Section 4 we consider when the central measure is an empirical measure.

2. Main results

For a metric space (X,d), the Prokhorov metric d_{Pr} on the space $\mathcal{M}(X)$ of Borel probability measures is defined by

$$d_{Pr}(\mu_1,\mu_2) := \inf\{\epsilon : \mu_1(A) \leq \mu_2(A^\epsilon)+\epsilon, A\in\mathcal{B}(X)\}, \quad \mu_1,\mu_2\in\mathcal{M}(X), \tag{2.1}$$

where

$$A^\epsilon := \{x'\in X : d(x,x') < \epsilon \text{ for some } x\in A\}.$$

According to Dudley [8, Thm. 11.3.3], when X is separable the Prokhorov metric metrizes weak convergence. Note that this definition produces the same metric if we were to use the "closed" inflated sets $A^\epsilon := \{x'\in X : d(x,x') \leq \epsilon \text{ for some } x\in A\}$ instead. On the other hand, the Kantorovich distance d_K on the space $\mathcal{M}(X)$ of Borel probability measures on a separable metric space X is defined as follows (see Vershik [16] for a historical review): Let

$$\|f\|_L := \sup_{x_1\neq x_2}\frac{|f(x_1)-f(x_2)|}{d(x_1,x_2)}$$

denote the Lipschitz norm of a real valued function on X. Then the Kantorovich distance is defined by

$$d_K(\mu_1,\mu_2) := \sup_{\|f\|_L\leq 1}\int f d(\mu_1-\mu_2). \tag{2.2}$$

According to the remark after [8, Lem. 11.8.3], d_K is an extended metric on $\mathcal{M}(X)$. Let $\Delta_n(X)\subset\mathcal{M}(X)$ denote the set of probability measures whose supports have at most n points, and let $\text{ext}(A)$ denote the set of extreme points of a set A. We can now state our result for the Prokhorov metric and Kantorovich extended metric. For either of these ($\hat{d} := d_K$ or $\hat{d} := d_{Pr}$, respectively), for $\mu\in\mathcal{M}(X)$ we define $B_\epsilon(\mu_n) := \{\mu'\in\mathcal{M}(X) : \hat{d}(\mu',\mu)\leq\epsilon\}$.

THEOREM 2.1. *Let X be a Borel subset of a Polish metric space and consider the space $\mathcal{M}(X)$ of Borel probability measures equipped with the Prokhorov metric or the Kantorovich extended metric. For $n\in\mathbb{N}$, $\epsilon>0$ and $\mu_n\in\Delta_n(X)$, consider the ball $B_\epsilon(\mu_n)$ about the measure μ_n. Then*

$$\text{ext}\big(B_\epsilon(\mu_n)\big)\subset\Delta_{n+2}(X).$$

Our path to Theorem 2.1 requires the development of more useful results which we now describe. At the heart of the matter is a result of Winkler regarding the existence of extreme points of closed convex sets of probability measures that is implicit in the results of Winkler [18, 19]. Since this result is more modest than Winkler's goal of developing integral representations, the proof we present appears somewhat simpler, in particular it is different in that it does not utilize Lusin's theorem.

THEOREM 2.2 (Winkler). *Let X be a Borel subset of a Polish metric space and consider the set $\mathcal{M}(X)$ of probability measures equipped with the weak topology. Then every nontrivial closed convex subset of $\mathcal{M}(X)$ has an extreme point.*

Winkler's Theorem 2.2 is fundamental in the proof of our second main result, Theorem 2.3, regarding the extreme points of the Monge–Wasserstein distance. This result combined with the duality results of Strassen and Kantorovich–Rubinstein are then used to establish Theorem 2.1. Moreover, in Section 3, Corollary 3.1 to Theorem 2.3 establishes the main results to be used towards the computation of supersets of the extreme points $\mathrm{ext}\big(B_\epsilon(\mu_n)\big)$, useful for convex maximization, in particular linear programming, over the ball $B_\epsilon(\mu)$.

For any two probability measures $\mu_1,\mu_2 \in \mathcal{M}(X)$, let $M(\mu_1,\mu_2) \subset \mathcal{M}(X\times X)$ denote those probability measures with marginals μ_1 and μ_2. Then for a non-negative lower semicontinuous real-valued cost function $c: X\times X \to \mathbb{R}$, the Monge–Wasserstein distance d_W on $\mathcal{M}(X)$ is defined by

$$d_W(\mu_1,\mu_2) := \inf_{\nu\in M(\mu_1,\mu_2)} \int c(x,x')d\nu(x,x').$$

Let $P_1 : \mathcal{M}(X\times X) \to \mathcal{M}(X)$ denote the marginal map corresponding to the first component and P_2 the marginal map with respect to the second component.

THEOREM 2.3. *Let X be a Borel subset of a Polish metric space and $c: X\times X \to \mathbb{R}$ a non-negative real-valued lower semicontinuous function. For $n\in\mathbb{N}$, $\epsilon>0$, and $\mu_n \in \Delta_n(X)$, consider the subset*

$$\Gamma_{\mu_n,\epsilon} := \{\nu \in \mathcal{M}(X\times X) : P_1\nu = \mu_n, \int c(x,x')d\nu(x,x') \le \epsilon\}.$$

Then

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon}) \subset \Delta_{n+2}(X\times X)$$

and

$$P_2\big(\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\big) \supset \mathrm{ext}\big(P_2(\Gamma_{\mu_n,\epsilon})\big).$$

In particular, we have

$$\mathrm{ext}\big(P_2(\Gamma_{\mu_n,\epsilon})\big) \subset \Delta_{n+2}(X).$$

3. Computation of supersets

Now we show how the duality results of Strassen and Kantorovich–Rubinstein combined with Theorem 2.3 can be used in the computation of supersets of the extreme points of $B_\epsilon(\mu_n)$. To begin we introduce some terminology. We say that a set B is a *superset for $B_\epsilon(\mu_n)$* if

$$\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \subset B \subset B_\epsilon(\mu_n). \tag{3.1}$$

For any function F which achieves its maximum at the extreme points, that is

$$\max_{\mu\in B_\epsilon(\mu_n)} F(\mu) = \max_{\mu\in \mathrm{ext}(B_\epsilon(\mu_n))} F(\mu),$$

it follows that

$$\max_{\mu\in B_\epsilon(\mu_n)} F(\mu) = \max_{\mu\in B} F(\mu)$$

for any superset B for $B_\epsilon(\mu_n)$. Consequently, efficiently constructed supersets facilitate the efficient solution to optimization problems over $B_\epsilon(\mu_n)$. To fix terms, we restrict our attention to the Prokhorov case, the Kantorovich case being essentially the same. For fixed $\epsilon > 0$ and $\mu_n \in \Delta_n$, let us consider the Prokhorov ball $B_\epsilon(\mu_n)$. Then it is clear that, since $\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \subset B_\epsilon(\mu_n)$, we obtain from Theorem 2.1 that

$$\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \subset B_\epsilon(\mu_n) \cap \Delta_{n+2}(X).$$

Moreover, since $\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \subset \partial B_\epsilon(\mu_n)$, where $\partial B_\epsilon(\mu_n) := \{\mu \in \mathcal{M}(X) : d_{Pr}(\mu,\mu_n) = \epsilon\}$ is the sphere, we also conclude that

$$\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \subset \partial B_\epsilon(\mu_n) \cap \Delta_{n+2}(X).$$

However, these supersets may be difficult to compute, so we look to Theorem 2.3 for sets generated by linear programming. To that end, write $\{d > \epsilon\}$ for the subset of elements $(x,y) \in X \times X$ such that $d(x,y) > \epsilon$, and consider the subset $\Gamma_{\mu_n,\epsilon} \subset \mathcal{M}(X \times X)$ defined in the proof of Theorem 2.1 by

$$\Gamma_{\mu_n,\epsilon} := \Big\{\nu \in \mathcal{M}(X \times X) : \nu\{d > \epsilon\} \le \epsilon, P_1\nu = \mu_n\Big\}.$$

The proof of Theorem 2.1 used Strassen's theorem to assert in (B.8) that

$$P_2\big(\Gamma_{\mu_n,\epsilon}\big) = B_\epsilon(\mu_n).$$

Then Theorem 2.3 implies

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon}) \subset \Delta_{n+2}(X \times X) \tag{3.2}$$

and the string of inequalities

$$\begin{aligned}\mathrm{ext}\big(B_\epsilon(\mu_n)\big) &= \mathrm{ext}\big(P_2(\Gamma_{\mu_n,\epsilon})\big)\\ &\subset P_2\big(\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\big)\\ &\subset \Delta_{n+2}(X).\end{aligned}$$

Consequently, we obtain the following.

COROLLARY 3.1. *Consider the situation of Theorem 2.1 and the set* $\Gamma_{\mu_n,\epsilon}$ *defined in Theorem 2.3 by* $c := d$ *in the Kantorovich case and* $c := \mathbb{1}_{d>\epsilon}$ *in the Prokhorov case. Then we have*

$$\begin{array}{lcl}\mathrm{ext}\big(B_\epsilon(\mu_n)\big) \;\subset & P_2\big(\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\big) & \subset\; B_\epsilon(\mu_n) \cap \Delta_{n+2}(X)\\ \mathrm{ext}\big(B_\epsilon(\mu_n)\big) \;\subset & P_2\big(\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X \times X)\big) & \subset\; B_\epsilon(\mu_n) \cap \Delta_{n+2}(X).\end{array}$$

The statement of Corollary 3.1 captures the mechanism by which we obtain the improvement from $n+3$ to $n+2$ Dirac masses in the description of the extreme points in Theorem 2.1. Indeed, since the set $\Gamma_{\mu_n,\epsilon}$ is a set of measures subject to $n+1$ constraints, its extreme points are a convex combination of $n+2$ Dirac masses on the product space $X \times X$. Then the fact that the extreme points of $B_\epsilon(\mu_n)$ consists of the convex combination of $n+2$ Dirac masses follows from the fact that Corollary 3.1 implies that the projection onto the second component of these extreme points covers all the extreme points of $B_\epsilon(\mu_n)$, and the fact that projection of Dirac masses on $X \times X$ are Dirac masses on X.

Corollary 3.1 also says that both

$$P_2\big(\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X \times X)\big) \quad \text{and} \quad P_2\big(\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\big)$$

are supersets for $B_\epsilon(\mu_n)$. Although the latter is smaller in that

$$P_2\big(\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\big) \subset P_2\big(\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X\times X)\big),$$

the computation of the former is useful in the computation of the latter, so we consider the computation of both.

3.1. Computing $\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X\times X)$. Since, by Equation (3.2), both $\mathrm{ext}(\Gamma_{\mu_n,\epsilon})$ and $\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X\times X)$ are subsets of $P_1^{-1}\mu_n \cap \Delta_{n+2}(X\times X)$, it will be convenient to compute $P_1^{-1}\mu_n \cap \Delta_{n+2}(X\times X)$ first. Let us proceed inductively, and assume that $\mu_n \in \Delta_n(X)$ but is not in $\Delta_{n-1}(X)$. Then $\mu_n := \sum_{i=1}^n \beta_i \delta_{y_i}$ with $\beta_i > 0, y_i \in X, i=1,\dots,n$, $\sum_{i=1}^n \beta_i = 1$, and $y_i \neq y_j, i\neq j$. Fixing this $y=(y_i)$ and (β_i), we now define some subsets of $\mathcal{M}(X\times X)$. For $x\in X^m, n\le m\le n+2$, denote

$$\delta_{y,x} := \sum_{k=1}^n \beta_k \delta_{y_k,x_k},$$

and let

$$\Pi_0 := \big\{\delta_{y,x} x \in X^n\big\}. \tag{3.3}$$

For $i=1,\dots,n$ and $x\in X^{n+1}$, define

$$\Pi_i(x) := \delta_{y,x} + \big\{\gamma(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}), 0<\gamma<\beta_i\big\} \tag{3.4}$$

and

$$\Pi_i := \{\Pi_i(x), x\in X^{n+1}\}. \tag{3.5}$$

Moreover, for $x\in X^{n+2}$ and for $i<j$, define

$$\Pi_{i,j}(x) := \delta_{y,x} + \Big\{\gamma_i(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}) + \gamma_j(\delta_{y_j,x_{n+2}} - \delta_{y_j,x_j}), 0<\gamma_i<\beta_i, 0<\gamma_j<\beta_j\Big\} \tag{3.6}$$

while for $i=j$, define

$$\Pi_{i,i}(x) := \delta_{y,x} + \Big\{\gamma_1(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}) + \gamma_2(\delta_{y_i,x_{n+2}} - \delta_{y_i,x_i}), \gamma_1>0, \gamma_2>0, \gamma_1+\gamma_2<\beta_i\Big\} \tag{3.7}$$

and then, for $i\le j$, again take the union

$$\Pi_{i,j} := \{\Pi_{i,j}(x), x\in X^{n+2}\}. \tag{3.8}$$

Lemma 3.1. *In terms of the sets defined in Equations* (3.3)*,* (3.5)*, and* (3.8)*, we have*

$$P_1^{-1}\mu_n \cap \Delta_{n+2}(X\times X) = \Pi_0 \cup_{k=1}^n \Pi_k \cup_{i\le j} \Pi_{i,j}.$$

Using Lemma 3.1, we can now obtain an almost explicit representation of $\Gamma_{\mu_n,\epsilon} \cap \Delta_{n+2}(X\times X)$, almost in the sense that it will amount to an explicitly represented set subject to the constraint of a single explicitly computable function. To that end, let us combine the definitions (3.3), (3.5), and (3.8) of Π_0, Π_i, and $\Pi_{i,j}$ into one symbol with the introduction of a multi-index $\imath$ that can take the values $\imath=0$, $\imath=i$ for $i\in\{1,n\}$, or $\imath=(i,j)$ with $i\le j$. Then, in this notation $\Pi_\imath(x)$ will denote $\Pi_0(x)$ and imply $x\in X^n$

when $\imath=0$, it will denote $\Pi_i(x)$ and imply $x\in X^{n+1}$ when $\imath=i$, and denote $\Pi_{i,j}(x)$ and imply $x\in X^{n+2}$ when $\imath=(i,j)$.

Since, in general, for $\nu:=\sum_{k=1}^m \alpha_k\delta_{x_k,x'_k}$ we have

$$\nu\{d>\epsilon\}=\sum_{k=1}^m \alpha_k \mathbb{1}_{d(x_k,x'_k)>\epsilon}, \tag{3.9}$$

it follows that the function $\nu\mapsto\nu\{d>\epsilon\}$ restricted to $\Delta_{n+2}(X\times X)$ is explicitly computable. Then, since

$$\Gamma_{\mu_n,\epsilon}\cap\Delta_{n+2}(X\times X)=P_1^{-1}\mu_n\cap\Delta_{n+2}(X\times X)\cap\big\{\nu\in\mathcal{M}(X\times X):\nu\{d>\epsilon\}\le\epsilon\big\}, \tag{3.10}$$

if we incorporate the constraint $\nu\{d>\epsilon\}\le\epsilon$ by defining

$$\bar{\Pi}_\imath(x):=\Pi_\imath(x)\cap\big\{\nu\in\mathcal{M}(X\times X):\nu\{d>\epsilon\}\le\epsilon\big\}, \tag{3.11}$$

along with their unions $\bar{\Pi}_\imath$ over X^n, X^{n+1}, and X^{n+2}, respectively, then, from the distributive law of set theory, Lemma 3.1, and Equation (3.10), we conclude that

$$\Gamma_{\mu_n,\epsilon}\cap\Delta_{n+2}(X\times X)=\bar{\Pi}_0\cup_{k=1}^n\bar{\Pi}_k\cup_{i\le j}\bar{\Pi}_{i,j}. \tag{3.12}$$

3.2. Computing $\mathrm{ext}(\Gamma_{\mu_n,\epsilon})$**.** To compute $\mathrm{ext}(\Gamma_{\mu_n,\epsilon})$ we use a stronger version of the characterization of the extreme points found in Winkler [19, Thm. 2.1] than we used in Theorem 2.3, along with the computation of $P_1^{-1}\mu_n\cap\Delta_{n+2}(X\times X)$ from Lemma 3.1. To that end, consider the constraint functions $f_i:=\mathbb{1}_{y_i\times X}, i=1,\dots,n$ (where $\mathbb{1}_{y_i\times X}(a,b)=1$ if $a=y_i$ and $\mathbb{1}_{y_i\times X}(a,b)=0$ if $a\ne y_i$) and $f_{n+1}:=\mathbb{1}_{d>\epsilon}$. Then Winkler's [19, Thm. 2.1] assertion

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\subset\Big\{\nu\in\Gamma_{\mu_n,\epsilon}:\nu=\sum_{i=1}^m\alpha_i\delta_{x_i,x'_i}, 1\le m\le n+2, \alpha_i>0, x_i,x'_i\in X, i=1,\dots,m,$$
$$\text{the vectors } \big(f_1(x_i,x'_i),\dots,f_{n+1}(x_i,x'_i),1\big), i=1,\dots,m \text{ are linearly independent}\Big\}$$

amounts to

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\subset\Big\{\nu\in\Gamma_{\mu_n,\epsilon}:\nu=\sum_{i=1}^m\alpha_i\delta_{x_i,x'_i}, 1\le m\le n+2, \alpha_i>0, x_i,x'_i\in X, i=1,\dots,m, \tag{3.13}$$
$$\text{the vectors } \big(\mathbb{1}_{y_1}(x_i),\dots,\mathbb{1}_{y_n}(x_i),\mathbb{1}_{d(x_i,x'_i)>\epsilon},1\big), i=1,\dots,m \text{ are linearly independent}\Big\}.$$

Since Theorem 2.3 asserts that $\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\subset\Delta_{n+2}(X\times X)$, it follows that we can replace $\Gamma_{\mu_n,\epsilon}$ by $\Gamma_{\mu_n,\epsilon}\cap\Delta_{n+2}(X\times X)$ in the right-hand side of Equation (3.13). Having done so, let us define

$$\bar{\Theta}:=\Big\{\nu\in\Gamma_{\mu_n,\epsilon}\cap\Delta_{n+2}(X\times X):\nu=\sum_{i=1}^m\alpha_i\delta_{x_i,x'_i}, 1\le m\le n+2, \alpha_i>0, x_i,x'_i\in X, i=1,\dots,m,$$
$$\text{the vectors } \big(\mathbb{1}_{y_1}(x_i),\dots,\mathbb{1}_{y_n}(x_i),\mathbb{1}_{d(x_i,x'_i)>\epsilon},1\big), i=1,\dots,m \text{ are linearly independent}\Big\}. \tag{3.14}$$

to be the right-hand side of Equation (3.13). Then we have

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon}) \subset \bar{\Theta} \subset \Gamma_{\mu_n,\epsilon}$$

and therefore $\bar{\Theta}$ is a superset for $\Gamma_{\mu_n,\epsilon}$. To compute it, for $i\in\{1,\dots,n\}$, let us define

$$\Lambda_i := \{x\in X^{n+1} : \mathbb{1}_{d(y_i,x_{n+1})>\epsilon} \neq \mathbb{1}_{d(y_i,x_i)>\epsilon}\}. \tag{3.15}$$

and for $i<j$ define

$$\Lambda_{i,j} := \{x\in X^{n+2} : \mathbb{1}_{d(y_i,x_{n+1})>\epsilon} \neq \mathbb{1}_{d(y_i,x_i)>\epsilon}, \mathbb{1}_{d(y_j,x_{n+2})>\epsilon} \neq \mathbb{1}_{d(y_j,x_j)>\epsilon}\}. \tag{3.16}$$

LEMMA 3.2. *With Λ_i defined in Equation* (3.15), *$\Lambda_{i,j}$ defined in Equation* (3.16), *and $\bar{\Pi}_0$, $\bar{\Pi}_i$, and $\bar{\Pi}_{i,j}$ defined in Equation* (3.11), *we have*

$$\bar{\Theta} = \bar{\Pi}_0 \cup_{k=1}^n (\bar{\Pi}_i \cap \Lambda_i) \cup_{i<j} (\bar{\Pi}_{i,j} \cap \Lambda_{i,j}).$$

REMARK 3.1. For a reference measure $\mu := \sum_{k=1}^n \beta_k \delta_{y_k}$, it is interesting to note that the condition that a measure

$$\delta_{y,x} + \{\gamma(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}), 0<\gamma<\beta_i\}$$

is a member of $\Pi_i \cap \Lambda_i$ amounts to the splitting off of the mass β_i on the Dirac situated at y_i into the convex sum of two Dirac masses, one situated at (y_i,x_i) and one at (y_i,x_{n+1}), such that, between x_i and x_{n+1}, one is inside the ball of radius ϵ about y_i and the other is outside it. Moreover, to be a member of $\Pi_{i,j}$ with $i<j$ amounts to two such splits.

3.3. Equivalence classes determined by the adjacency matrix. For $x\in X^m, n\le m\le n+2$, let its adjacency matrix $A(x)$ be defined by

$$A^{i,j}(x) := \mathbb{1}_{d(y_i,x_j)>\epsilon}, \quad i=1,\dots,n, j=1,\dots,m.$$

Commensurate with our introduction of the multi-index $\imath$, we use the expression $A(x)$ to mean the $n\times m$ adjacency matrix when $x\in X^m$, for any $m=n,n+1,n+2$. Since, by Lemma 3.2, $\bar{\Theta} = \bar{\Pi}_0 \cup_{k=1}^n \bar{\Pi}_k \cup_{i\le j} \bar{\Pi}_{i,j}$ and the latter are determined by conditions $\Lambda_i, i=1,\dots,n$, $\Lambda_{i,j}$ for $i<j$, and $\nu\{(z,z')\in X\times X : d(z,z')>\epsilon\}\le\epsilon$, all of which, by the the evaluation (3.9), only depend on the values of the adjacency matrix, we obtain the following lemma. It asserts that, for any point in $\bar{\Pi}_0$, $\bar{\Pi}_i$ or $\bar{\Pi}_{i,j}$, if the second components x of the Dirac masses are changed to x' with the same adjacency matrix, then the resulting sum of Dirac masses remains in $\bar{\Pi}_0$, $\bar{\Pi}_i$ or $\bar{\Pi}_{i,j}$ respectively. Consequently, it will be useful in the efficient exploration of the set $\bar{\Theta}$.

LEMMA 3.3. *For $n\le m\le n+2$, $x\in X^m$, $z\in X^m$ and $\alpha\in\mathbb{R}^m$, consider $\mu(x) := \sum_{k=1}^m \alpha_k \delta_{z_k,x_k}$. If $\mu(x)\in\bar{\Pi}_\imath(x)$, then for all x' such that $A(x')=A(x)$, we have $\mu(x')\in\bar{\Pi}_\imath(x')$.*

4. Extreme points of a ball about an empirical measure

Empirical measures take the form $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$, with $y_i\in X, i=1,\dots,n$. When all the points y_i are unique, we can define $\beta_i := \frac{1}{n}, i=1,\dots,n$ in the expressions of Section 3, when the points have duplicates things will be more complicated. In the unique case, the definitions (3.3), (3.4), (3.6), and (3.7) of Π_0, $\Pi_i(x)$, and $\Pi_{i,j}(x)$ take on a more symmetrical form, and since the case when the central measure is an empirical measure is an important application, we spell them out. To begin with, we have

$$\delta_{y,x} = \frac{1}{n}\sum_{k=1}^{n} \delta_{y_k,x_k}.$$

Moreover, the evaluation of the constraint $\nu(d > \epsilon) \le \epsilon$ also takes a simpler form, so that constrained sets $\bar{\Pi}_0$, $\bar{\Pi}_i(x)$, and $\bar{\Pi}_{i,j}(x)$ appear as follows:

$$\bar{\Pi}_0 = \{\delta_{y,x}, x \in X^n\}$$

subject to the constraint

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{d(y_k,x_k)>\epsilon} \le \epsilon,$$

while for $i \in \{1,\dots,n\}$ we have

$$\bar{\Pi}_i(x) = \delta_{y,x} + \frac{1}{n}\{\gamma(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}), 0 < \gamma < 1\}$$

subject to the constraint

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{d(y_k,x_k)>\epsilon} + \gamma(\mathbb{1}_{d(y_i,x_{n+1})>\epsilon} - \mathbb{1}_{d(y_i,x_i)>\epsilon}) \le \epsilon,$$

and for $i < j$ we have

$$\bar{\Pi}_{i,j}(x) = \delta_{y,x} + \frac{1}{n}\Big\{\gamma_i(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}) + \gamma_j(\delta_{y_j,x_{n+2}} - \delta_{y_j,x_j}), 0 < \gamma_i < 1, 0 < \gamma_j < 1\Big\}$$

subject to the constraint

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{d(y_k,x_k)>\epsilon} + \gamma_i(\mathbb{1}_{d(y_i,x_{n+1})>\epsilon} - \mathbb{1}_{d(y_i,x_i)>\epsilon}) + \gamma_j(\mathbb{1}_{d(y_j,x_{n+2})>\epsilon} - \mathbb{1}_{d(y_j,x_j)>\epsilon}) \le \epsilon,$$

and for $i = j$ we have

$$\bar{\Pi}_{i,i}(x) = \delta_{y,x} + \Big\{\gamma_1(\delta_{y_i,x_{n+1}} - \delta_{y_i,x_i}) + \gamma_2(\delta_{y_i,x_{n+2}} - \delta_{y_i,x_i}), \gamma_1 > 0, \gamma_2 > 0, \gamma_1 + \gamma_2 < 1\Big\}$$

subject to the constraint

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{d(y_k,x_k)>\epsilon} + \gamma_1(\mathbb{1}_{d(y_i,x_{n+1})>\epsilon} - \mathbb{1}_{d(y_i,x_i)>\epsilon}) + \gamma_2(\mathbb{1}_{d(y_i,x_{n+2})>\epsilon} - \mathbb{1}_{d(y_i,x_i)>\epsilon}) \le \epsilon.$$

Appendix A.

A.1. Extreme subsets. We begin by establishing a fundamental identity regarding the extreme subsets of extreme subsets[1] of an affine space. Since this terminology varies in the literature, we fix it now. Following [2, Def. 7.61], we say that a set E is an *extreme subset* of a subset $A \subset L$ of a real linear space L if $E \subset A$ and $\theta x + (1-\theta)y \in E$ with $x, y \in A, \theta \in (0,1)$ implies that $x, y \in E$. Note that this definition does not require convexity. An *extreme point* of A is an extreme subset of A consisting of a single point. We say that a set F is a *face* of a subset $A \subset L$ of a real linear space L if it is a convex extreme subset of A. The following lemma implies that Simon [14, Prop. 8.6] is valid *without assuming compactness or convexity.*

Lemma A.1. *Let A be a subset of a real linear space L and let E be an extreme subset of A. Then B is an extreme subset of E if and only if $B \subset E$ and it is an extreme subset of A. In particular,*

$$\operatorname{ext}(E) = E \cap \operatorname{ext}(A).$$

Proof. The proof is identical to that of [14, Prop. 8.6], but we reproduce it here so that the reader can confirm that it is valid without compactness or convexity assumptions. First suppose that $B \subset E$ and B is an extreme subset of A. Then, by definition, if $\theta x + (1-\theta)y \in B$, with $x, y \in A, \theta \in (0,1)$, then $x, y \in B$. Since $E \subset A$, it follows that if we have $\theta x + (1-\theta)y \in B$, with $x, y \in E, \theta \in (0,1)$, that $x, y \in B$. Consequently, since $B \subset E$, B is an extreme subset of E. Now assume that B is an extreme subset of E. Then, if we have $\theta x + (1-\theta)y \in B$, with $x, y \in A, \theta \in (0,1)$, the fact that $B \subset E$ and E is an extreme subset of A implies that $x, y \in E$. Then, since B is an extreme subset of E, it follows that $x, y \in B$. Since clearly $B \subset A$, we conclude that B is an extreme subset of A. ∎

A.2. Affine images of extreme points. Here we establish a fundamental result for affine transformations and extreme points of, possibly non-convex, subsets.

Lemma A.2. *Let L and L' be real linear spaces and $K \subset L$ a subset. Suppose that $G : K \to L'$ is the restriction of an affine transformation $G : L \to L'$ to K such that $\operatorname{ext}(G^{-1}(k')) \neq \emptyset$ for all $k' \in \operatorname{ext}(G(K))$. Then $G(\operatorname{ext}(K)) \supset \operatorname{ext}(G(K))$.*

Proof. Let $k' \in \operatorname{ext}(G(K))$ and consider any point $k \in G^{-1}(k')$. Then if $k = \theta k_1 + (1-\theta)k_2$, with $k_1, k_2 \in K, \theta \in (0,1)$, then $k' = G(k) = G(\theta k_1 + (1-\theta)k_2) = \theta G(k_1) + (1-\theta)G(k_2)$, so that, since k' is an extreme point, it follows that $G(k_1) = G(k_2) = G(k)$. That is, $G^{-1}(k')$ is an extreme subset of K. Therefore, Lemma A.1 implies that

$$\operatorname{ext}\big(G^{-1}(k')\big) = G^{-1}(k') \cap \operatorname{ext}(K),$$

so that any extreme point of $G^{-1}(k')$ is an extreme point of K. Since, by assumption, $G^{-1}(k')$ has an extreme point, it follows that any such extreme point is an extreme point of K. Since the image under G of any such point is k', and $k' \in \operatorname{ext}(G(K))$ was arbitrary, the assertion follows. ∎

[1] The repetition here is not a typo.

A.3. Integrals of extended real-valued lower semicontinuous functions. Here we formulate a generalization to extended real-valued functions of [2, Thm. 15.5], that the integral of a bounded lower semicontinuous function forms a lower semicontinuous function in the weak topology.

LEMMA A.3. *Let (X,d) be a metric space and $f:X\rightarrow \bar{R}_+$ a non-negative lower semicontinuous extended real-valued function. For $\mu\in\mathcal{M}(X)$ define $\int f d\mu$ to be the integral if f is μ-integrable and ∞ if it is not. Then the function $F:\mathcal{M}(X)\rightarrow \bar{R}$ defined by $F(\mu):=\int f d\mu$ is lower semicontinuous in the weak topology.*

Proof. We follow Aliprantis and Border [2, Thm. 15.5]. First let us clip the function f at the level s by $f^s(x):=\min(f(x),s), x\in X$. Then, since for all c we have $\{x: f^s(x)\leq c\}=\{x:f(x)\leq c\}$ for $s>c$ and $\{x:f^s(x)\leq c\}=\{x:f(x)\leq s\}$ for $s\leq c$, it follows that f^s is a real-valued semicontinuous function. Consequently, by [2, Thm. 3.13] for each s, f^s is the increasing pointwise limit of a sequence f^s_n of Lipschitz continuous functions. By further clipping from below at 0, sending $f^s_n\mapsto \max(f^s_n,0)$ we obtain that we can assume that for each s, f^s is the increasing pointwise limit of a sequence f^s_n of non-negative bounded continuous functions. Therefore, setting $s:=n$ and defining $f_n:=f^n_n$, we conclude that f is the increasing pointwise limit of a sequence f_n of bounded continuous nonnegative real-valued functions.

Now let μ_α be a net such that $\mu_\alpha\rightarrow\mu$ in the weak topology and let us utilize the integration theory for extended real-valued functions as found in Ash [3, Sec. 1]. Then it follows that

$$\int f_n d\mu_\alpha \xrightarrow{\alpha} \int f_n d\mu \tag{A.1}$$

and

$$\int f_n d\mu_\alpha \leq \int f d\mu_\alpha \tag{A.2}$$

so that we conclude that

$$\int f_n d\mu \leq \liminf_{\alpha} \int f d\mu_\alpha,$$

for each n. Therefore, from the monotone convergence theorem for extended valued functions (see e.g. Ash [3, 1.6.2]) we have

$$\int f d\mu = \lim_{n\rightarrow\infty} \int f_n d\mu$$

and we conclude that

$$\int f d\mu = \lim_{n\rightarrow\infty} \int f_n d\mu \leq \liminf_{\alpha} \int f d\mu_\alpha,$$

so that the assertion follows from the alternative characterization of lower semicontinuous extended real-valued functions [2, Lem. 2.42]. ∎

Appendix B.

B.1. Proof of Theorem 2.2. We follow the proof of the main result in [18], simplifying it according to our more modest goal. Let t denote the topology of X. Since X is a Borel subset of a Polish space, it follows that it is Suslin and, therefore, all finite Borel measures on (X,t) are tight. Let $C \subset \mathcal{M}(X)$ be a nontrivial closed convex subset and consider $\mu^* \in C$. Since μ^* is tight, using a recursive argument, we obtain a sequence $K_n \subset X, n \in \mathbb{N}$ of disjoint compact subsets such that if we define $X_1 := \cup_{n\in\mathbb{N}} K_n$ we have $\mu^*(X_1) = 1$. Let the relative topology of the subspace $X_1 \subset X$ be denoted by t_0 and introduce a finer topology $t_1 \supset t_0$ defined by $A \in t_1$ if, for every $n \in \mathbb{N}$, we have $A \cap K_n = B_n \cap K_n$ for some $B_n \in t$. It follows that $K_n \in t_1$ for all $n \in \mathbb{N}$, so that (X_1,t_1) is locally compact. Moreover, since (X_1,t_0) is metric, it is Hausdorff, and since t_1 is finer than t_0 it follows that (X_1,t_1) is Hausdorff. Let us show that (X_1,t_1) is also completely regular. To that end, recall (see e.g. Willard [17, Thm. 14.12]) that a space is completely regular if and only if its topology is the initial topology corresponding to the bounded continuous functions. Since (X_1,t_0) is metric it is completely regular. Consequently the topology t_1 amounts to the initial topology corresponding to the addition of the set of indicator functions $\mathbb{1}_{K_n}, n \in \mathbb{N}$ to the collection of continuous functions on (X_1,t_0). Therefore, (X_1,t_1) is also completely regular. Since (X,t) is Suslin it is second countable and therefore (X_1,t_0) is second countable. Since a base for the topology t_1 can be constructed by taking a base for (X_1,t_0) and taking all intersections with the sets $K_n, n \in \mathbb{N}$, it follows that (X_1,t_1) is second countable. Consequently, all the spaces (X,t), (X_1,t_0), and (X_1,t_1) are second countable.

Now observe that for $A \in t_1$ we have $A = \cup_{n\in\mathbb{N}} A \cap K_n$ and for each n, we have $A \cap K_n = B_n \cap K_n$ for some $B_n \in t$. Since both B_n and K_n are in $\mathcal{B}(t)$ it follows that the intersection is also and therefore also the countable union $A = \cup_{n\in\mathbb{N}} A \cap K_n$. That is, $A \in \mathcal{B}(t)$ and since $A \subset X_1$ it follows that $A \in \mathcal{B}(t_0)$. Since t_1 is finer than t_0, we conclude that

$$\mathcal{B}(t_0) = \mathcal{B}(t_1)$$

and therefore

$$\mathcal{M}(X_1,t_0) = \mathcal{M}(X_1,t_1) \tag{B.1}$$

as sets.

Since (X_1,t_1) is locally compact and Hausdorff, we consider the Alexandroff one-point compactification (X_2,t_2) of (X_1,t_1). Since (X_1,t_1) is second countable, it follows (see e.g. [2, Thm. 3.44]) that the compactification (X_2,t_2) is metrizable. Consequently, (X_2,t_2) is a compact metrizable Hausdorff space, and so it follows (see e.g. [2, Thm. 15.11]) that $\mathcal{M}(X_2,t_2)$ is compact and metrizable. Moreover, since by e.g. [2, Lem. 3.26 & Thm. 3.28], all compact metrizable spaces are separable and therefore second countable, it follows that $\mathcal{M}(X_2,t_2)$ is second countable.

Define

$$\mathcal{M}_{X_1}(X,t) = \{\mu \in \mathcal{M}(X,t) : \mu(X_1) = 1\}$$

$$\mathcal{M}_{X_1}(X_2,t_2) = \{\mu \in \mathcal{M}(X_2,t_2) : \mu(X_1) = 1\}$$

where $X_1 \subset X_2$ is the subset identification corresponding to the compactification. Since both $\mathcal{M}(X,t)$ and $\mathcal{M}(X_2,t_2)$ are second countable, it follows that the subspaces

$\mathcal{M}_{X_1}(X,t)$ and $\mathcal{M}_{X_1}(X_2,t_2)$ are second countable. Since (X_2,t_2) is compact and Hausdorff it follows from [17, Thm. 17.10 & Cor. 15.7] that (X_2,t_2) is completely regular. Consequently, if we let

$$i^0:(X_1,t_0)\to(X,t)$$
$$i^1:(X_1,t_1)\to(X_2,t_2)$$

denote the two subset injections, then since both (X_1,t_0) and (X_2,t_2) are completely regular, Bourbaki [4, Prop. 8, Sec. 5.3] implies that the pushforward maps

$$i^0_*:\mathcal{M}(X_1,t_0)\to\mathcal{M}_{X_1}(X,t),$$
$$i^1_*:\mathcal{M}(X_1,t_1)\to\mathcal{M}_{X_1}(X_2,t_2),$$

are homeomorphisms, because of the identity (B.1) it is natural to define

$$\iota:\mathcal{M}_{X_1}(X,t)\to\mathcal{M}_{X_1}(X_2,t_2)$$

by

$$\iota:=i^1_*(i^0_*)^{-1}.$$

Although each component i^0_* and i^1_* of ι is a homeomorphism, since we have $\mathcal{M}(X_1,t_0)=\mathcal{M}(X_1,t_1)$ only as sets, ι may not be a homeomorphism. However, since t_1 is finer than t_0 it follows that the identity map $\acute{\iota}:\mathcal{M}(X_1,t_1)\to\mathcal{M}(X_1,t_0)$ is continuous, and if we more properly write

$$\iota:=i^1_*(\acute{\iota})^{-1}(i^0_*)^{-1}$$

as a composition of three maps on topological spaces, it follows from the continuity of $\acute{\iota}$ and the fact that i^0_* and i^1_* are homeomorphisms that

$$\iota\text{ is a closed map.}\tag{B.2}$$

Now define

$$C_0:=C\cap\mathcal{M}_{X_1}(X,t)$$
$$C_2:=\iota C_0$$

and

$$\bar{C}_2:=\text{the closure of}\,C_2\,\text{in}\,\mathcal{M}(X_2,t_2).$$

Since ι is affine it follows that C_2 is convex. Moreover, since C_0 is relatively closed in $\mathcal{M}_{X_1}(X,t)$ and by Equation (B.2) ι is a closed map, it follows that $C_2=\iota C_0$ is relatively closed in $\mathcal{M}_{X_1}(X_2,t_2)$. Consequently, there exists a closed set $\acute{C}_2\subset\mathcal{M}(X_2,t_2)$ such that $C_2=\acute{C}_2\cap\mathcal{M}_{X_1}(X_2,t_2)$. Since it follows that $\acute{C}_2\supset C_2$ we obtain

$$C_2\subset\bar{C}_2\subset\acute{C}_2$$

and therefore

$$\begin{aligned}C_2&=C_2\cap\mathcal{M}_{X_1}(X_2,t_2)\\&\subset\bar{C}_2\cap\mathcal{M}_{X_1}(X_2,t_2)\end{aligned}$$

$$\begin{aligned}&\subset \acute{C}_2 \cap \mathcal{M}_{X_1}(X_2,t_2)\\&= C_2\end{aligned}$$

so that we conclude that

$$C_2 = \bar{C}_2 \cap \mathcal{M}_{X_1}(X_2,t_2). \tag{B.3}$$

It is easy to show that both $\mathcal{M}_{X_1}(X,t) \subset \mathcal{M}(X,t)$ and $\mathcal{M}_{X_1}(X_2,t_2) \subset \mathcal{M}(X_2,t_2)$ are extreme subsets. Therefore, it follows from Lemma A.2 that

$$\mathrm{ext}(C_0) = \mathrm{ext}(C) \cap \mathcal{M}_{X_1}(X,t) \tag{B.4}$$

and

$$\mathrm{ext}(C_2) = \mathrm{ext}(\bar{C}_2) \cap \mathcal{M}_{X_1}(X_2,t_2). \tag{B.5}$$

Since ι is a composition of affine bijections, it is an affine bijection, so that we have

$$\mathrm{ext}(C_2) = \iota\,\mathrm{ext}(C_0).$$

Finally, observe that μ^*, selected at the beginning of the proof, satisfies $\mu* \in \mathcal{M}_{X_1}(X,t)$. Therefore, it follows that C_0 and therefore $C_2 := \iota C_0$ and $\bar{C}_2$ are not empty. Consequently, since $\bar{C}_2 \subset \mathcal{M}(X_2,t_2)$ is closed and $\mathcal{M}(X_2,t_2)$ compact it follows that $\bar{C}_2$ is compact, and since $\mathcal{M}(X_2,t_2)$ is locally convex and metrizable, it follows from Choquet's theorem for metrizable compact convex sets, see Alfsen [1, Cor. I.4.9], that each element $\mu \in \bar{C}_2$ has an integral representation over the boundary $\mathrm{ext}(\bar{C}_2)$. That is, $\mathrm{ext}(\bar{C}_2) \neq \emptyset$ is measurable, and for $\mu \in \bar{C}_2$ there exists a probability measure p on $\mathrm{ext}(\bar{C}_2)$ such that, for all continuous functions f on $\bar{C}_2$, we have

$$\mu(f) = \int_{\mathrm{ext}(\bar{C}_2)} \nu(f)\,dp(\nu),$$

where $\mu(f)$ and $\nu(f)$ denote the integrals $\int f d\mu$ and $\int f d\nu$.

Consider the open subset $X_1 \subset X_2$. Since X_1 is a metric space, it follows (see e.g. [2, Cor. 3.14]) that the indicator function $\mathbb{1}_{X_1}$ is the increasing pointwise limit of a sequence of continuous functions $f_n, n \in \mathbb{N}$ with values in $[0,1]$. Since $\bar{C}_2$ is a subset of a metrizable second countable space, it too is metrizable and second countable, and therefore it follows from [2, Lem. 3.4] that it is separable. Consequently, [2, Thm. 15.13] implies that the function $\nu \mapsto \nu(f)$ is measurable for all bounded measurable functions f. Therefore, by the monotone convergence theorem [3, Thm. 1.6.2] applied three times: to the left hand side, to the integrand of the right-hand side, and to the integral on the right-hand side, we conclude that

$$\mu(X_1) = \int_{\mathrm{ext}(\bar{C}_2)} \nu(X_1)\,dp(\nu). \tag{B.6}$$

Since $C_2 \subset \bar{C}_2$, it follows that $\mu \in C_2$ has a representing measure p such that integral formula (B.6) holds. Since $\mu \in C_2$, the equality $\mu(X_1) = 1$ implies that $\nu(X_1) = 1$ p-almost everywhere. In particular, there exists a $\nu \in \bar{C}_2$ such that $\nu(X_1) = 1$. That is, $\mathrm{ext}(\bar{C}_2) \cap \mathcal{M}_{X_1}(X_2,t_2) \neq \emptyset$. Since by Equation (B.5) $\mathrm{ext}(C_2) = \mathrm{ext}(\bar{C}_2) \cap \mathcal{M}_{X_1}(X_2,t_2)$ it follows that $\mathrm{ext}(C_2) \neq \emptyset$. Furthermore, the relation $\iota\,\mathrm{ext}(C_0) = \mathrm{ext}(C_2)$ implies that $\mathrm{ext}(C_0) \neq \emptyset$, and the relation $\mathrm{ext}(C_0) = \mathrm{ext}(C) \cap \mathcal{M}_{X_1}(X,t)$ implies that $\mathrm{ext}(C) \neq \emptyset$, which is the assertion of the theorem.

B.2. Proof of Theorem 2.3. It is straightforward to show that $X\times X$ is a Borel subset of the Polish metric space determined by the product of the ambient Polish metric spaces. Therefore, Suslin's theorem (see e.g. Kechris [11, Thm. 14.2]) implies that both X and $X\times X$ are Suslin, and therefore by Dellacherie and Meyer [6, III.69], it follows that all probability measures in both $\mathcal{M}(X)$ and $\mathcal{M}(X\times X)$ are tight. This tightness facilitates both the existence of extreme points for convex sets of measures, useful in obtaining the assertion, and the duality theorems of Strassen and Kantorovich–Rubinstein used in the proof of Theorem 2.1.

Lemma A.3 implies that $\{\nu\in\mathcal{M}(X\times X): \int c(x,x')d\nu(x,x')\le\epsilon\}$ is closed and convex in the weak topology. Moreover, by Aliprantis and Border [2, Thm. 15.14] the marginal maps P_1 and P_2 are continuous in the weak topologies. Since singletons in $\mathcal{M}(X)$ are closed, for $\mu\in\mathcal{M}(X)$, it follows that $\{\nu\in\mathcal{M}(X\times X): P_1\nu=\mu_n\}$, $\{\nu\in\mathcal{M}(X\times X): P_2\nu=\mu\}$ are also closed and convex, and therefore $\Gamma_{\mu_n,\epsilon}\cap P_2^{-1}\mu$ is closed and convex in the weak topology. Since $\Gamma_{\mu_n,\epsilon}\cap P_2^{-1}\mu$ is nonemtpy, Winkler's theorem 2.2 implies that it possesses an extreme point. Therefore Lemma A.2 implies that

$$P_2(\mathrm{ext}(\Gamma_{\mu_n,\epsilon}))\supset \mathrm{ext}\big(P_2(\Gamma_{\mu_n,\epsilon})\big),$$

establishing the second assertion.

For the first, let us describe $\mathrm{ext}(\Gamma_{\mu_n,\epsilon})$. To that end, write $\mu_n=\sum_{i=1}^n\alpha_i\delta_{x_i}$ with $\alpha_i\ge 0, x_i\in X, i=1,\dots,n$ and $\sum_{i=1}^n\alpha_i=1$. Then consider the $n+1$ constraint functions c and $\mathbb{1}_{\{x_i\}\times X}, i=1,\dots,n$ to define $\Gamma_{\mu_n,\epsilon}$ as inequality/equality constraints defined by integrals of measurable functions on $\mathcal{M}(X\times X)$. Then [12, Thm. 4.1, Rmk. 4.2] (derived from Winkler [19, Thm. 2.1], which is a consequence of Dubins [7]) implies that

$$\mathrm{ext}(\Gamma_{\mu_n,\epsilon})\subset\Delta_{n+2}(X\times X),$$

establishing the first assertion. The third assertion follows by combining the first two and $P_2\big(\Delta_{n+2}(X\times X)\big)=\Delta_{n+2}(X)$.

B.3. Proof of Theorem 2.1. Since X is a Borel subset in a Polish metric space, Suslin's theorem, see e.g. Kechris [11, Thm. 14.2], implies that X is Suslin, and therefore by Dellacherie and Meyer [6, III.69], it follows that all probability measures in $\mathcal{M}(X)$ are tight.

Let us first begin with the Prokhorov case. We use the Prokhorov metric on $\mathcal{M}(X\times X)$. Consider the subset $\Gamma_{\mu_n,\epsilon}\subset\mathcal{M}(X\times X)$ defined by

$$\Gamma_{\mu_n,\epsilon}:=\Big\{\nu\in\mathcal{M}(X\times X):\nu\{d>\epsilon\}\le\epsilon, P_1\nu=\mu_n\Big\}.$$

For any $\nu\in\Gamma_{\mu_n,\epsilon}$, for $\mu':=P_2\nu$ it follows that $P_1\nu=\mu_n$, $P_2\nu=\mu'$ and $\nu\{d>\epsilon\}\le\epsilon$, so that by the Prokhorov-Ky Fan inequality [8, Thm. 11.3.5] it follows that $d_{Pr}(\mu',\mu_n)\le\epsilon$, that is $\mu'\in B_\epsilon(\mu_n)$, so that we conclude that

$$P_2(\Gamma_{\mu_n,\epsilon})\subset B_\epsilon(\mu_n). \tag{B.7}$$

To obtain the reverse inequality, let us first note that the inf in the definition (2.1) of the Prokhorov metric can be replaced by a min. To see this, observe that for fixed $A\in\mathcal{B}(X)$, that the parametrized family of open sets $A^\epsilon,\epsilon>0$ is increasing. Consequently, if $\epsilon_n\downarrow\epsilon'$, then for any $\mu\in\mathcal{M}(X)$, we have $\mu(A^{\epsilon_n})\downarrow\mu(A^{\epsilon'})$, so that, for fixed $A\in\mathcal{B}(X)$ and $\mu_1,\mu_2\in\mathcal{M}(X)$, the interval $\{\epsilon:\mu_1(A)\le\mu_2(A^\epsilon)+\epsilon\}$ is closed. It follows that the intersection of these closed intervals $\{\epsilon:\mu_1(A)\le\mu_2(A^\epsilon)+\epsilon, A\in\mathcal{B}(X)\}$ over all $A\in\mathcal{B}(X)$ is closed. Therefore the infimum in the definition (2.1) is attained.

Now consider $\mu \in B_\epsilon(\mu_n)$ and define $\epsilon^* := d_{Pr}(\mu_n, \mu)$. Then by the previous remark we have

$$\mu(A) \le \mu_n(A^{\epsilon^*}) + \epsilon^*, \quad A \in \mathcal{B}(X)$$

and the inequality $\epsilon^* \le \epsilon$ implies that

$$\mu(A) \le \mu_n(A^{\epsilon}) + \epsilon, \quad A \in \mathcal{B}(X).$$

Moreover, if we denote $d(x,A) := \inf_{y \in A} d(x,y)$ then it is easy to see that $A^\epsilon = \{x \in X : d(x,A) < \epsilon\}$ and defining $A^{\epsilon]} = \{x \in X : d(x,A) \le \epsilon\}$ we obtain that

$$\mu(A) \le \mu_n(A^{\epsilon]}) + \epsilon, \quad A \in \mathcal{B}(X).$$

Then, since both μ and μ_n are tight, Dudley's [8, Thm. 11.6.2] extension of Strassen's theorem to tight measures on separable metric spaces implies that there exists a probability measure $\nu \in \mathcal{M}(X \times X)$ such that $P_1\nu = \mu_n$, $P_2\nu = \mu$, and $\nu\{d > \epsilon\} \le \epsilon$, that is, there exists a $\nu \in \Gamma_{\mu_n,\epsilon}$ such that $P_2\nu = \mu$, so that we obtain

$$P_2\big(\Gamma_{\mu_n,\epsilon}\big) \supset B_\epsilon(\mu_n)$$

and, so by Equation (B.7), conclude that

$$P_2\big(\Gamma_{\mu_n,\epsilon}\big) = B_\epsilon(\mu_n). \tag{B.8}$$

Since the metric d is a continuous function, it follows that the set $\{(x,x') \in X \times X : d(x,x') > \epsilon\}$ is open and therefore the indicator function $\mathbb{1}_{d>\epsilon}$ is lower semicontinuous. Therefore, we can apply Theorem 2.3 to obtain

$$\begin{aligned}\operatorname{ext}\big(B_\epsilon(\mu_n)\big) &= \operatorname{ext}\big(P_2(\Gamma_{\mu_n,\epsilon})\big)\\ &\subset \Delta_{n+2}(X)\end{aligned}$$

establishing the assertion.

Now let us consider the Kantorovich case. To that end, let $\mathcal{M}_1(X) \subset \mathcal{M}(X)$ denote those Borel probability measures μ such that $\int d(x',x)d\mu(x) < \infty$ for some $x' \in X$, and consider the Monge–Wasserstein distance d_W on $\mathcal{M}_1(X)$ defined by

$$d_W(\mu_1,\mu_2) := \inf_{\nu \in M(\mu_1,\mu_2)} \int d(x,x')d\nu(x,x').$$

Then the Kantorovich–Rubinstein theorem [8, Thm. 11.8.2] states that for all $\mu_1,\mu_2 \in \mathcal{M}_1(X)$ we have

$$d_K(\mu_1,\mu_2) = d_W(\mu_1,\mu_2),$$

and, if μ_1 and μ_2 are tight, that there is a measure in $\mathcal{M}(X \times X)$ at which the infimum in the definition of d_W is attained.

Define $\Gamma_{\mu_n,\epsilon} \subset \mathcal{M}(X \times X)$ by

$$\Gamma_{\mu_n,\epsilon} := \Big\{\nu \in \mathcal{M}(X \times X) : \int d(x,x')d\nu(x,x') \le \epsilon, P_1\nu = \mu_n\Big\},$$

and for $\nu \in \Gamma_{\mu_n,\epsilon}$, consider $\mu := P_2\nu$. Then, for $y \in X$, we have

$$\int d(y,x')d\mu(x') = \int d(y,x')d\nu(x,x')$$

$$\leq \int \big(d(y,x)+d(x,x')\big)d\nu(x,x')$$
$$= \int d(y,x)d\nu(x,x') + \int d(x,x')d\nu(x,x')$$
$$= \int d(y,x)d\mu_n(x) + \int d(x,x')d\nu(x,x')$$
$$\leq \int d(y,x)d\mu_n(x) + \epsilon,$$

and, since μ_n is a finite convex sum of Dirac masses, it follows that $\int d(y,x')d\mu(x') < \infty$, that is, $P_2\nu \in \mathcal{M}_1(X)$, so that we conclude that

$$P_2(\Gamma_{\mu_n,\epsilon}) \subset \mathcal{M}_1(X).$$

Since all measures in $\mathcal{M}_1(X)$ are tight, the Kantorovich–Rubinstein theorem then implies that

$$P_2(\Gamma_{\mu_n,\epsilon}) = B_\epsilon(\mu_n)$$

in the same way that the Strassen theorem implied it in Equation (B.8) for the Prokhorov metric. Moreover, since d is a metric, it is non-negative, real-valued and continuous, so it follows that it is a non-negative semicontinuous real-valued function. As in the Prokhorov case, Theorem 2.3 then yields the assertion.

B.4. Proof of Lemma 3.1. Since an element $\nu \in \Delta_{n+2}(X\times X)$ may have support smaller than $n+2$, we represent it by $\nu = \sum_{i=1}^m \alpha_i \delta_{x_i,x'_i}$, $\alpha_i > 0, x_i, x'_i \in X, i = 1,\dots,m, \sum_{i=1}^m \alpha_i = 1$, for $m \leq n+2$, where we also require $(x_i,x'_i) \neq (x_j,x'_j), i \neq j$. Such an element $\nu \in \Delta_{n+2}(X\times X)$ is a member of $P_1^{-1}\mu_n \cap \Delta_{n+2}(X\times X)$ if and only if $P_1\nu = \mu_n$. Therefore, we conclude that $\nu \in P_1^{-1}\mu_n \cap \Delta_{n+2}(X\times X)$ if and only if

$$\sum_{j=1}^m \alpha_j \delta_{x_j} = \sum_{i=1}^n \beta_i \delta_{y_i}.$$

Since $\beta_i > 0, i = 1,\dots,n$ and $\alpha_j > 0, j = 1,\dots,m$ it follows that

$$\{x_j, j = 1,\dots,m\} = \{y_i, i = 1,\dots,n\}.$$

In particular, m must satisfy $n \leq m \leq n+2$. Moreover, the three possible cases $m = n, n+1, n+2$ appear as follows: when $m = n$, there is a relabeling of the indices of $(x_j,x'_j), j = 1,\dots,n$ so that $x_i = y_i, \alpha_i = \beta_i, i = 1,\dots,n$. When $m = n+1$, there is a $j_1 \in \{1,\dots,n\}$ and a relabeling so that $x_i = y_i, i = 1,\dots,n$ and $x_{n+1} = y_{j_1}$. Then we also have $\alpha_i = \beta_i, i \neq j_1$ and $\alpha_{j_1} + \alpha_{n+1} = \beta_{j_1}$. When $m = n+2$, then there is a relabeling so that $x_i = y_i, i = 1,\dots,n$ and either 1) there is a $j_1 \in \{1,\dots,n\}$ such that $x_{n+1} = x_{x+2} = y_{j_1}$ and $\alpha_i = \beta_i, i \neq j_1$ and $\alpha_{j_1} + \alpha_{n+1} + \alpha_{n+2} = \beta_{j_1}$ or 2) there are two distinct values $j_1, j_2 \in \{1,\dots,n\}$ such that $x_{n+1} = y_{j_1}$, $x_{n+2} = y_{j_2}$, $\alpha_i = \beta_i, i \neq j_1 i \neq j_2$, $\alpha_{j_1} + \alpha_{n+1} = \beta_{j_1}$, and $\alpha_{j_2} + \alpha_{n+2} = \beta_{j_2}$. It is clear the $m = n$ case amounts to the statement $\nu \in \Pi_0$ defined in Equation (3.3). Let us now show that the $m = n+1$ and $m = n+2$ cases amount to the statements $\nu \in \Pi_i$ for some i and $\nu \in \Pi_{i,j}$ for some $i \leq j$ defined in Equation (3.5) and Equation (3.8), respectively, establishing the assertion.

To that end, for the $m = n+1$ case, the above assertion states that there is an $i \in \{1,\dots,n\}$ and an $x \in X^{n+1}$ such that

$$\nu = \sum_{k\neq i, k \in \{1,n\}} \beta_k \delta_{y_k,x_k} + \alpha_i \delta_{y_i,x_i} + \alpha_{n+1}\delta_{y_i,x_{n+1}}$$

with $\alpha_i+\alpha_{n+1}=\beta_i$. Since

$$\sum_{k\neq i,k\in\{1,n\}}\beta_k\delta_{y_k,x_k}+\alpha_i\delta_{y_i,x_i}+\alpha_{n+1}\delta_{y_i,x_{n+1}}=\delta_{y,x}+(\alpha_i-\beta_i)\delta_{y_i,x_i}+\alpha_{n+1}\delta_{y_i,x_{n+1}}$$
$$=\delta_{y,x}+\alpha_{n+1}(\delta_{y_i,x_{n+1}}-\delta_{y_i,x_i}),$$

by the identification $\gamma:=\alpha_{n+1}$, we conclude that $\nu\in\Pi_i$ defined in Equation (3.5). The proof in the $m=n+2$ case is essentially the same.

B.5. Proof of Lemma 3.2. Let us define

$$\Theta:=\Big\{\nu\in P_1^{-1}\mu_n\cap\Delta_{n+2}(X\times X):\nu=\sum_{i=1}^{m}\alpha_i\delta_{x_i,x_i'},1\leq m\leq n+2,\alpha_i>0,x_i,x_i'\in X,i=1,\dots,m,$$
$$\text{the vectors }\big(\mathbb{1}_{y_1}(x_i),\dots,\mathbb{1}_{y_n}(x_i),\mathbb{1}_{d(x_i,x_i')>\epsilon},1\big),i=1,\dots,m\text{ are linearly independent}\Big\}. \tag{B.9}$$

Then the identity

$$\Gamma_{\mu_n,\epsilon}=P_1^{-1}\mu_n\cap\big\{\nu\in\mathcal{M}(X\times X):\nu\{d>\epsilon\}\leq\epsilon\big\}$$

implies that

$$\bar{\Theta}=\Theta\cap\big\{\nu\in\mathcal{M}(X\times X):\nu\{d>\epsilon\}\leq\epsilon\big\}. \tag{B.10}$$

As in Section 3.1, let us compute $\bar{\Theta}$ by first computing Θ and then using the identity (B.10). To that end, observe that the definition (B.9) of Θ implies that the support points $(x_i,x_i'),i=1,\dots,m$ contain no duplicates so that we can apply Lemma 3.1 which implies that we can constrain the values of m in the definition of Θ to $n\leq m\leq n+2$. Moreover, Θ is defined in terms of $P_1^{-1}\mu_n\cap\Delta_{n+2}(X\times X)$, and by Lemma 3.1 we have $P_1^{-1}\mu_n\cap\Delta_{n+2}(X\times X)=\Pi_0\cup_{k=1}^{n}\Pi_k\cup_{i\leq j}\Pi_{i,j}$. Consequently, using the multi-index $\imath$ introduced above in Equation (3.11), it is natural to define

$$\Theta_\imath:=\Theta\cap\Pi_\imath$$

and observe that

$$\Theta=\Theta_0\cup_{k=1}^{n}\Theta_k\cup_{i\leq j}\Theta_{i,j}.$$

First consider Θ_0. Since the definition of Π_0 implies that $\{x_j,j=1,\dots,n\}$ must be a permutation of $\{y_i,i=1,\dots,n\}$, it follows that the linear independence condition of Equation (B.9) is satisfied in this case. That is,

$$\Theta_0=\Pi_0. \tag{B.11}$$

Now consider Π_i for $i\in\{1,\dots,n\}$. Then the definition (3.5) of Π_i implies that, upon relabeling, the linear independence of the set $\big(\mathbb{1}_{y_1}(x_i),\dots,\mathbb{1}_{y_n}(x_i),\mathbb{1}_{d(x_i,x_i')>\epsilon},1\big),i=1,\dots,n+1$ amounts to the linear independence of the set

$$\Big(I_{n\times n},z_n,I_n\Big)$$

together with

$$\Big(0,\dots,1_i,\dots,0,\mathbb{1}_{d(y_i,x'_{n+1})>\epsilon},1\Big)$$

where z_n has components $\mathbb{1}_{d(y_i,x'_i)>\epsilon}, i=1,\dots,n$, $I_{n\times n}$ is the identity matrix, I_n is the vector of 1s, and 1_i indicates a 1 in the ith position. Because the first row has the identity matrix, this set of vectors is linearly independent if and only if

$$\begin{gathered}\Big(0,\dots,1_i,\dots,0,\ \mathbb{1}_{d(y_i,x'_i)>\epsilon}\ ,1\Big)\\ \Big(0,\dots,1_i,\dots,0,\mathbb{1}_{d(y_i,x'_{n+1})>\epsilon},1\Big)\end{gathered}$$

is linearly independent, which is equivalent to the assertion that $x'\in\Lambda_i$ defined in Equation (3.15). Consequently, we obtain

$$\Theta_i=\Pi_i\cap\Lambda_i. \tag{B.12}$$

For $\Theta_{i,j}$ with $i\le j$, let us first show that $\Theta_{i,i}=\emptyset$. To that end, let $x'\in X^{n+2}$ and consider $\nu\in\Pi_{i,i}(x')$. Then using the same reasoning as above, it follows that the linear independence condition is equivalent to the linear independence of the three vectors

$$\begin{gathered}\Big(0,\dots,1_i,\dots,0,\ \mathbb{1}_{d(y_i,x'_i)>\epsilon}\ ,1\Big)\\ \Big(0,\dots,1_i,\dots,0,\mathbb{1}_{d(y_i,x'_{n+1})>\epsilon},1\Big)\\ \Big(0,\dots,1_i,\dots,0,\mathbb{1}_{d(y_i,x'_{n+2})>\epsilon},1\Big).\end{gathered}$$

Since the last row is identically 1, the independence of this set is not possible regardless of the values of $\mathbb{1}_{d(y_i,x'_i)>\epsilon}, \mathbb{1}_{d(y_i,x'_{n+1})>\epsilon}$ and $\mathbb{1}_{d(y_i,x'_{n+2})>\epsilon}$. Therefore,

$$\Theta_{i,i}=\emptyset, \quad i=1,\dots,n. \tag{B.13}$$

So let us consider $\Theta_{i,j}$ with $i<j$. Then, upon relabeling, the linear independence of the set $\big(\mathbb{1}_{y_1}(x_i),\dots,\mathbb{1}_{y_n}(x_i),\mathbb{1}_{d(x_i,x'_i)>\epsilon},1\big), i=1,\dots,n+2$ amounts to the linear independence of the set

$$\Big(I_{n\times n}, z_n, I_n\Big)$$

together with

$$\begin{gathered}\Big(0,\dots,1_i,\dots,0,\dots,0,\mathbb{1}_{d(y_i,x'_{n+1})>\epsilon},1\Big)\\ \Big(0,\dots,0,\dots,1_j,\dots,0,\mathbb{1}_{d(y_j,x'_{n+2})>\epsilon},1\Big).\end{gathered}$$

Because the first row has the identity matrix, this set of vectors is linearly independent if and only if both

$$\begin{gathered}\Big(0,\dots,1_i,\dots,0,\ \mathbb{1}_{d(y_i,x'_i)>\epsilon}\ ,1\Big)\\ \Big(0,\dots,1_i,\dots,0,\mathbb{1}_{d(y_i,x'_{n+1})>\epsilon},1\Big)\end{gathered}$$

and

$$\begin{gathered}\Big(0,\dots,1_j,\dots,0,\ \mathbb{1}_{d(y_j,x'_j)>\epsilon}\ ,1\Big)\\ \Big(0,\dots,1_j,\dots,0,\mathbb{1}_{d(y_j,x'_{n+2})>\epsilon},1\Big)\end{gathered}$$

are linearly independent. Then, as in the Θ_i case above, the linear independence of these two sets is equivalent to requiring that $x' \in \Lambda_{i,j}$ defined in Equation (3.16). That is, we have

$$\Theta_{i,j} = \Pi_{i,j} \cap \Lambda_{i,j}. \tag{B.14}$$

Therefore, we have established that

$$\Theta = \Pi_0 \cup_{k=1}^{n} (\Pi_i \cap \Lambda_i) \cup_{i<j} (\Pi_{i,j} \cap \Lambda_{i,j}),$$

and the assertion then easily follows.

Acknowledgments. The authors thank the referees for a thorough and thoughtful review of the manuscript providing many substantial improvements in its presentation. The authors gratefully acknowledge this work supported by the Air Force Office of Scientific Research and the DARPA EQUiPS Program under Award Number FA9550-12-1-0389 (Scientific Computation of Optimal Statistical Estimators) and Number FA9550-16-1-0054 (Computational Information Games).

REFERENCES

[1] E.M. Alfsen, *Compact Convex Sets and Boundary Integrals*, Springer Berlin-Heidelberg-New York, 71, 1971.
[2] C.D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, Berlin, Third Edition, 2006.
[3] R.B. Ash, *Real Analysis and Probability*, Probability and Mathematical Statistics, Academic Press, New York, 11, 1972.
[4] N. Bourbaki and S.K. Berberian, *Integration II*, Springer, 2004.
[5] G. Choquet, *Lectures on Analysis: Vol. III: Infinite Dimensional Measures and Problem Solutions*, W.A. Benjamin, 1969.
[6] C. Dellacherie and P.-A. Meyer, *Probabilités et Potentiel*, Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.
[7] L.E. Dubins, *On extreme points of convex sets*, J. Math. Anal. Appl., 5(2), 237–244, 1962.
[8] R.M. Dudley, *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 74, 2002. Revised reprint of the 1989 original.
[9] J. Dupačová, *Uncertainties in minimax stochastic programs*, Optimization, 60(10-11), 1235–1250, 2011.
[10] P.M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*, arXiv:1505.05116, 2015.
[11] A.S. Kechris, *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1995.
[12] H. Owhadi, C. Scovel, T.J. Sullivan, M. McKerns, and M. Ortiz, *Optimal uncertainty quantification*, SIAM Review, 55(2), 271–345, 2013.
[13] W.W. Rogosinski, *Moments of non-negative mass*, Proceedings of the Royal Society of London. Series A. Math. Phys. Sci., 245(1240), 1–27, 1958.
[14] B. Simon, *Convexity: An Analytic Viewpoint*, Cambridge Univ. Press, Cambridge, 2011.
[15] S. Straszewicz, *Über exponierte punkte abgeschlossener Punktmengen*, Fundamenta Mathematicae, 24(1), 139–143, 1935.
[16] A.M. Vershik, *Kantorovich metric: Initial history and little-known applications*, J. Math. Sci., 133(4), 1410–1417, 2006.
[17] S. Willard, *General Topology*, Addison-Wesely Publishing Company, London, 1970.
[18] G. Winkler, *On the integral representation in convex noncompact sets of tight measures*, Mathematische Zeitschrift, 158(1), 71–77, 1978.
[19] G. Winkler, *Extreme points of moment sets*, Math. Oper. Res., 13(4), 581–587, 1988.
[20] D. Wozabal, *A framework for optimization under ambiguity*, Annals of Operations Research, 193(1), 21–47, 2012.