

# BOUNDS ON CHOQUET RISK MEASURES IN FINITE PRODUCT SPACES WITH AMBIGUOUS MARGINALS

MARIO GHOSSOUB  
UNIVERSITY OF WATERLOO  
[MARIO.GHOSSOUB@UWATERLOO.CA](mailto:MARIO.GHOSSOUB@UWATERLOO.CA)

DAVID SAUNDERS  
UNIVERSITY OF WATERLOO  
[DSAUNDERS@UWATERLOO.CA](mailto:DSAUNDERS@UWATERLOO.CA)

KELVIN SHUANGJIAN ZHANG\*  
UNIVERSITY OF WATERLOO  
[KS3ZHANG@UWATERLOO.CA](mailto:KS3ZHANG@UWATERLOO.CA)

ABSTRACT. We investigate the problem of finding upper and lower bounds for a Choquet risk measure of a nonlinear function of two risk factors, when the marginal distributions of the risk factors are ambiguous and represented by nonadditive measures on the marginal spaces and the joint nonadditive distribution on the product space is unknown. We treat this problem as a generalization of the optimal transport problem to the setting of nonadditive measures. We provide explicit characterizations of the optimal solutions for finite marginal spaces, and we investigate some of their properties. We further discuss the connections with linear programming, showing that the optimal transport problems for capacities are linear programs, and we also characterize their duals explicitly. Finally, we investigate a series of numerical examples, including a comparison with the classical optimal transport problem, and applications to counterparty credit risk.

*Key Words and Phrases:* Risk Management, Optimal Transport, Non-Additive Measures, Risk Measures, Cooperative Games.

---

\* Corresponding author.

*2020 Mathematics Subject Classification:* 49Q22, 90C08, 91A12, 91A70, 91G40, 91G70.

Mario Ghossoub and David Saunders acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada in the form of Discovery Grants (NSERC Grant Nos. 2018-03961 and 2017-04220, respectively).

## 1. INTRODUCTION

An important problem in the literature on credit risk management is that of determining bounds on the Credit Valuation Adjustment (CVA), that is, the price adjustment on a given derivatives portfolio to account for potential counterparty credit risk losses (e.g., [Garcia-Cespedes, de Juan Herrero, Rosen, and Saunders \(2010\)](#); [Glasserman and Yang \(2018\)](#); [Rosen and Saunders \(2010, 2012\)](#)). A portfolio's counterparty credit risk exposure depends on market risk factors, and the likelihood of a counterparty default depends on credit risk factors. Consequently, the computation of CVA requires the modelling of potential portfolio losses as functions of these two sets of dependent risk factors. There is a large literature on the required credit risk models (e.g., [McNeil, Frey, and Embrechts \(2015\)](#) and the references therein). In practice, counterparty exposures often depend on a large number of risk factors (equity prices, interest rates, exchange rates, etc.), leading to several challenges with their measurement and management (e.g., [Brigo, Morini, and Pallavicini \(2013\)](#); [Gregory \(2020\)](#)).

Joint models of market and credit risk are, in general, very difficult to develop and estimate in practice. Hence, even when the marginal distributions of the market and credit risk factors are known, there is still uncertainty about their joint distribution and about the ensuing CVA computation. [Glasserman and Yang \(2018\)](#) examine bounds on CVA arising from the uncertainty about the dependence structure. They formulated the problem of finding the worst-case CVA with respect to the dependence structure between the risk factors as an Optimal Transport (OT) problem. In related work, [Memartoluie, Saunders, and Wirjanto \(2017\)](#) considered in a formal way the problem of finding the worst-case Expected Shortfall (ES) of a nonlinear function of market risk and credit risk, given the marginal distributions of the factors, and they showed that in the case of finite sample spaces, the problem is equivalent to a linear program. Recently, [Ghossoub, Hall, and Saunders \(2023\)](#) extended the problem to general spaces and to spectral risk measures. They examined the problem of finding a worst-case spectral risk measure of a nonlinear function of two risk factors with known marginals, with respect to their dependence structure. They formulated the problem as a generalized OT problem and provided a strong duality theory similar to the Kantorovich duality in classical OT theory.

OT is the subject of a large literature, dating back to the seminal work of [Monge \(1781\)](#) and [Kantorovich \(1942\)](#). [Monge \(1781\)](#) considered the problem of minimizing the total cost (measured using the Euclidean distance between the source and the target) of moving one mass distribution to another among all volume-preserving maps. [Kantorovich \(1942, 1948\)](#) later relaxed this problem by expanding the feasible set to all measure couplings with given marginal distributions and developed a duality theory for the relaxed problem. Modern OT is a large and rapidly developing field (e.g., [Santambrogio \(2015\)](#); [Villani \(2008\)](#)) with applications to several areas within mathematics (e.g., [S. Rachev and Rüschendorf \(1998\)](#); [Villani \(2003\)](#)), and applied fields such as physics (e.g., [Guillen and Kitagawa \(2017\)](#); [R. McCann \(2020\)](#)), statistics (e.g., [Panaretos and Zemel \(2022\)](#); [Zhang, Peyré, Fadili, and Pereyra \(2020\)](#)), economics (e.g., [Carlier and Zhang \(2020\)](#); [Galichon \(2016\)](#); [R. J. McCann and Zhang \(2019\)](#)), finance (e.g., [Eckstein, Guo, Lim, and Obłój \(2021\)](#); [Henry-Labordère \(2017\)](#)), and machine learning (e.g., [Peyré and Cuturi \(2019\)](#); [Torres, Pereira, and Amini \(2021\)](#)), for instance.

In the aforementioned literature, the marginal distributions of risk factors are assumed to be given and known, but their dependence structure is unknown. In particular, the marginals are (additive) probability measures. As a result, problems of bounding risk measures of loss functions can be formulated as (generalized) OT problems, with various cost functions, depending on the particular application. In many such applications, particularly related to the modelling of decision-making under ambiguity or vagueness in beliefs, a decision-maker's attitude toward, and sensitivity to ambiguity in beliefs is represented by monotone set functions that lack additivity. Such objects are called *capacities* or *nonadditive measures*. See, for example, the work of [Quiggin \(1982, 1993\)](#); [Schmeidler \(1986, 1989\)](#); [Yaari \(1987\)](#) for theoretical foundations. In particular, the seminal contribution of [Schmeidler \(1986, 1989\)](#) axiomatized models of decision-making under ambiguity in which the decision-maker's preferences admit a representation in terms of an expected utility with respect to a nonadditive measure. Such expectations are defined through the notion of a Choquet integral with respect to a capacity.<sup>1</sup> We refer to [Denneberg \(1994\)](#) or [Marinacci and Montrucchio \(2004\)](#) for more about capacities and Choquet integration.

In this paper, we are interested in the problem of bounding a risk measure of a nonlinear function of two risk factors, but where (i) the marginal distributions of the risk factors are ambiguous, and represented by nonadditive measures on the marginal spaces; and, (ii) the objective function is a Choquet integral. As in [Glasserman and Yang \(2018\)](#), we consider the case of two risk factors defined on finite spaces. We assume given (marginal) capacities on these spaces, representing the ambiguous distributions of the risk factors, and consider the problem of finding the joint capacity on the product space with these given marginals, which maximizes or minimizes the Choquet integral of a given portfolio loss function. We treat this problem as a generalization of the OT problem to the setting of nonadditive measures. We provide explicit characterizations of the optimal solutions for finite marginal spaces, and we investigate some of their properties. Additionally, we explore connections to linear programming and present a version of the Kantorovich duality.

The remainder of the paper is organized as follows. Section 2 presents definitions and background material needed for the rest of the paper. Section 3 formulates the problem of bounding Choquet risk measures as an OT problem with nonadditive marginals. Section 4 presents a mathematical formulation of the OT problem for capacities, investigates properties of its feasible set, and gives characterizations and explicit formulas for its solution. In addition, we further study properties of the optimal capacities (in particular, non-emptiness of the core) in terms of the corresponding properties of the marginal capacities. The explicit formula for the core of the minimizer can be found in that section. Moreover, as in the case of measures, the OT problem for capacities can be formulated as a linear program (see [Torra \(2023\)](#) for a related result), and we characterize its dual in Section 5. Section 6 presents numerical examples comparing our problem to the classical OT problem and illustrating its use in a counterparty credit risk application. Finally, Section 7 concludes.

---

<sup>1</sup>Note that our use of the word capacity here is distinct from the usage in the literature on optimal transport with capacity constraints (e.g., [Korman and McCann \(2015\)](#); [Korman, McCann, and Seis \(2015\)](#); [Pennanen and Perkkiö \(2019\)](#)), where the “capacity constraint” imposes an upper bound on the density of the coupling.

## 2. PRELIMINARIES

**2.1. Capacities and Choquet Integration.** Denote by  $B(\Sigma)$  the vector space of all bounded and  $\Sigma$ -measurable real-valued functions on a given measurable space  $(S, \Sigma)$ . Then  $(B(\Sigma), \|\cdot\|_{sup})$  is a Banach space (Dunford & Schwartz, 1958, IV.5.1), where  $\|\cdot\|_{sup}$  denotes the supnorm.

Let  $ba(\Sigma)$  denote the linear space of all bounded finitely additive set functions on  $(S, \Sigma)$ . When equipped with the variation norm  $\|\cdot\|_v$ ,  $ba(\Sigma)$  is a Banach space, and  $(ba(\Sigma), \|\cdot\|_v)$  is isometrically isomorphic to the norm-dual of the Banach space  $(B(\Sigma), \|\cdot\|_{sup})$  (e.g., (Dunford & Schwartz, 1958, IV.5.1)) via the duality  $\langle \phi, \lambda \rangle = \int \phi d\lambda$ ,  $\forall \lambda \in ba(\Sigma)$ ,  $\forall \phi \in B(\Sigma)$ . Denote by  $ca(\Sigma)$  the collection of all countably additive elements of  $ba(\Sigma)$ . Then  $ca(\Sigma)$  is a  $\|\cdot\|_v$ -closed (and hence complete) linear subspace of  $ba(\Sigma)$ . Henceforth, a collection of probability measures will be called weak\*-compact if it is compact in the weak\* topology  $\sigma(ba(\Sigma), B(\Sigma))$ .

**Definition 2.1.** A capacity (nonadditive measure) on a measurable space  $(S, \Sigma)$  is a finite set function  $\gamma : \Sigma \rightarrow [0, \nu(S)]$  such that  $\gamma(\emptyset) = 0$  and  $\gamma$  is monotone; that is, for any  $A, B \in \Sigma$ ,  $\gamma(A) \leq \gamma(B)$  whenever  $A \subseteq B$ . When  $\gamma(S) = 1$ , the capacity  $\gamma$  is said to be normalized.

The conjugate of a capacity  $\gamma$  on  $(S, \Sigma)$  is the finite set function  $\bar{\gamma} : \Sigma \rightarrow [0, \nu(S)]$  defined by  $\bar{\gamma}(A) := \gamma(S) - \gamma(A^c)$ , for all  $A \in \Sigma$ . Then  $\bar{\gamma}$  is also a capacity, and if  $\gamma$  is normalized then so is  $\bar{\gamma}$ .

A capacity  $\gamma$  is called supermodular (resp. submodular) if

$$\gamma(A \cup B) + \gamma(A \cap B) \geq (\text{resp. } \leq) \gamma(A) + \gamma(B), \quad \forall A, B \in \Sigma.$$

The core of a capacity  $\gamma$  on  $(S, \Sigma)$ , denoted by  $\mathcal{C}(\gamma)$ , is the collection of all bounded finitely additive measures  $\eta$  on  $(S, \Sigma)$  such that  $\eta(A) \geq \gamma(A)$ , for all  $A \in \Sigma$ . When nonempty,  $\text{core}(\gamma)$  is weak\*-compact and convex.

**Definition 2.2.** Let  $\gamma$  be a capacity on  $(S, \Sigma)$ . The Choquet integral of  $Y \in B(\Sigma)$  with respect to  $\gamma$  is defined by

$$\int Y d\gamma := \int_0^{+\infty} \gamma(\{s \in S : Y(s) \geq t\}) dt + \int_{-\infty}^0 [\gamma(\{s \in S : Y(s) \geq t\}) - 1] dt,$$

where the integrals are taken in the sense of Riemann.

**Definition 2.3.** Two functions  $Y_1, Y_2 \in B(\Sigma)$  are said to be comonotonic if

$$\left[ Y_1(s) - Y_1(s') \right] \left[ Y_2(s) - Y_2(s') \right] \geq 0, \quad \text{for all } s, s' \in S.$$

If  $\gamma \in ca(\Sigma)$ , then the Choquet integral with respect to  $\gamma$  is the usual Lebesgue integral with respect to  $\gamma$  (e.g., (Marinacci & Montrucchio, 2004, p. 59)). Unlike the Lebesgue integral, the Choquet integral is not an additive operator on  $B(\Sigma)$ . However, the Choquet integral is additive over comonotonic functions.

**Proposition 2.4.** Let  $\gamma$  be a capacity on  $(S, \Sigma)$ .

- (1) If  $\phi_1, \phi_2 \in B(\Sigma)$  are comonotonic, then  $\int (\phi_1 + \phi_2) d\gamma = \int \phi_1 d\gamma + \int \phi_2 d\gamma$ .
- (2) If  $\phi_1, \phi_2 \in B(\Sigma)$  are such that  $\phi_1 \leq \phi_2$ , then  $\int \phi_1 d\gamma \leq \int \phi_2 d\gamma$ .
- (3) For all  $\phi \in B(\Sigma)$  and all  $c \geq 0$ , then  $\int c\phi d\gamma = c \int \phi d\gamma$ .
- (4) If  $\gamma$  is submodular, then for any  $\phi_1, \phi_2 \in B(\Sigma)$ ,  $\int (\phi_1 + \phi_2) d\gamma \leq \int \phi_1 d\gamma + \int \phi_2 d\gamma$ .
- (5) If  $\gamma$  is supermodular, then for any  $\phi_1, \phi_2 \in B(\Sigma)$ ,  $\int (\phi_1 + \phi_2) d\gamma \geq \int \phi_1 d\gamma + \int \phi_2 d\gamma$ .

**2.2. Risk Measures.** Risk measures are real-valued functionals defined on some collection of random variables on a given probability space. They are often used either as a quantification of riskiness of a given financial position, or as a way to determine adequate capital requirements (e.g., Föllmer and Schied (2016), McNeil et al. (2015), or Rüschendorf (2013)). Formally, a risk measure is a mapping  $\rho : \mathcal{X} \rightarrow \mathbb{R}$ , where  $\mathcal{X}$  is a prespecified collection of random variables on a given probability space  $(S, \Sigma, \mathbb{P})$ . Common properties of risk measures include:

- R.1** (Monotonicity)  $\rho(X) \leq \rho(Y)$ , for all  $X, Y \in \mathcal{X}$  such that  $X \leq Y$ ,  $\mathbb{P}$ -a.s.
- R.2** (Positive Homogeneity)  $\rho(\lambda X) = \lambda\rho(X)$ , for all  $X \in \mathcal{X}$  and all  $\lambda \in \mathbb{R}_+$ .
- R.3** (Cash Invariance)  $\rho(X + c) = \rho(X) + c$ , for all  $X \in \mathcal{X}$  and  $c \in \mathbb{R}$ .
- R.4** (Subadditivity)  $\rho(X + Y) \leq \rho(X) + \rho(Y)$  for all  $X, Y \in \mathcal{X}$ .
- R.5** (Comonotonic Additivity)  $\rho(X + Y) = \rho(X) + \rho(Y)$  for all  $X, Y \in \mathcal{X}$  that are comonotonic.
- R.6** (Law Invariance)  $\rho(X) = \rho(Y)$  when  $X$  and  $Y$  have the same distribution under  $\mathbb{P}$ .

A coherent risk measure Artzner, Delbaen, Eber, and Heath (1999); Delbaen (2002) is a risk measure that satisfies Axioms **R.1-R.4**, which are considered desirable for effective risk management. A practically relevant example of a coherent risk measure, frequently used in the banking and insurance industries, is the Expected Shortfall (ES), also known as the Conditional Value-at-Risk (CVaR). If  $F_X^{\leftarrow}(t)$  is the left-continuous quantile of  $X$ , and  $\alpha \in (0, 1)$ , then the expected shortfall of  $X$  at the confidence level  $\alpha$  is:

$$\text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 F_X^{\leftarrow}(t) dt.$$

If the space  $(S, \Sigma, \mathbb{P})$  is nonatomic, then a coherent, comonotonic additive, and law-invariant risk measure admits a representation as a *spectral risk measure* (e.g., (Föllmer & Schied, 2016, Theorem 4.93), Kusuoka (2001), (McNeil et al., 2015, Proposition 8.18), Shapiro (2013)), that is, as a risk measure of the form

$$\rho(X) = \int_0^1 \text{ES}_u(X) d\Gamma(u),$$

for some probability measure  $\Gamma$  on  $[0, 1]$ . Moreover, by a classical result on Choquet integration (e.g., Schmeidler (1986)), monotone and comonotonic additive risk measures admit

a representation in terms of a Choquet integral of the form

$$\rho(X) = \int X d\gamma,$$

for some capacity  $\gamma$  on  $(S, \Sigma)$ .

The class of spectral risk measures is flexible and rich enough to encompass several of the most popular and practically relevant risk measures. Additionally, there is a tight relationship between spectral risk measures and the subclass of Choquet risk measures called Distortion Risk Measures (DRM). These are Choquet risk measures for which the capacity  $\gamma$  is of the form  $T \circ \mathbb{P}$ , for some increasing function  $T : [0, 1] \rightarrow [0, 1]$  such that  $T(0) = 1 - T(1) = 0$ . The function  $T$  is called a distortion function, or a probability weighting function. Indeed, it can be shown (e.g., Acerbi (2002); Föllmer and Schied (2016); Kusuoka (2001); McNeil et al. (2015); Shapiro (2013)) that a spectral risk measure also admits the representation

$$\rho(X) = \int_0^1 F_X^{\leftarrow}(t) \kappa(t) dt,$$

where  $\kappa : [0, 1] \rightarrow \mathbb{R}_+$  is a nonnegative and increasing function that satisfies  $\int_0^1 \kappa(t) dt = 1$ . This function is called the spectral function. For instance, for the Expected Shortfall (ES) at level  $\alpha$ , the spectral function is given by  $(1 - \alpha)^{-1} \mathbf{1}_{[\alpha, 1]}(t)$ , that is,  $\text{ES}_\alpha(X) = (1 - \alpha)^{-1} \int_\alpha^1 F_X^{\leftarrow}(t) dt$ . Moreover, letting

$$T(x) = 1 - \int_0^{1-x} \kappa(t) dt, \quad \forall x \in [0, 1],$$

it follows that  $T$  is a distortion function, and it can be shown that  $\rho$  is DRM with respect to  $T \circ \mathbb{P}$ , that is,

$$\rho(X) = \int X dT \circ \mathbb{P}.$$

**2.3. Finite State Spaces.** Suppose that  $\mathcal{Z}$  is a nonempty finite set, and let  $\Sigma = 2^{\mathcal{Z}}$  be the collection of all of its subsets. Throughout, we identify measures on any nonempty finite set  $\mathcal{Z}$  with vectors  $v \in \mathbb{R}^{|\mathcal{Z}|}$  through  $v(A) = \sum_{i \in A} v_i$ . Let  $\gamma$  be a capacity on  $(\mathcal{Z}, \Sigma)$ .

**Definition 2.5.** The Möbius transform of a capacity  $\gamma$  is defined as

$$m^\gamma(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \gamma(B).$$

The Choquet integral of a function  $f$  with respect to the capacity  $\gamma$  can be represented in terms of the Möbius transform as follows:

$$\begin{aligned} \gamma(f) &= \sum_{A \subseteq \mathcal{X}} m^\gamma(A) \bigwedge_{x \in A} f_x = \sum_{A \subseteq \mathcal{X}} \sum_{B \subseteq A} (-1)^{|A \setminus B|} \gamma(B) \bigwedge_{x \in A} f_x \\ (2.1) \quad &= \sum_{B \subseteq \mathcal{X}} \gamma(B) \left( \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x \right) = \sum_{B \subseteq \mathcal{X}} K_f(B) \gamma(B), \end{aligned}$$

with

$$(2.2) \quad K_f(B) := \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x,$$

where  $f_x = f(x)$ , and  $\bigwedge_{x \in A} f_x$  represents the minimum of  $f$  on  $A$  (e.g., (Grabisch, 2016, Theorem 4.95)). See Grabisch (2016) and Marinacci and Montrucchio (2004) for more information about the Möbius transform.

**Definition 2.6.** Let  $\mathcal{Z}$  be a nonempty finite set, and let  $\mathcal{G} \subseteq 2^{\mathcal{Z}}$  be a collection of subsets containing  $\mathcal{Z}$  and the empty set. Suppose that a function  $G : \mathcal{G} \rightarrow \mathbb{R}_+$  satisfies  $G(\emptyset) = 0$ , and  $G(A) \leq G(B)$  whenever  $A, B \in \mathcal{G}$ ,  $A \subseteq B$ . The capacity on  $\mathcal{Z}$  defined by

$$G^*(B) := \inf_{\substack{A \in \mathcal{G} \\ A \supseteq B}} G(A), \text{ for all } B \in 2^{\mathcal{Z}},$$

is called the outer envelope of  $G$ . The capacity defined by

$$G_*(B) := \sup_{\substack{A \in \mathcal{G} \\ A \subseteq B}} G(A), \text{ for all } B \in 2^{\mathcal{Z}},$$

is called the inner envelope of  $G$ .

When it is necessary to make  $\mathcal{G}$  explicit in the notation, we will write  $G^*(B) = G^*(B; \mathcal{G})$  for the outer envelope, and  $G_*(B) = G_*(B; \mathcal{G})$  for the inner envelope. It is easy to see that  $G_* \leq G^*$ .<sup>2</sup>

**Definition 2.7.** Given nonempty finite sets  $\mathcal{X}, \mathcal{Y}$ , we define  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}$  to be the collection of all subsets of  $\mathcal{X} \times \mathcal{Y}$  of the form  $A \times B$  with  $A \subseteq \mathcal{X}$  and  $B \subseteq \mathcal{Y}$ . We define  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$  to be the collection of all subsets of  $\mathcal{X} \times \mathcal{Y}$  of the form  $A \times B$  with  $A \subseteq \mathcal{X}$  and  $B \subseteq \mathcal{Y}$ , and either  $A = \mathcal{X}$  or  $B = \mathcal{Y}$  (or both). That is  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$  is the collection of all sets either of the form  $\mathcal{X} \times B$  with  $B \subseteq \mathcal{Y}$  or  $A \times \mathcal{Y}$  with  $A \subseteq \mathcal{X}$ .

Sets in product spaces and their projections will feature prominently in the optimal solutions of our optimization problems. The notation in the next definition will be convenient.

**Definition 2.8.** For a set  $M \subseteq \mathcal{X} \times \mathcal{Y}$ , define:

$$\begin{aligned} M_{\mathcal{X}} &:= \{x \in \mathcal{X} : \exists z = (x, y) \in M\}, & M_{\mathcal{Y}} &:= \{y \in \mathcal{Y} : \exists z = (x, y) \in M\}, \\ \widetilde{M}_{\mathcal{X}} &:= \{x \in \mathcal{X} : (x, y) \in M, \forall y \in \mathcal{Y}\}, & \widetilde{M}_{\mathcal{Y}} &:= \{y \in \mathcal{Y} : (x, y) \in M, \forall x \in \mathcal{X}\}. \end{aligned}$$

It is easy to see that  $\widetilde{M}_{\mathcal{X}} = ((M^c)_{\mathcal{X}})^c$ , and  $\widetilde{M}_{\mathcal{Y}} = ((M^c)_{\mathcal{Y}})^c$ .

**Definition 2.9.** Let  $k \geq 2$  be an integer. A capacity  $\gamma$  on  $\mathcal{Z}$  is called  $k$ -monotone if for any sets  $A_1, \dots, A_k \in \mathcal{Z}$ ,

$$\gamma \left( \bigcup_{j=1}^k A_j \right) \geq \sum_{\substack{J \subseteq \{1, \dots, k\} \\ J \neq \emptyset}} (-1)^{|J|+1} \gamma \left( \bigcap_{j \in J} A_j \right).$$

<sup>2</sup>Fix  $M \subseteq \mathcal{Z}$ , and  $A, B \in \mathcal{G}$  with  $A \subseteq M \subseteq B$ . Then  $G(A) \leq G(B)$ . Minimizing over  $B$  containing  $M$  yields  $G(A) \leq G^*(M)$ , and then maximizing over  $A$  contained in  $M$  gives that  $G_*(M) \leq G^*(M)$ .

The capacity is called  $k$ -alternating if the above inequality is reversed. A 2-monotone capacity is supermodular, while a 2-alternating capacity is submodular. If  $\gamma$  is  $k$ -monotone for all  $k \geq 2$ , it is called totally monotone, and if it is  $k$ -alternating for all  $k \geq 2$ , it is called totally alternating.

### 3. BOUNDS ON CHOQUET RISK MEASURES

**3.1. Problem Formulation.** We consider the case of a portfolio whose loss depends on two risk factors defined on two finite spaces. We assume given (marginal) capacities on these spaces, representing the ambiguous distributions of the risk factors, and we consider the problem of finding the joint capacity on the product space with these given marginals that maximizes or minimizes the Choquet integral of a given portfolio loss function.

Specifically, let  $\mathcal{X}$  and  $\mathcal{Y}$  be non-empty finite sets, and let  $X$  and  $Y$  be random variables on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. We are given a function  $L : X(\mathcal{X}) \times Y(\mathcal{Y}) \rightarrow \mathbb{R}$  representing the loss on a portfolio consisting of the risk factors  $X$  and  $Y$ . The distributional uncertainty, or ambiguity, about the risk factors is represented by capacities  $\mu$  on  $\mathcal{X}$  and  $\nu$  on  $\mathcal{Y}$ , to be interpreted as ambiguous beliefs about the distributions of  $X$  and  $Y$ , respectively.

A joint distribution for  $X$  and  $Y$  is represented by a capacity on the product space  $\mathcal{X} \times \mathcal{Y}$ , such that the projections onto  $\mathcal{X}$  and  $\mathcal{Y}$  are  $\mu$  and  $\nu$ , respectively.

**Definition 3.1.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be nonempty finite sets,  $\mu$  a capacity on  $\mathcal{X}$ ,  $\nu$  a capacity on  $\mathcal{Y}$ , and  $\pi$  a capacity on  $\mathcal{X} \times \mathcal{Y}$ .

- (1) The marginal capacities of  $\pi$  on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, are defined by

$$\pi_{\mathcal{X}}(A) := \pi(A \times \mathcal{Y}) \quad \text{and} \quad \pi_{\mathcal{Y}}(B) := \pi(\mathcal{X} \times B), \quad \text{for all } A \subseteq \mathcal{X} \text{ and } B \subseteq \mathcal{Y}.$$

- (2) The set of all capacities  $\pi$  on  $\mathcal{X} \times \mathcal{Y}$  such that  $\pi_{\mathcal{X}} = \mu$  and  $\pi_{\mathcal{Y}} = \nu$  is denoted by  $\Pi_{\text{Ch}}(\mu, \nu)$ .

We are interested in evaluating a risk measure  $\rho(L(X, Y))$  of the portfolio loss function in the case where  $\rho$  is a Choquet integral of  $L(X, Y)$  with respect to a capacity  $\pi$  on  $\mathcal{X} \times \mathcal{Y}$ :

$$\rho_{\pi}(L(X, Y)) = \int_{\mathcal{X} \times \mathcal{Y}} L(X, Y) d\pi.$$

In our framework, while the capacities  $\mu$  and  $\nu$  are given, no information about the dependence structure (and hence the joint distribution) of the two risk factors is available. Therefore, computing a Choquet risk measure of the portfolio loss function is not possible without further information. A natural question that arises is whether we are able to establish upper and lower bounds on the value of such a risk measures with respect to the uncertainty about the joint capacity  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ . Specifically, our problem is that of finding capacities that maximize or minimize the Choquet integral of  $L(X, Y)$  among all capacities in  $\Pi_{\text{Ch}}(\mu, \nu)$ :

$$(3.1) \quad \mathcal{L}(L; \Pi_{\text{Ch}}(\mu, \nu)) := \inf_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \rho_{\pi}(L(X, Y)) \leq \sup_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \rho_{\pi}(L(X, Y)) =: \mathcal{U}(L; \Pi_{\text{Ch}}(\mu, \nu)).$$



Problem (3.1) can be seen as a generalization of the optimal transport problem to the setting of nonadditive measures.

#### 4. THE OPTIMAL TRANSPORT PROBLEM FOR CAPACITIES

In this section, we formulate the optimal transport problem for capacities.<sup>3</sup> Once the problem is formulated, we investigate properties of the feasible set. Understanding the lattice structure of the feasible set leads immediately to explicit formulas for the optimizers.

**Definition 4.1.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be non-empty finite sets, and let  $u$  and  $v$  be probability measures on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Denote by  $\Pi_a(u, v)$  the set of measures on  $\mathcal{X} \times \mathcal{Y}$  that have the marginals  $u$  on  $\mathcal{X}$  and  $v$  on  $\mathcal{Y}$ . That is,

$$\Pi_a(u, v) := \left\{ \pi \mid \pi \text{ is a measure on } \mathcal{X} \times \mathcal{Y} \text{ such that } \pi(A \times \mathcal{Y}) = u(A), \text{ for any } A \subseteq \mathcal{X}, \right. \\ \left. \text{and } \pi(\mathcal{X} \times B) = v(B), \text{ for any } B \subseteq \mathcal{Y}. \right\}$$

Given a function  $f$ , the optimal transport minimization problem is:

$$(4.1) \quad \inf_{\pi \in \Pi_a(u, v)} \pi(f) = \inf_{\pi \in \Pi_a(u, v)} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} f(x, y) \pi(\{(x, y)\}).$$

Similarly, given a function  $g$ , the optimal transport maximization problem is:

$$(4.2) \quad \sup_{\pi \in \Pi_a(u, v)} \pi(g).$$

Both the maximization and minimization problems are linear in  $\pi$ . Because  $\Pi_a(u, v)$  is convex and compact, optimal solutions exist, and the set of optimal solutions contains at least one extreme point of the feasible set. For instance, when  $|\mathcal{X}| = |\mathcal{Y}|$  and both  $u$  and  $v$  are uniform measures, by Birkhoff's Theorem there exists an optimal solution supported on  $\bigcup_{i=1}^{|\mathcal{X}|} \{(x_i, y_{\sigma(i)})\}$ , for some permutation  $\sigma$ .

**Definition 4.2.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be nonempty finite sets and  $\pi$  be a capacity on  $\mathcal{X} \times \mathcal{Y}$ . The marginal capacities of  $\pi$  on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, are defined by

$$\pi_{\mathcal{X}}(A) := \pi(A \times \mathcal{Y}) \quad \text{and} \quad \pi_{\mathcal{Y}}(B) := \pi(\mathcal{X} \times B),$$

for all  $A \subseteq \mathcal{X}$ ,  $B \subseteq \mathcal{Y}$ .

In particular, for two probability measures  $u$  and  $v$ ,  $\Pi_a(u, v) \subseteq \Pi_{\text{Ch}}(u, v)$ , where the latter is defined in Definition 3.1. The proof of the following result is straightforward.

**Lemma 4.3.** *Let  $\mu$  and  $\nu$  be normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Then  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  if and only if  $\bar{\pi} \in \Pi_{\text{Ch}}(\bar{\mu}, \bar{\nu})$ .*

Given a function  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ , consider the analogue of the optimal transport problem on capacity couplings, i.e. finding capacities to maximize or minimize the Choquet integral

<sup>3</sup> The transshipment problem is an optimization problem closely related to the optimal transport problem, see S. T. Rachev and Rüschemdorf (1998), Chen, Hu, and Jiang (2022), and the references therein. A formulation of the transshipment problem for capacities is contained in an appendix that is available from the authors upon request.

of  $f$  among all capacities in  $\Pi_{\text{Ch}}(\mu, \nu)$ :

$$\mathcal{L}(f; \Pi_{\text{Ch}}(\mu, \nu)) := \inf_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) \leq \sup_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) =: \mathcal{U}(f; \Pi_{\text{Ch}}(\mu, \nu)).$$

We note that, since  $\pi(-f) \neq -\pi(f)$  in general, it is worthwhile to develop the theories for the minimum and maximum problems in parallel.

**4.1. The Feasible Set and Its Properties.** The first thing to observe about the feasible set is that it is nonempty.

**Proposition 4.4.** *Let  $\mu$  and  $\nu$  be normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$  respectively. Then  $\Pi_{\text{Ch}}(\mu, \nu) \neq \emptyset$ .*

*Proof.* Define the function  $G : \mathcal{P}_{\mathcal{X}, \mathcal{Y}} \rightarrow \mathbb{R}_+$  by  $G(A \times B) := \mu(A) \cdot \nu(B)$  for  $A \times B \in \mathcal{P}_{\mathcal{X}, \mathcal{Y}}$  with  $A \subseteq \mathcal{X}$  and  $B \subseteq \mathcal{Y}$ . It is easy to verify that both  $G_*$  and  $G^*$  are in  $\Pi_{\text{Ch}}(\mu, \nu)$ .  $\square$

We note that we could have used  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$  in place of  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}$  in the above argument, and reached the same conclusion. Since  $\Pi_{\text{Ch}}(\mu, \nu)$  is defined by a finite system of linear equalities and inequalities, and  $0 \leq \pi(B) \leq 1$  for any set  $B$ , we in fact have the following result.

**Proposition 4.5.** *Let  $\mu$  and  $\nu$  be normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$ . Then  $\Pi_{\text{Ch}}(\mu, \nu)$  is a compact, convex polyhedron in  $\mathbb{R}^{2^{|\mathcal{X}|+|\mathcal{Y}|}}$ .*

**Remark 4.6.**

- A capacity  $\gamma$  is called the unanimity game associated with the set  $F$  if  $\gamma(G) = 1$  if  $G \supseteq F$ , and  $\gamma(G) = 0$  otherwise. If  $\mu$  is the unanimity game associated with  $A \subseteq \mathcal{X}$ , and  $\nu$  is the unanimity game associated with  $B \subseteq \mathcal{Y}$ , then the unanimity game  $\pi$  associated with  $A \times B \subseteq \mathcal{X} \times \mathcal{Y}$  is in  $\Pi_{\text{Ch}}(\mu, \nu)$ .
- Suppose that  $\mu$  is a totally monotone capacity on  $\mathcal{X}$  with Möbius transform  $m^\mu$ , and  $\nu$  is a totally monotone capacity on  $\mathcal{Y}$  with Möbius transform  $m^\nu$ , then  $\pi$  defined to be the capacity on  $\mathcal{X} \times \mathcal{Y}$  with Möbius transform given by

$$m^\pi(F) = \begin{cases} m^\mu(A) \cdot m^\nu(B), & F = A \times B, A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}; \\ 0, & \text{otherwise,} \end{cases}$$

is a totally monotone capacity in  $\Pi_{\text{Ch}}(\mu, \nu)$ .<sup>4</sup> For further information on this construction, see Bauer (2012); Destercke (2013); Ghirardato (1997); Hendon, Jacobsen, Sloth, and Tran (1991); Koshevoy (1998); Walley and Fine (1982). Combining the above argument with Lemma 4.3, it is easy to see that if  $\mu$  and  $\nu$  are totally alternating, then there exists a totally alternating capacity  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ .

- A *possibility measure*  $\gamma$  is defined as a normalized capacity such that  $\gamma(A \cup B) = \max(\gamma(A), \gamma(B))$ , for any sets  $A$  and  $B$ . From this definition, it is easy to see that  $\gamma(A) = \max_{z \in A} \gamma(\{z\})$  (and by normalization, there must exist  $z$  such that  $\gamma(\{z\}) =$

<sup>4</sup>It should be noted that if  $\mu$  and  $\nu$  are capacities, this construction does not in general result in a capacity. A counterexample is given by  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ ,  $\mu = \nu$ , with  $\mu(\emptyset) = 0$ ,  $\mu(\{0\}) = \mu(\{1\}) = 0.7$ ,  $\mu(\mathcal{X}) = 1$  (see Dyckerhoff (2022)).

1). If  $\mu$  and  $\nu$  are possibility measures, then  $\pi(A) := \max_{(x,y) \in A} \mu(\{x\}) \cdot \nu(\{y\})$  defines a possibility measure in  $\Pi_{\text{Ch}}(\mu, \nu)$ . The conjugate of a possibility measure is called a necessity measure (which satisfies  $\gamma(A \cap B) = \min(\gamma(A), \gamma(B))$ ). Again, using Lemma 4.3 one can show that if  $\mu$  and  $\nu$  are necessity measures, then there exists a necessity measure  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ .

A capacity is said to be balanced if its core is nonempty. The next result demonstrates that there exists a balanced  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  if and only if both  $\mu$  and  $\nu$  are balanced.

**Proposition 4.7.** *Let  $\mu$  and  $\nu$  be normalized capacities on nonempty finite sets  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Then the following are equivalent:*

- (1) *Both  $\mu$  and  $\nu$  have nonempty cores (i.e.,  $\mathcal{C}(\mu) \neq \emptyset$  and  $\mathcal{C}(\nu) \neq \emptyset$ ).*
- (2) *There exists  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  with a nonempty core.*

*Proof.* Suppose that  $u \in \mathcal{C}(\mu)$  and  $v \in \mathcal{C}(\nu)$ . Define a measure  $w$  on  $\mathcal{X} \times \mathcal{Y}$  by  $w(\{(x, y)\}) := u(\{x\})v(\{y\})$  and additivity. Further, define  $G : \mathcal{P}_{\mathcal{X}, \mathcal{Y}} \rightarrow \mathbb{R}_+$  by  $G(A \times B) := \mu(A) \cdot \nu(B)$  for  $A \subseteq \mathcal{X}$  and  $B \subseteq \mathcal{Y}$ , and take  $\pi = G_* \in \Pi_{\text{Ch}}(\mu, \nu)$ . It is easy to see that  $\pi(\mathcal{X} \times \mathcal{Y}) = w(\mathcal{X} \times \mathcal{Y})$ . Let  $M \subseteq \mathcal{X} \times \mathcal{Y}$ , and consider  $K = A \times B \in \mathcal{P}_{\mathcal{X}, \mathcal{Y}}$ ,  $K \subseteq M$ . Then:

$$\begin{aligned} G(K) = \mu(A) \nu(B) &\leq \sum_{x \in A} \sum_{y \in B} u(\{x\})v(\{y\}) = \sum_{z=(x,y) \in K} w(\{(x, y)\}) \\ &\leq \sum_{z=(x,y) \in M} w(\{(x, y)\}) = w(M). \end{aligned}$$

This implies that  $\pi(M) = G_*(M) \leq w(M)$ , for all  $M \subseteq \mathcal{X} \times \mathcal{Y}$ . Therefore,  $w \in \mathcal{C}(\pi)$ .

Conversely, let  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  and  $w \in \mathcal{C}(\pi)$ , and define for  $y \in \mathcal{Y}$ ,  $v(\{y\}) := \sum_{x \in \mathcal{X}} w(\{(x, y)\})$ . With  $B \subseteq \mathcal{Y}$ , we have

$$v(B) = \sum_{y \in B} v(\{y\}) = \sum_{x \in \mathcal{X}, y \in B} w(\{(x, y)\}) = w(\mathcal{X} \times B) \geq \pi(\mathcal{X} \times B) = \nu(B),$$

with equality when  $B = \mathcal{Y}$ , and therefore  $v \in \mathcal{C}(\nu) \neq \emptyset$ . The same argument yields  $\mathcal{C}(\mu) \neq \emptyset$ .  $\square$

**Remark 4.8.** It should be noted that there can exist capacities  $\mu$  on  $\mathcal{X}$  and  $\nu$  on  $\mathcal{Y}$  with nonempty cores and an element  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  with an empty core. Consider  $\mathcal{X} = \{x_1, x_2\}$ ,  $\mathcal{Y} = \{y_1, y_2\}$ , and take  $\mu$  and  $\nu$  to be probability measures on  $\mathcal{X}$  and  $\mathcal{Y}$  respectively, giving equal weight to each point. Define  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  to give value zero to the empty set, 1 to  $\mathcal{X} \times \mathcal{Y}$ ,  $\frac{1}{4}$  to any subset consisting of a single point,  $\frac{1}{2}$  to any subset consisting of two points, and  $\frac{7}{8}$  to any subset consisting of three points. Any element  $w \in \mathcal{C}(\pi)$  would have to satisfy  $w(\{(x_1, y_1)\}) \geq \frac{1}{4}$ , and  $w(\mathcal{X} \times \mathcal{Y} \setminus \{(x_1, y_1)\}) \geq \frac{7}{8}$ , and thus  $w(\mathcal{X} \times \mathcal{Y}) \geq \frac{9}{8} > 1$ , contradicting  $w(\mathcal{X} \times \mathcal{Y}) = \pi(\mathcal{X} \times \mathcal{Y}) = 1$ .

**4.2. Lattice Structure of the Feasible Set and Characterization of the Optimal Solutions.** If we think of normalized capacities on  $\mathcal{Z}$  as functions on the collection of subsets  $2^{\mathcal{Z}}$ , then given two capacities  $\gamma$  and  $\pi$ , we can define, for  $A \subseteq \mathcal{Z}$ :

$$(\pi \wedge \gamma)(A) := \min(\pi(A), \gamma(A)), \quad (\pi \vee \gamma)(A) := \max(\pi(A), \gamma(A)).$$

With these definitions,  $\pi \wedge \gamma$  and  $\pi \vee \gamma$  are both capacities, and the collection of all normalized capacities is a bounded distributive lattice, with largest element giving value 1 to all nonempty sets, and smallest element giving value 0 to all sets except  $\mathcal{Z}$ , which has value 1.<sup>5</sup>

Since all capacities in  $\Pi_{\text{Ch}}(\mu, \nu)$  have the same values for sets of the form  $A \times \mathcal{Y}$ , for  $A \subseteq \mathcal{X}$ , and  $\mathcal{X} \times B$ , for  $B \subseteq \mathcal{Y}$ , we have that  $\Pi_{\text{Ch}}(\mu, \nu)$  is a distributive sublattice. Furthermore,  $\Pi_{\text{Ch}}(\mu, \nu)$  is bounded (as a lattice) with maximum and minimum elements given by taking setwise maxima and minima:

$$\pi^*(A) = \sup_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(A) \quad \text{and} \quad \pi_*(A) = \inf_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(A).$$

The next result follows from the definition of the Choquet integral.

**Theorem 4.9.** *For  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ , and  $\pi_*$  and  $\pi^*$  described above, we have*

$$\min_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) = \pi_*(f) \quad \text{and} \quad \max_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) = \pi^*(f).$$

*Proof.* We first verify that both  $\pi_*$  and  $\pi^*$  are indeed feasible. Note that if  $N = A \times \mathcal{Y}$  for  $A \subseteq \mathcal{X}$ , then  $\pi(N) = \mu(A)$  for all  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ , and therefore  $\pi^*(N) = \pi_*(N) = \mu(A)$ . Similarly, if  $N = \mathcal{X} \times B$  with  $B \subseteq \mathcal{Y}$ , then  $\pi^*(N) = \pi_*(N) = \nu(B)$ . Furthermore, by their definitions, both  $\pi^*$  and  $\pi_*$  are non-negative non-decreasing set functions, i.e. capacities. In other words, we have that  $\pi^*, \pi_* \in \Pi_{\text{Ch}}(\mu, \nu)$ .

Now, by the definition in (4.2),  $\pi_*$  and  $\pi^*$  achieve the set-wise infimum and supremum among  $\Pi_{\text{Ch}}(\mu, \nu)$ , respectively. Let  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ . Then:

$$\begin{aligned} \pi(f) &= \int_0^\infty \pi(\{f \geq t\}) dt + \int_{-\infty}^0 (\pi(\{f \geq t\}) - \pi(\mathcal{Z})) dt \\ &= \int_0^\infty \pi(\{f \geq t\}) dt + \int_{-\infty}^0 (\pi(\{f \geq t\}) - 1) dt \\ &\geq \int_0^\infty \pi_*(\{f \geq t\}) dt + \int_{-\infty}^0 (\pi_*(\{f \geq t\}) - 1) dt = \pi_*(f). \end{aligned}$$

The proof for  $\pi^*$  is similar. □

It is possible to find explicit expressions for  $\pi_*$  and  $\pi^*$ .

**Theorem 4.10.** *For any  $N \subseteq \mathcal{X} \times \mathcal{Y}$ ,*

$$\pi_*(N) = \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})) \quad \text{and} \quad \pi^*(N) = \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})).$$

*Proof.* Define  $G : \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^* \rightarrow \mathbb{R}$  by

$$G(M) := \begin{cases} \mu(A), & \text{if } M = A \times \mathcal{Y}; \\ \nu(B), & \text{if } M = \mathcal{X} \times B. \end{cases}$$

Let  $G^*$  and  $G_*$  be the outer and inner envelope of  $G$  as defined in Definition 2.6 with  $\mathcal{G} = \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$ . From the monotonicity of  $\mu$  on  $2^{\mathcal{X}}$  (with the inclusion order), it is not hard to

<sup>5</sup>We note that there is another way of defining lattice operations on capacities, involving setwise maxima and minima of their Möbius transforms. See Grabisch (2016); Marinacci and Montrucchio (2004) for details.

see that, for any  $N \in \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$  with  $N = A \times \mathcal{Y}$ , one has  $G^*(N) = G_*(N) = \mu(A)$ . Similarly, for any  $N = \mathcal{X} \times B$  with  $B \subseteq \mathcal{Y}$ , we have  $G^*(N) = G_*(N) = \nu(B)$ . By definition,  $G^*$  and  $G_*$  are clearly non-negative and non-decreasing, so  $G^*, G_* \in \Pi_{\text{Ch}}(\mu, \nu)$ .

For any  $N \subseteq \mathcal{X} \times \mathcal{Y}$ ,  $\tilde{N}_{\mathcal{X}} \times \mathcal{Y} \subseteq N \subseteq N_{\mathcal{X}} \times \mathcal{Y}$  and  $\mathcal{X} \times \tilde{N}_{\mathcal{Y}} \subseteq N \subseteq \mathcal{X} \times N_{\mathcal{Y}}$ . Therefore,  $G^*(N) \leq \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}}))$  and  $G_*(N) \geq \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}}))$ . If  $N \subseteq A \times \mathcal{Y}$ , then  $N_{\mathcal{X}} \subseteq A$ , and if  $A' \times \mathcal{Y} \subseteq N$ , then  $A' \subseteq \tilde{N}_{\mathcal{X}}$ . The monotonicity of  $\mu$  and  $\nu$  then imply that

$$\begin{aligned} G^*(N) &= \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})), \\ G_*(N) &= \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})). \end{aligned}$$

To complete the proof, we will show that  $\pi_* = G_*$  and  $\pi^* = G^*$ . For any  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  and  $N \subseteq \mathcal{X} \times \mathcal{Y}$ , the relation  $\tilde{N}_{\mathcal{X}} \times \mathcal{Y} \subseteq N \subseteq N_{\mathcal{X}} \times \mathcal{Y}$  implies that

$$\mu(\tilde{N}_{\mathcal{X}}) = \pi(\tilde{N}_{\mathcal{X}} \times \mathcal{Y}) \leq \pi(N) \leq \pi(N_{\mathcal{X}} \times \mathcal{Y}) = \mu(N_{\mathcal{X}}),$$

and  $\mathcal{X} \times \tilde{N}_{\mathcal{Y}} \subseteq N \subseteq \mathcal{X} \times N_{\mathcal{Y}}$  implies that

$$\nu(\tilde{N}_{\mathcal{Y}}) = \pi(\mathcal{X} \times \tilde{N}_{\mathcal{Y}}) \leq \pi(N) \leq \pi(\mathcal{X} \times N_{\mathcal{Y}}) = \nu(N_{\mathcal{Y}}).$$

Therefore,

$$G_*(N) = \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})) \leq \pi(N) \leq \min(\mu(N_{\mathcal{X}}), \nu(N_{\mathcal{Y}})) = G^*(N).$$

This implies,  $G_* \leq \pi_*$  and  $\pi^* \leq G^*$ . The equalities hold because  $G_*, G^* \in \Pi_{\text{Ch}}(\mu, \nu)$ .  $\square$

**Remark 4.11.** If we explicitly include the dependence of the optimizers on the marginal capacities, i.e. when given  $\mu, \nu$  write  $\pi_*(\cdot; \mu, \nu)$  and  $\pi^*(\cdot; \mu, \nu)$  for the smallest and largest elements of  $\Pi_{\text{Ch}}(\mu, \nu)$ , then it is easy to show that  $\bar{\pi}_*(\cdot; \mu, \nu) = \pi^*(\cdot; \bar{\mu}, \bar{\nu})$  and  $\bar{\pi}^*(\cdot; \mu, \nu) = \pi_*(\cdot; \bar{\mu}, \bar{\nu})$ .

**Remark 4.12.**

- Suppose that  $\mu$  is the unanimity game associated with  $A \subseteq \mathcal{X}$  and  $\nu$  is the unanimity game associated with  $B \subseteq \mathcal{Y}$ , and  $N \subseteq \mathcal{X} \times \mathcal{Y}$ . Then  $\pi_*(N) = 1$  if either  $A \times \mathcal{Y} \subseteq N$  or  $\mathcal{X} \times B \subseteq N$ , and zero otherwise. On the other hand,  $\pi^*(N) = 1$  if for all  $x_0 \in A$  there exists  $y(x_0) \in \mathcal{Y}$  such that  $(x_0, y(x_0)) \in N$  and for all  $y_0 \in B$  there exists  $x(y_0) \in \mathcal{X}$  such that  $(x(y_0), y_0) \in N$ , and  $\pi^*(N) = 0$  otherwise.
- Suppose that  $\mu$  and  $\nu$  are possibility measures, and define  $M : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$  by  $M(x, y) := \max(\mu(\{x\}), \nu(\{y\}))$ . Then given  $N \subseteq \mathcal{X} \times \mathcal{Y}$ ,

$$\pi_*(N) = \max(\max_{x \in \tilde{N}_{\mathcal{X}}} \mu(\{x\}), \max_{y \in \tilde{N}_{\mathcal{Y}}} \nu(\{y\})) = \max_{(x, y) \in \tilde{N}_{\mathcal{X}} \times \tilde{N}_{\mathcal{Y}}} M(x, y).$$

Define  $m : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$  by  $m(x, y) := \min(\mu(\{x\}), \nu(\{y\}))$ , then

$$\pi^*(N) = \min(\max_{x \in N_{\mathcal{X}}} \mu(\{x\}), \max_{y \in N_{\mathcal{Y}}} \nu(\{y\})) = \max_{(x, y) \in N_{\mathcal{X}} \times N_{\mathcal{Y}}} m(x, y).$$

When  $\mu$  and  $\nu$  are necessity measures, then  $\pi_*$  and  $\pi^*$  can be calculated using the previous remark.

Consider  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ . For a fixed  $x \in \mathcal{X}$ , define

$$f_y(x) := \min\{f(x, y) : y \in \mathcal{Y}\} \quad \text{and} \quad f^y(x) := \max\{f(x, y) : y \in \mathcal{Y}\},$$

with  $f_x, f^x : \mathcal{Y} \rightarrow \mathbb{R}$  defined similarly. Then

$$\begin{aligned} \widetilde{\{f \geq t\}}_{\mathcal{X}} &= \{x \in \mathcal{X} : (x, y) \in \{f \geq t\} \forall y \in \mathcal{Y}\} \\ &= \{x \in \mathcal{X} : \min_{y \in \mathcal{Y}} f(x, y) \geq t\} = \{f_y \geq t\}. \end{aligned}$$

Similarly  $\widetilde{\{f \geq t\}}_{\mathcal{Y}} = \{f_x \geq t\}$ , and therefore

$$\pi_*(\{f \geq t\}) = \max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})),$$

and

$$\pi_*(f) = \int_0^\infty \max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})) dt + \int_{-\infty}^0 (\max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})) - 1) dt,$$

using the fact that we have assumed  $\mu$  and  $\nu$  to be normalized.

Using a similar argument,

$$\begin{aligned} \{f \geq t\}_{\mathcal{X}} &= \{x \in \mathcal{X} : \exists y \in \mathcal{Y}, f(x, y) \geq t\} \\ &= \{x \in \mathcal{X} : \max_{y \in \mathcal{Y}} f(x, y) \geq t\} = \{f^y \geq t\}, \end{aligned}$$

and  $\{f \geq t\}_{\mathcal{Y}} = \{f^x \geq t\}$ . Thus,

$$\pi^*(\{f \geq t\}) = \min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})),$$

and

$$\pi^*(f) = \int_0^\infty \min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})) dt + \int_{-\infty}^0 (\min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})) - 1) dt.$$

To conclude, we have

$$\begin{aligned} &\mathcal{L}(f; \Pi_{\text{Ch}}(\mu, \nu)) \\ &= \min_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) = \pi_*(f) \\ &= \int_0^\infty \max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})) dt + \int_{-\infty}^0 (\max(\mu(\{f_y \geq t\}), \nu(\{f_x \geq t\})) - 1) dt \\ &\leq \int_0^\infty \min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})) dt + \int_{-\infty}^0 (\min(\mu(\{f^y \geq t\}), \nu(\{f^x \geq t\})) - 1) dt \\ &= \pi^*(f) = \max_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(f) \\ &= \mathcal{U}(f; \Pi(\mu, \nu)). \end{aligned}$$

**4.3. Balancedness and Cores of the Optimal Solutions.** Since  $\pi_*(N) \leq \pi(N) \leq \pi^*(N)$ , for all  $N \subseteq \mathcal{X} \times \mathcal{Y}$  and  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ , we immediately obtain the following result.

**Proposition 4.13.** *Let  $\mu$  and  $\nu$  be normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. The following statements regarding the cores hold.*

- (1) *If  $\mathcal{C}(\pi^*) \neq \emptyset$ , then  $\mathcal{C}(\pi) \neq \emptyset$  for all  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ .*
- (2) *If  $\mathcal{C}(\pi_*) = \emptyset$ , then  $\mathcal{C}(\pi) = \emptyset$  for all  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$ .*
- (3) *In particular,  $\mathcal{C}(\pi_*) \neq \emptyset$  iff  $\mathcal{C}(\mu) \neq \emptyset$  and  $\mathcal{C}(\nu) \neq \emptyset$ .*

*Proof.* Suppose  $p \in \mathcal{C}(\pi^*)$ , then for any fixed  $\pi \in \Pi(\mu, \nu)$  and any  $N \subseteq \mathcal{X} \times \mathcal{Y}$ , one has  $p(N) \geq \pi^*(N) \geq \pi(N)$ , with both equalities hold at  $N = \mathcal{X} \times \mathcal{Y}$ . Therefore,  $p \in \mathcal{C}(\pi)$ . Using the same argument, one can show (2). Proposition 4.7 together with (2) implies (3).  $\square$

However,  $\mathcal{C}(\pi^*)$  is typically empty, as per the following result.

**Proposition 4.14.** *Suppose that  $\mu$  and  $\nu$  are normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, and  $|\mathcal{X}| \geq 2$ ,  $|\mathcal{Y}| \geq 2$ . Then  $\mathcal{C}(\pi^*) = \emptyset$ .*

*Proof.* Let  $\{A_1, A_2\}$  and  $\{B_1, B_2\}$  be partitions of  $\mathcal{X}$  and  $\mathcal{Y}$  respectively, and define:

$$N^1 = (A_1 \times B_1) \cup (A_2 \times B_2) \quad \text{and} \quad N^2 = (A_1 \times B_2) \cup (A_2 \times B_1).$$

Then  $N_{\mathcal{X}}^1 = N_{\mathcal{X}}^2 = \mathcal{X}$ ,  $N_{\mathcal{Y}}^1 = N_{\mathcal{Y}}^2 = \mathcal{Y}$ , so that for the disjoint sets  $N^1$  and  $N^2$ ,  $\pi^*(N^1) = \pi^*(N^2) = 1$ .  $\square$

We can in fact explicitly identify  $\mathcal{C}(\pi_*)$  in terms of  $\mathcal{C}(\mu)$  and  $\mathcal{C}(\nu)$ .

**Proposition 4.15.** *Let  $\mu$  and  $\nu$  be normalized capacities on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Then*

$$\mathcal{C}(\pi_*) = \bigcup_{u \in \mathcal{C}(\mu), v \in \mathcal{C}(\nu)} \Pi_a(u, v).$$

*Proof.* Let  $w \in \mathcal{C}(\pi_*)$ , and for each fixed  $x_0 \in \mathcal{X}$ ,  $y_0 \in \mathcal{Y}$  define  $u_w(\{x_0\}) := \sum_{y \in \mathcal{Y}} w(\{x_0, y\})$ , and  $v_w(\{y_0\}) := \sum_{x \in \mathcal{X}} w(\{x, y_0\})$ . Clearly  $w \in \Pi_a(u_w, v_w)$ . Furthermore, for  $A \subseteq \mathcal{X}$ , we have

$$u_w(A) = w(A \times \mathcal{Y}) \geq \pi_*(A \times \mathcal{Y}) = \mu(A),$$

since  $\pi_* \in \Pi_{\text{Ch}}(\mu, \nu)$ . Thus,  $u_w \in \mathcal{C}(\mu)$ , and similarly  $v_w \in \mathcal{C}(\nu)$ .

Conversely, suppose that  $w \in \Pi_a(u, v)$  with  $u \in \mathcal{C}(\mu)$  and  $v \in \mathcal{C}(\nu)$ . Clearly,  $w(\mathcal{X} \times \mathcal{Y}) = u(\mathcal{X}) = \mu(\mathcal{X}) = 1$ . Let  $N \subseteq \mathcal{X} \times \mathcal{Y}$ , and note that  $\tilde{N}_{\mathcal{X}} \times \mathcal{Y} \subseteq N$  and  $\mathcal{X} \times \tilde{N}_{\mathcal{Y}} \subseteq N$ . Then  $\pi_*(N) = \max(\mu(\tilde{N}_{\mathcal{X}}), \nu(\tilde{N}_{\mathcal{Y}})) \leq \max(u(\tilde{N}_{\mathcal{X}}), v(\tilde{N}_{\mathcal{Y}})) = \max(w(\tilde{N}_{\mathcal{X}} \times \mathcal{Y}), w(\mathcal{X} \times \tilde{N}_{\mathcal{Y}})) \leq w(N)$ . That is,  $w \in \mathcal{C}(\pi_*)$ .  $\square$

**Remark 4.16.** By (Grabisch, 2016, Corollary 2.23 (ii)),  $\gamma$  is supermodular if and only if for every  $A \subseteq B \subseteq \mathcal{X} \times \mathcal{Y}$  and  $z \notin B$ ,  $\Delta_z \gamma(A) \leq \Delta_z \gamma(B)$ , where  $\Delta_z \gamma(A) := \gamma(A \cup \{z\}) - \gamma(A)$ , and  $\Delta_z \gamma(B)$  is defined similarly. It is well-known that if  $\gamma$  is supermodular, then  $\mathcal{C}(\gamma) \neq \emptyset$  (e.g., (Grabisch, 2016, Theorem 3.15)).

Let  $\mathcal{X} = \{x_1, x_2, x_3\}$ , and  $\mathcal{Y} = \{y_1, y_2, y_3\}$ , and let  $\mu$  be the additive (and therefore supermodular) capacity with  $\mu(\{x_1\}) = \mu(\{x_2\}) = 0.1$ , and  $\mu(\{x_3\}) = 0.8$ , with  $\nu$  defined on

$\mathcal{Y}$  in the same way. Define:

$$A := \{(x_1, y_2), (x_1, y_3)\} \text{ and } B := \{(x_1, y_2), (x_1, y_3), (x_2, y_3), (x_3, y_3)\},$$

and  $z := (x_1, y_1)$ . Note that  $\tilde{A}_{\mathcal{X}} = \emptyset$ ,  $\tilde{A}_{\mathcal{Y}} = \emptyset$ , so  $\pi_*(A) = 0$ . Also,  $(\widetilde{A \cup z})_{\mathcal{X}} = \{x_1\}$ ,  $(\widetilde{A \cup z})_{\mathcal{Y}} = \emptyset$ , so  $\Delta_z \pi_*(A) = \pi_*(A \cup z) = \mu(\{x_1\}) = 0.1$ . Furthermore,  $\tilde{B}_{\mathcal{X}} = \emptyset$ ,  $\tilde{B}_{\mathcal{Y}} = \{y_3\}$ ,  $(\widetilde{B \cup z})_{\mathcal{X}} = \{x_1\}$ , and  $(\widetilde{B \cup z})_{\mathcal{Y}} = \{y_3\}$ , so  $\pi_*(B) = \pi_*(B \cup z) = \nu(\{y_3\}) = 0.8$ , and  $\Delta_z \pi_*(B) = 0$ . Thus, we conclude that while  $\pi_*$  has a nonempty core, it is not supermodular.

**Definition 4.17.** A capacity  $\gamma$  on  $\mathcal{Z}$  is said to be exact if for every  $S \in 2^{\mathcal{Z}} \setminus \emptyset$ , there exists a core element  $p \in \mathcal{C}(\gamma)$  such that  $p(S) = \gamma(S)$ .

We have seen that  $\mathcal{C}(\pi^*)$  is typically empty, so that  $\pi^*$  will not be exact. In the case when  $\mu$  and  $\nu$  are exact, we may ask whether  $\pi_*$  is exact. That is, we define the capacity  $\tilde{\pi} \in \Pi_{\text{Ch}}(\mu, \nu)$  by:

$$\tilde{\pi}(N) := \min \left\{ p(N) : p \in \bigcup_{u \in \mathcal{C}(\mu), v \in \mathcal{C}(\nu)} \Pi_a(u, v) \right\}, \text{ for any } N \subseteq \mathcal{X} \times \mathcal{Y},$$

and we ask whether  $\pi_* = \tilde{\pi}$ .

**Remark 4.18.** In general  $\tilde{\pi}$  as defined above need not be either submodular or supermodular. To see this, consider the case  $\mathcal{X} = \mathcal{Y} = \{1, 2, \dots, n\}$  for some  $n \geq 3$ , with  $\mu$  and  $\nu$  being uniform probability measures, and let  $\pi'$  be the conjugate of  $\tilde{\pi}$ .<sup>6</sup> Then

$$\pi'(A) = 1 - \tilde{\pi}(A^c) = 1 - \min_{p \in \Pi_a(\mu, \nu)} p(A^c) = \max_{p \in \Pi_a(\mu, \nu)} p(A).$$

By Birkhoff's Theorem, the optimum  $\tilde{\pi}(A)$  (and similarly  $\pi'(A)$ ) is achieved by measures that put mass  $\frac{1}{n}$  on points  $\{x_i, y_{\sigma(i)}\}$  for some permutation  $\sigma$ . Consider  $A_1 = \{(1, 1)\}$ ,  $z = (n, n)$  and  $B_1 = \mathcal{X} \times \mathcal{Y} \setminus \{z\}$ . Then it is easy to see that  $\Delta_z \pi'(A_1) = \frac{2}{n} - \frac{1}{n} = \frac{1}{n}$ , while  $\Delta_z \pi'(B_1) = 1 - 1 = 0$ . Thus  $\Delta_z \pi'(A_1) > \Delta_z \pi'(B_1)$ , and  $A_1 \subseteq B_1$ , so  $\pi'$  is not supermodular (and therefore  $\tilde{\pi}$  is not submodular, see (Grabisch, 2016, Theorem 2.20)). On the other hand, consider  $A_2 = \{(1, 1)\}$ ,  $B_2 = \{(1, 1), (2, 1)\}$  and  $z = (1, 2)$ . Then  $\Delta_z \pi'(A_2) = 0$ , and  $\Delta_z \pi'(B_2) = \frac{1}{n}$ . We therefore have that  $(B_2 \cup \{z\})^c \subseteq (A_2 \cup \{z\})^c$ , and  $\Delta_z \tilde{\pi}((B_2 \cup \{z\})^c) = \Delta_z \pi'(B_2) > \Delta_z \pi'(A_2) = \Delta_z \tilde{\pi}((A_2 \cup \{z\})^c)$  (e.g., (Grabisch, 2016, Theorem 2.16)). Thus  $\tilde{\pi}$  is not supermodular (and  $\pi'$  is not submodular).

**Remark 4.19.** Let  $n \geq 2$ ,  $\mathcal{X} = \{1, \dots, n\}$  and  $\mathcal{Y} = \mathcal{X}$ , and take  $\mu$  and  $\nu$  to be two probability measures on  $\mathcal{X}$  that are not equal. Then  $\mathcal{C}(\mu) = \{\mu\}$ , and  $\mathcal{C}(\nu) = \{\nu\}$ , so that  $\mathcal{C}(\pi_*) = \Pi_a(\mu, \nu)$ . Notice that any element of  $\Pi_a(\mu, \nu)$  is also in  $\Pi_{\text{Ch}}(\mu, \nu)$ .  $\Pi_a(\mu, \nu)$  is compact, and for any fixed  $B$ ,  $p(B) = \sum_{\{x, y\} \in B} p(\{x, y\})$  is a continuous function on  $\Pi_a(\mu, \nu)$  and therefore its minimum is attained. Consider the set  $D = \{(1, 1), (2, 2), \dots, (n, n)\}$  and  $M = D^c$ . We have that  $\tilde{M}_{\mathcal{X}} = \tilde{M}_{\mathcal{Y}} = \emptyset$ , and therefore  $\pi_*(M) = 0$ . Suppose that  $\pi_*$  was exact. Then there is a  $\pi \in \Pi_a(\mu, \nu)$  such that  $\pi(M) = 0$ . But then  $\pi$  is concentrated on the diagonal  $D$ , contradicting the fact that  $\mu \neq \nu$ . This implies that  $\pi_*$  is not exact.

<sup>6</sup>We prefer to avoid the cumbersome notation  $\tilde{\pi}$ .



## 5. LINEAR PROGRAMMING AND THE KANTOROVICH DUALITY FOR CAPACITIES

In this section, we formulate the optimal transport problem for capacities as a linear program, and we present its dual. Recall that the Choquet integral of  $f$  with respect to a capacity  $\gamma$  on  $\mathcal{Z}$  can be written as

$$\int f d\gamma = \sum_{B \subseteq \mathcal{Z}} K_f(B) \gamma(B),$$

where

$$K_f(B) = \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{x \in A} f_x.$$

While this expression is not linear in  $f$ , it is linear in  $\gamma$ , and since the constraints defining  $\Pi_{\text{Ch}}(\mu, \nu)$  are all linear (see Proposition 4.5), the problem of minimizing  $\pi(c)$  over all  $\pi \in \Pi_{\text{Ch}}(\mu, \nu)$  becomes a linear program:

$$(5.1) \quad \min_{\pi} \sum_{B \subseteq \mathcal{X} \times \mathcal{Y}} K_c(B) \pi(B), \quad \text{subject to}$$

$$\pi(G \times \mathcal{Y}) = \mu(G), \quad \emptyset \neq G \subseteq \mathcal{X};$$

$$(5.2) \quad \pi(\mathcal{X} \times F) = \nu(F), \quad \emptyset \neq F \subseteq \mathcal{Y};$$

$$\pi(A \cup w) \geq \pi(A), \quad A \subset \mathcal{X} \times \mathcal{Y}, w = \{(x, y)\} \notin A;$$

$$\pi(\emptyset) = 0,$$

(e.g., (Grabisch, 2016, pp. 81-82)). Recall that a subset  $B$  of  $\mathcal{X} \times \mathcal{Y}$  is in  $\mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*$  if  $B = G \times \mathcal{Y}$  for some  $G \subseteq \mathcal{X}$  or  $B = \mathcal{X} \times F$  for some  $F \subseteq \mathcal{Y}$ .

The dual of the above linear program is given by

$$(5.3) \quad \max_{\hat{\varphi}, \hat{\psi}, \hat{\rho}} \sum_{G \subseteq \mathcal{X}} \hat{\varphi}(G) \mu(G) + \sum_{F \subseteq \mathcal{Y}} \hat{\psi}(F) \nu(F), \quad \text{subject to}$$

$$(5.4) \quad \hat{\varphi}(G) - \sum_{w \notin G \times \mathcal{Y}} \hat{\rho}(G \times \mathcal{Y}, w) + \sum_{w \in G \times \mathcal{Y}} \hat{\rho}((G \times \mathcal{Y}) \setminus \{w\}, w) = K_c(G \times \mathcal{Y}), \quad \emptyset \neq G \subsetneq \mathcal{X};$$

$$\hat{\psi}(F) - \sum_{w \notin \mathcal{X} \times F} \hat{\rho}(\mathcal{X} \times F, w) + \sum_{w \in \mathcal{X} \times F} \hat{\rho}((\mathcal{X} \times F) \setminus \{w\}, w) = K_c(\mathcal{X} \times F), \quad \emptyset \neq F \subsetneq \mathcal{Y};$$

$$\hat{\varphi}(\mathcal{X}) + \hat{\psi}(\mathcal{Y}) + \sum_w \hat{\rho}((\mathcal{X} \times \mathcal{Y}) \setminus \{w\}, w) = K_c(\mathcal{X} \times \mathcal{Y});$$

$$- \sum_{w \notin B} \hat{\rho}(B, w) + \sum_{w \in B} \hat{\rho}(B \setminus \{w\}, w) = K_c(B), \quad B \notin \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*;$$

$$\hat{\rho} \geq 0.$$

Let  $(\hat{\varphi}_*, \hat{\psi}_*, \hat{\rho}_*)$  be an optimal solution to (5.3 - 5.4). Then complementary slackness implies that, for any  $(A, w) \in \{(A, w) \in 2^{\mathcal{X} \times \mathcal{Y}} \times (\mathcal{X} \times \mathcal{Y}) : w \notin A\}$ ,

$$(5.5) \quad \hat{\rho}_*(A, w) (\pi_*(A \cup w) - \pi_*(A)) = 0.$$

**Remark 5.1.** The dual of the maximization problem

$$(5.6) \quad \max_{\pi \in \Pi_{\text{Ch}}(\mu, \nu)} \pi(c)$$

is given by

$$(5.7) \quad \min_{\hat{\varphi}, \hat{\psi}, \hat{\rho}} \sum_{G \subseteq \mathcal{X}} \hat{\varphi}(G) \mu(G) + \sum_{F \subseteq \mathcal{Y}} \hat{\psi}(F) \nu(F), \quad \text{subject to}$$

$$(5.8) \quad \begin{aligned} \hat{\varphi}(G) - \sum_{w \notin G \times \mathcal{Y}} \hat{\rho}(G \times \mathcal{Y}, w) + \sum_{w \in G \times \mathcal{Y}} \hat{\rho}((G \times \mathcal{Y}) \setminus \{w\}, w) &= K_c(G \times \mathcal{Y}), & \emptyset \neq G \subsetneq \mathcal{X}; \\ \hat{\psi}(F) - \sum_{w \notin \mathcal{X} \times F} \hat{\rho}(\mathcal{X} \times F, w) + \sum_{w \in \mathcal{X} \times F} \hat{\rho}((\mathcal{X} \times F) \setminus \{w\}, w) &= K_c(\mathcal{X} \times F), & \emptyset \neq F \subsetneq \mathcal{Y}; \\ \hat{\varphi}(\mathcal{X}) + \hat{\psi}(\mathcal{Y}) + \sum_w \hat{\rho}((\mathcal{X} \times \mathcal{Y}) \setminus \{w\}, w) &= K_c(\mathcal{X} \times \mathcal{Y}); \\ - \sum_{w \notin B} \hat{\rho}(B, w) + \sum_{w \in B} \hat{\rho}(B \setminus \{w\}, w) &= K_c(B), & B \notin \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^*; \\ \hat{\rho} &\leq 0. \end{aligned}$$

Suppose that  $(\hat{\varphi}^*, \hat{\psi}^*, \hat{\rho}^*)$  is an optimal solution to (5.7 - 5.8). Then by complementary slackness, for any  $(A, w) \in \{(A, w) \in 2^{\mathcal{X} \times \mathcal{Y}} \times (\mathcal{X} \times \mathcal{Y}) : w \notin A\}$ ,

$$\hat{\rho}^*(A, w) (\pi^*(A \cup w) - \pi^*(A)) = 0.$$

**Remark 5.2.** The dual of the minimization Optimal Transport problem is equivalent to the problem

$$(5.9) \quad \max_{L_\varphi, L_\psi, \hat{\rho}} \sum_{G \subseteq \mathcal{X}} m^\mu(G) L_\varphi(G) + \sum_{F \subseteq \mathcal{Y}} m^\nu(F) L_\psi(F), \quad \text{subject to}$$

$$(5.10) \quad \begin{aligned} L_\varphi(A\mathcal{X}) + L_\psi(A\mathcal{Y}) + \sum_{D \supseteq A} \sum_{w \in A} \hat{\rho}(D \setminus \{w\}, w) &= \bigwedge_{(x,y) \in A} c(x, y), & \emptyset \neq A \subseteq \mathcal{X} \times \mathcal{Y}; \\ \hat{\rho} &\geq 0. \end{aligned}$$

To see this, we will show, by the following change of variables<sup>7</sup>

$$\hat{\varphi}(G) := \sum_{B \supseteq G} (-1)^{|B \setminus G|} L_\varphi(B); \quad \hat{\psi}(F) := \sum_{A \supseteq F} (-1)^{|A \setminus F|} L_\psi(A),$$

that the objectives are equal and that the constraints can be derived from each other.

First, the objective function becomes

$$\begin{aligned} & \sum_{G \subseteq \mathcal{X}} \hat{\varphi}(G) \mu(G) + \sum_{F \subseteq \mathcal{Y}} \hat{\psi}(F) \nu(F) \\ &= \sum_{G \subseteq \mathcal{X}} \sum_{B \supseteq G} (-1)^{|B \setminus G|} L_\varphi(B) \mu(G) + \sum_{F \subseteq \mathcal{Y}} \sum_{A \supseteq F} (-1)^{|A \setminus F|} L_\psi(A) \nu(F) \\ &= \sum_{B \subseteq \mathcal{X}} \left( \sum_{G \subseteq B} (-1)^{|B \setminus G|} \mu(G) \right) L_\varphi(B) + \sum_{A \subseteq \mathcal{Y}} \left( \sum_{F \subseteq A} (-1)^{|A \setminus F|} \nu(F) \right) L_\psi(A) \end{aligned}$$

<sup>7</sup>This corresponds to the situation derived from a set function  $\xi_\varphi$ , where  $\hat{\varphi} = m^{\xi_\varphi}$ , and  $L_\varphi = \tilde{m}^{\xi_\varphi}$ , the co-Möbius transform, with similar conventions for  $\psi$ , see (Grabisch, 2016, Table A.2, p. 440).

$$= \sum_{B \subseteq \mathcal{X}} m^\mu(B) L_\varphi(B) + \sum_{A \subseteq \mathcal{Y}} m^\nu(A) L_\psi(A).$$

To see that the constraints are equivalent, notice that the above transformation can be inverted as

$$L_\varphi(G) = \sum_{G' \supseteq G} \hat{\varphi}(G'); \quad L_\psi(F) = \sum_{F' \supseteq F} \hat{\psi}(F').$$

Furthermore, for any  $B \subseteq \mathcal{X} \times \mathcal{Y}$ , recall

$$K_c(B) = \sum_{A \supseteq B} (-1)^{|A \setminus B|} \bigwedge_{(x,y) \in A} c(x,y).$$

Using the same inversion formula, we obtain

$$\bigwedge_{(x,y) \in A} c(x,y) = \sum_{B \supseteq A} K_c(B).$$

For any non-empty set  $A \subseteq \mathcal{X} \times \mathcal{Y}$ , sum all constraints with a right-hand side involving  $K_c(B)$  with  $B \supseteq A$ . The right-hand side term of (5.4) becomes

$$\sum_{B \supseteq A} K_c(B) = \bigwedge_{(x,y) \in A} c(x,y).$$

The sum of terms on the left-hand side of (5.4) will yield a sum involving  $\hat{\varphi}$ , which is

$$\sum_{G' \supseteq A_{\mathcal{X}}} \hat{\varphi}(G') = L_\varphi(A_{\mathcal{X}}),$$

and a sum involving  $\hat{\psi}$ , which is

$$\sum_{F' \supseteq A_{\mathcal{Y}}} \hat{\psi}(F') = L_\psi(A_{\mathcal{Y}}).$$

Lastly, denoting the sum of all terms involving  $\hat{\rho}$  in (5.4) by  $S$ , we obtain

$$S = J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + J_7,$$

where

$$\begin{aligned} J_1 &:= - \sum_{\substack{G' \supseteq A_{\mathcal{X}} \\ G' \neq \mathcal{X}}} \sum_{\substack{x \notin G' \\ y \in \mathcal{Y}}} \hat{\rho}(G' \times \mathcal{Y}, (x, y)); & J_2 &:= \sum_{\substack{G' \supseteq A_{\mathcal{X}} \\ G' \neq \mathcal{X}}} \sum_{\substack{x \in G' \\ y \in \mathcal{Y}}} \hat{\rho}(G' \times \mathcal{Y} \setminus \{(x, y)\}, (x, y)); \\ J_3 &:= - \sum_{\substack{F' \supseteq A_{\mathcal{Y}} \\ F' \neq \mathcal{Y}}} \sum_{\substack{y \notin F' \\ x \in \mathcal{X}}} \hat{\rho}(\mathcal{X} \times F', (x, y)); & J_4 &:= \sum_{\substack{F' \supseteq A_{\mathcal{Y}} \\ F' \neq \mathcal{Y}}} \sum_{\substack{y \in F' \\ x \in \mathcal{X}}} \hat{\rho}(\mathcal{X} \times F' \setminus \{(x, y)\}, (x, y)); \\ J_5 &:= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \hat{\rho}(\mathcal{X} \times \mathcal{Y} \setminus \{(x, y)\}, (x, y)); \\ J_6 &:= - \sum_{\substack{B \supseteq A \\ B \notin \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^* \\ (x, y) \notin B}} \hat{\rho}(B, (x, y)); & J_7 &:= \sum_{\substack{B \supseteq A \\ B \notin \mathcal{P}_{\mathcal{X}, \mathcal{Y}}^* \\ (x, y) \in B}} \hat{\rho}(B \setminus \{(x, y)\}, (x, y)). \end{aligned}$$

By summing the above terms, we obtain

$$\begin{aligned}
(5.11) \quad S &= - \sum_{B \supseteq A} \sum_{w \notin B} \hat{\rho}(B, w) + \sum_{D \supseteq A} \sum_{w \in D} \hat{\rho}(D \setminus \{w\}, w) \\
&= - \sum_{B \supseteq A} \sum_{w \notin B} \hat{\rho}(B, w) + \sum_{D \supseteq A} \left[ \sum_{w \in A} \hat{\rho}(D \setminus \{w\}, w) + \sum_{w \in D \setminus A} \hat{\rho}(D \setminus \{w\}, w) \right] \\
&= \sum_{D \supseteq A} \sum_{w \in A} \hat{\rho}(D \setminus \{w\}, w).
\end{aligned}$$

The last equality comes from the observation that there exists an one-to-one mapping between  $\{(B, w) : A \subset B, w \notin B\}$  and  $\{(D, w) : A \subset D, w \in D \setminus A\}$  by the map  $D := B \cup \{w\}$ , and thus the first and third terms in the second line of (5.11) cancel out. Therefore, one can derive the equations in (5.10) from those in (5.4). Similarly, one can also prove the opposite direction by using the above change of variables.  $\square$

**Remark 5.3.** By a similar argument, the dual of the maximization Optimal Transport problem is equivalent to

$$\begin{aligned}
(5.12) \quad & \min_{L_\varphi, L_\psi, \hat{\rho}} \sum_{G \subseteq \mathcal{X}} m^\mu(G) L_\varphi(G) + \sum_{F \subseteq \mathcal{Y}} m^\nu(F) L_\psi(F), \quad \text{subject to} \\
(5.13) \quad & L_\varphi(A_\mathcal{X}) + L_\psi(A_\mathcal{Y}) + \sum_{D \supseteq A} \sum_{w \in A} \hat{\rho}(D \setminus \{w\}, w) = \bigwedge_{(x,y) \in A} c(x, y), \quad \emptyset \neq A \subseteq \mathcal{X} \times \mathcal{Y}; \\
& \hat{\rho} \leq 0.
\end{aligned}$$

## 6. NUMERICAL EXAMPLES

### 6.1. A Comparison with the Optimal Transport Problem for Additive Measures.

In this section, we compare the optimal transport problem for capacities with the classical optimal transport problem (for measures) via numerical simulations. Assume that  $\mathcal{X}$  and  $\mathcal{Y}$  are two finite subsets of  $\mathbb{R}$  with  $|\mathcal{X}| = 30$  and  $|\mathcal{Y}| = 20$ , and  $\mu$  and  $\nu$  are probability measures on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. Given the quadratic function  $c(x, y) = (x - y)^2$  on  $\mathcal{X} \times \mathcal{Y}$ , the classical optimal transport problem is to find

$$\min_{\pi \in \Pi_a(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y);$$

while the optimal transport minimization problem for capacities seeks

$$\min_{\gamma \in \Pi_{\text{Ch}}(\mu, \nu)} \gamma(c),$$

where  $\gamma(c)$  represents the Choquet integral of  $c$  with respect to  $\gamma$ . The latter problem will have a lower minimum since its feasible set is larger.

We use the Python package AMPL to solve the linear program for the classical optimal transport minimization. However, the linear program for the optimal transport for capacities is quite large when both sets have cardinality greater than 5. For example, when  $|\mathcal{X}| = |\mathcal{Y}| = 5$ , the number of variables in the linear program is 33,554,432, and the number of constraints is 419,430,437. These numbers will become astronomical if  $|\mathcal{X}|$  and  $|\mathcal{Y}|$  exceed 20. For

the case when  $|\mathcal{X}| = 30$  and  $|\mathcal{Y}| = 20$ , the number of variables for the linear program is larger than  $10^{180}$ , and the number of constraints is larger than  $10^{183}$ ; while the number of variables for the classical optimal transport problem is 600 and the number of constraints is only 50. Therefore, in this case, solving the classical optimal transport problem using linear programming methods is still fast, but the linear program for capacities cannot be solved using numerical methods. However, using the explicit solution provided in Theorem 4.10, the minimum can be computed in a few seconds even when  $|\mathcal{X}| = |\mathcal{Y}| = 100$ .

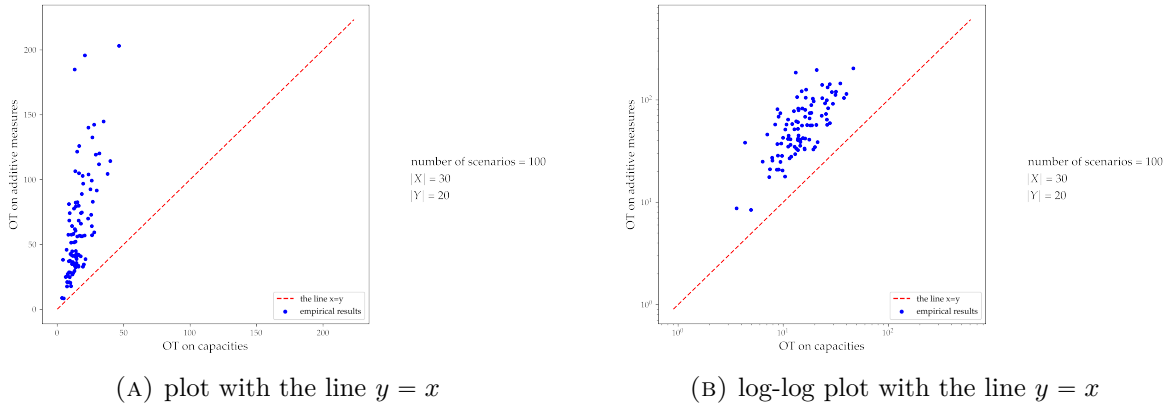


FIGURE 1. Comparison of the optimal values of the optimal transport problem for capacities with those of the optimal transport problem for additive measures, when both marginals are additive measures.

For the case  $|\mathcal{X}| = 30$  and  $|\mathcal{Y}| = 20$ , we run the following experiment 100 times. We consider the spaces  $\mathcal{X} = \{1, 2, 3, \dots, 30\}$  and  $\mathcal{Y} = \{0, 2.2, 4.4, 6.6, \dots, 41.8\}$ , as well as the cost function  $c(x, y) = (x - y)^2$ . To determine the marginal capacity  $\mu$ , we simulate  $|\mathcal{X}| - 1 = 29$  independent random variates from a uniform distribution on  $[0, 1]$ , and we let  $U_{(i)}, i = 1, \dots, 29$  be their order statistics (so that  $U_{(i)} \leq U_{(i+1)}$ , and  $U_{(1)}$  is the smallest observation). Then set  $U_{(0)} = 0$  and  $U_{(30)} = 1$ , and  $\mu(\{i\}) = U_{(i)} - U_{(i-1)}$ , for  $i = 1, \dots, 30$ . The  $\mu$  capacity (measure) of any other subset of  $\mathcal{X}$  is determined by additivity. An analogous method is used to simulate the marginal capacity  $\nu$  on  $\mathcal{Y}$ . For each pair of simulated capacities  $(\mu, \nu)$  generated in this fashion, we calculate the minima for both the optimal transport problem for measures and the one for capacities, and we compare the resulting optimal values.

The horizontal coordinates of the blue dots in Figure 1a represent the optimal values of the optimal transport problem for capacities; the vertical coordinates of the blue dots represent the minimum values of the classical optimal transport problem.

We observe a trend that the greater the distance between the two marginal distributions, the larger the ratio between OT minimum for measures over the OT minimum for capacities. This trend is better revealed by the log-log plot in Figure 1b, showing that the one optimal value appears to behave roughly like a power of the other. The difference between these two minima implies that the classical optimal transport minimum over probability measures is inaccurate in approximating the optimal transport minimum for capacities.

**6.2. An Application in Counterparty Credit Risk.** We consider a basic model in counterparty credit risk, similar to the one used in Ghossoub et al. (2023). Consider a bank that

trades with two counterparties whose credit exposures and the credit ratings at the end of the year determine the counterparty credit risk losses of the bank over the next year. For simplicity, we assume that there are four credit ratings, A, B, C, and D (default), with the transition probabilities in Table 1.<sup>8</sup>

Initial State	Year End State			
	A	B	C	D
A	0.990	0.007	0.002	0.001
B	0.030	0.950	0.015	0.005
C	0.015	0.020	0.960	0.005
D	0	0	0	1

TABLE 1. Transition probabilities for a simplified credit rating system.

Assume that the initial credit ratings of counterparties 1 and 2 are  $B$  and  $C$ , respectively. Due to ambiguity, we assume the joint rating  $Y = (Y_1, Y_2)$  of these two counterparties at the year-end is represented by a capacity on  $\mathcal{Y}$  of the form  $g \circ \mathbb{P}$  where  $g(x) := x^s$  is a concave distortion function with  $s \in (0, 1]$ <sup>9</sup>, and  $\mathbb{P}$  is the law of a joint probability distribution with a Gaussian copula. In particular, we let  $V = (V_1, V_2)$  be a two-dimensional Gaussian random vector with mean 0 and covariance matrix

$$\Sigma_V = \begin{pmatrix} 1 & \rho_y \\ \rho_y & 1 \end{pmatrix},$$

and define  $Y_i = F_i^{\leftarrow}(\Phi(V_i))$ ,  $i = 1, 2$ , where  $F_i$  is the marginal cumulative distribution function of  $Y_i$ . In particular, we have:

$$(6.1) \quad Y_1 = \begin{cases} D, & \text{if } V_1 \leq \Phi^{-1}(0.005); \\ C, & \text{if } \Phi^{-1}(0.005) \leq V_1 \leq \Phi^{-1}(0.02); \\ B, & \text{if } \Phi^{-1}(0.02) \leq V_1 \leq \Phi^{-1}(0.97); \\ A, & \text{if } V_1 \geq \Phi^{-1}(0.97), \end{cases}$$

where  $\Phi$  is the standard normal cumulative distribution function.  $Y_2$  is defined similarly.

The cardinality of  $\mathcal{Y}$  is 16. The probability of each pair of credit ratings can be calculated using the bivariate Gaussian distribution. For example

$$P(Y_1 = D, Y_2 = D) = \Phi_2(\Phi^{-1}(0.005), \Phi^{-1}(0.005); \rho_y),$$

where  $\Phi_2$  is the bivariate normal cumulative distribution function.

We assume that each counterparty exposure has a (marginal) binomial distribution. In particular, we suppose that counterparty 1 has exposure  $X_1$  that follows  $\text{binomial}(n_1, p_1)$  and counterparty 2 has exposure  $X_2$  that follows  $\text{binomial}(n_2, p_2)$ . The random vector  $(X_1, X_2)$  is taken to have a Gaussian copula with correlation  $\rho_x$ . We denote the corresponding probability distribution on  $\mathcal{X}$  by  $\mathbb{Q}$ , and we assume that the marginal capacity  $\mu = \mathbb{Q}$  (i.e., there is no distortion, or for the exposure capacity  $s = 1$ ).

<sup>8</sup>Table 1 is borrowed from [Hardy and Saunders \(2022\)](#).

<sup>9</sup>In particular, when  $s = 1$ , this capacity is the same as the additive measure  $P$ .

We take  $n_1 = 40, p_1 = 0.4, n_2 = 25,$  and  $p_2 = 0.7.$  Then the cardinality of  $\mathcal{X}$  is  $(n_1 + 1) \cdot (n_2 + 1) = 1066.$  Again, here the joint probabilities  $P((X_1, X_2) = (n, m))$  for  $0 \leq n \leq n_1$  and  $0 \leq m \leq n_2$  can be calculated using the cumulative bivariate Gaussian distribution.

Finally, we describe the loss function, which is the sum of the losses due to the credit migrations of each counterparty:

$$(6.2) \quad L(X, Y) = X_1 \cdot h(Y_1) + X_2 \cdot h(Y_2),$$

where the function  $h : \{A, B, C, D\} \rightarrow [0, 1]$  represents the fraction of total exposure that will be lost in the next year, given the credit rating at the year-end. In this example, we take  $h(A) = 0, h(B) = 0.1, h(C) = 0.2, h(D) = 1$  (default).

We look for the maximum risk represented by a Choquet integral of the loss function  $L$  against a capacity  $\gamma$  with prescribed marginal capacities  $\mu$  and  $\nu = g \circ \mathbb{P},$  as described above above. That is,

$$(6.3) \quad \max_{\gamma \in \Pi(\mu, \nu)} \gamma(L) = \max_{\gamma \in \Pi(\mu, g \circ P)} \gamma(L).$$

Note that, unlike the optimization problems in Section 6.1, one of the given marginals in (6.3) is non-additive.

Figure 2 shows how the maximum varies along with changes in (a) the power  $s$  in the distortion function, (b) the correlation  $\rho_x,$  or (c) the correlation  $\rho_y$  in the Gaussian copula. We observe that the maximum risk is a decreasing and convex function of the distortion parameter  $s.$  This accords with intuition, as the smaller the parameter  $s,$  the greater uncertainty there is regarding the marginal distribution of the credit risk factors. With the other parameters fixed, the maximum Choquet risk is an increasing and concave function of the correlation in the copula defining the distribution of the exposure factors  $\mu.$  Again, this makes financial sense given the nature of our loss function. If the exposures were negatively correlated, then an increase in the first term in the loss function  $L$  in (6.2) would tend to correspond to a decrease in the second term. This diversification effect is amplified the greater the magnitude of the negative correlation. Similarly, if  $\rho_x$  is large and positive, then the tail of the losses will be fatter due to the tendency for large exposures to both counterparties to occur simultaneously.

Perhaps most interesting is Figure 2c, which considers the impact of the correlation parameter of the copula of the credit risk factors on the maximum Choquet risk, as it reveals the nature (and underlying conservatism) of the Choquet risk measure corresponding to the capacity  $\pi^*.$  The most striking aspect of the plot in Figure 2c is that it is not monotone; there is an interior maximum of  $\pi^*(L)$  as a function of  $\rho_y.$  To understand this, we consider explicitly a simplified version of the model. In particular, we take  $X_1$  to be binomial with parameters  $n_1 = 2$  and  $p_1 = 0.4,$   $X_2$  to be binomial with  $n_2 = 2$  and  $p_2 = 0.7,$  and  $\rho_x = -0.3.$  Based on this specification, we can compute the probabilities for the joint distribution of the exposures  $(X_1, X_2)$  given in Table 2.

We further simplify the model by assuming only two credit states, default and no default, with both firms starting in the no-default state, and with default probability  $PD = 0.005.$

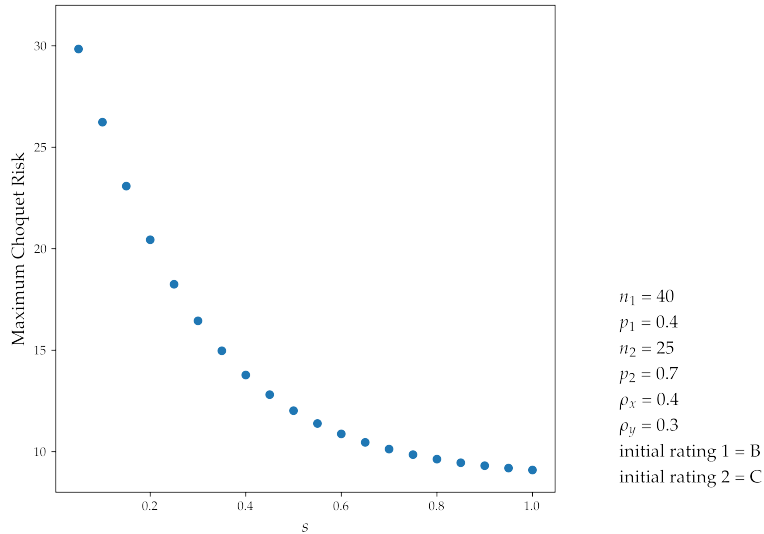
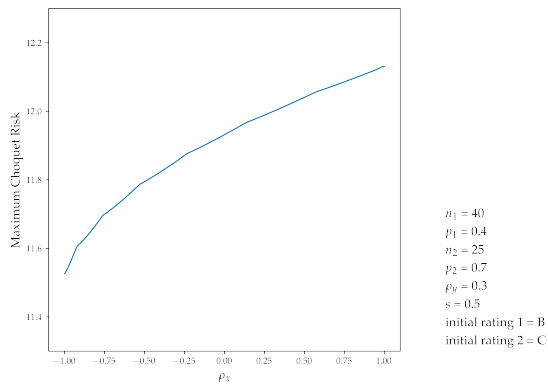
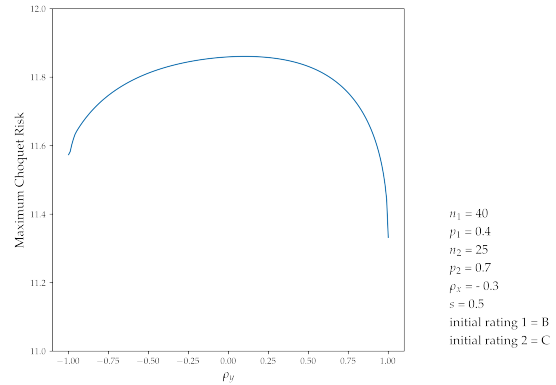
(A) Maximum Choquet risk as the distortion parameter  $s$  varies(B) Maximum Choquet risk as  $\rho_x$  varies(C) Maximum Choquet risk as  $\rho_y$  varies

FIGURE 2. Sensitivity Analysis of the Maximum Choquet Risk for the counterparty credit risk example.

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$X_1 = 0$	0.09	0.27	0
$X_1 = 1$	0	0.15	0.33
$X_1 = 2$	0	0	0.16

TABLE 2. Joint exposure probabilities for the simplified counterparty credit risk example.

The probability of both counterparties defaulting together is then

$$(6.4) \quad p_{DD}(\rho_y) = \Phi_2(\Phi^{-1}(0.005), \Phi^{-1}(0.005); \rho_y),$$

while the probability of at least one of the counterparties defaulting is  $p_D = 0.01 - p_{DD}(\rho_y)$ .

Let  $U^t = \{L \geq t\}$ , so that  $\pi^*(U^t) = \min(\mu(U_{\mathcal{X}}^t), \nu(U_{\mathcal{Y}}^t)) = \min(Q(U_{\mathcal{X}}^t), \sqrt{P(U_{\mathcal{Y}}^t)})$ . Here  $U_{\mathcal{X}}^t$  is the set of  $(x_1, x_2)$  for which there is *some* scenario for the credit factor  $Y$  such that  $L(X, Y) \geq t$ . Since we can take the credit scenario to be as extreme as possible (both counterparties default), reflecting the inherent conservatism in  $\pi^*$ , we see that



$U_{\mathcal{X}}^t = \{(x_1, x_2) : x_1 + x_2 \geq t\}$ . Simple calculations with the bivariate normal distribution with  $\rho_x = -0.3$  then lead to the data in Table 3.

$t$ values	$U_{\mathcal{X}}^t$	$\mu(U_{\mathcal{X}}^t)$
$t > 4$	$\emptyset$	0
$3 < t \leq 4$	$\{(2, 2)\}$	0.0494
$2 < t \leq 3$	$\{(1, 2), (2, 1), (2, 2)\}$	0.35
$1 < t \leq 2$	$\{(1, 0), (0, 1), (0, 0)\}^c$	0.8162
$0 < t \leq 1$	$\{(0, 0)\}^c$	0.9843
$t \leq 0$	$\mathcal{X}$	1

TABLE 3. Sets  $U_{\mathcal{X}}^t$  and their capacities for the simplified counterparty credit risk example.

Similarly, when considering  $U_{\mathcal{Y}}^t$ , we take the worst-case exposure scenario  $X_1 = X_2 = 2$ , and find that  $U_{\mathcal{Y}}^t = \{(y_1, y_2) : h(y_1) + h(y_2) \geq \frac{t}{2}\}$ . Recalling that  $\nu(U_{\mathcal{Y}}^t) = \sqrt{P(U_{\mathcal{Y}}^t)}$ , we obtain the data in Table 4.

$t$ values	$U_{\mathcal{Y}}^t$	$\nu(U_{\mathcal{Y}}^t)$
$t > 4$	$\emptyset$	0
$2 < t \leq 4$	$\{(D, D)\}$	$\sqrt{p_{DD}(\rho_y)}$
$0 < t \leq 2$	$\{(A, A)\}^c$	$\sqrt{0.01 - p_{DD}(\rho_y)}$
$t \leq 0$	$\mathcal{Y}$	1

TABLE 4. Sets  $U_{\mathcal{Y}}^t$  and their capacities for the simplified counterparty credit risk example.

A simple calculation then yields:

$$\begin{aligned}
\max_{\gamma \in \Pi_{\text{Ch}}(\mu, \nu)} \gamma(L) &= \int_0^\infty \pi^*(L \geq t) dt \\
&= \int_0^4 \min(\mu(U_{\mathcal{X}}^t), \nu(U_{\mathcal{Y}}^t)) dt \\
&= 2\sqrt{0.01 - p_{DD}(\rho_y)} + \sqrt{p_{DD}(\rho_y)} + 0.0494,
\end{aligned}$$

and it can be seen that this function has an interior maximum (as a function of  $\rho_y$  on  $[-1, 1]$ ). It is interesting to note that this behaviour depends on the parameters of our model, such as the probabilities of the most extreme exposure and credit scenarios. For example, with  $\rho_x = 1$  instead of  $\rho_x = -0.3$ , similar calculations give that  $\pi^*(L) = 2(\sqrt{0.01 - p_{DD}(\rho_y)} + \sqrt{p_{DD}(\rho_y)})$ , which is monotone increasing in  $\rho_y$ .

**6.3. Comparison of Maximum Expected Shortfall and Maximum Choquet Risk with Expected Shortfall Marginal Risks.** In this subsection, we will compare the Choquet risk measure defined in the current paper with the Maximum Expected Shortfall (MES) studied in Ghossoub et al. (2023).

For a given loss random variable  $L$  defined on  $\mathcal{X} \times \mathcal{Y}$ , and for prescribed marginal probability measures  $\mu$  on  $\mathcal{X}$  and  $\nu$  on  $\mathcal{Y}$ , the *maximum expected shortfall* at confidence level  $\alpha$

associated with  $L$  is defined as

$$\text{MES}_\alpha(L) := \sup_{\pi \in \Pi_\alpha(\mu, \nu)} \text{ES}_{\alpha, \pi}(L),$$

where  $\text{ES}_{\alpha, \pi}$  is the expected shortfall with respect to the probability measure  $\pi \in \Pi_\alpha(\mu, \nu)$ . In contrast to the maximum Choquet risk measure problem studied in this paper, when determining  $\text{MES}_\alpha$ :

- The marginal probability distributions of the risk factors on  $\mathcal{X}$  and  $\mathcal{Y}$  are assumed to be known with certainty (in contrast to the case of marginal capacities, which may represent ambiguity about these marginal distributions).
- The joint risk measure is restricted to be the expected shortfall computed with respect to some probability measure  $\pi \in \Pi_\alpha$  (in contrast to the maximum Choquet risk measure problem, in which we consider all possible Choquet risk measures on  $\mathcal{X} \times \mathcal{Y}$  that match the given marginal Choquet risk measures on  $\mathcal{X}$  and  $\mathcal{Y}$ ).

Since expected shortfall is a distortion risk measure, the MES can be written as:

$$(6.5) \quad \text{MES}_\alpha(L) = \sup_{\pi \in \Pi_\alpha(\mu, \nu)} \int L dg_\alpha(\pi),$$

where

$$g_\alpha(x) = \begin{cases} \frac{x}{1-\alpha}, & x \in [0, 1-\alpha), \\ 1, & x \in [1-\alpha, 1], \end{cases}$$

is the corresponding distortion function.

Explicitly, the Choquet Maximum Expected Shortfall (MES) can be defined as the maximum Choquet integral of the loss function against capacities with the same marginals as  $g_\alpha(\pi)$ .

$$(6.6) \quad \text{CMES}_\alpha(L) := \sup_{\gamma \in \Pi(g_\alpha(\mu), g_\alpha(\nu))} \int L d\gamma.$$

Since the feasible set for the maximum Choquet risk measure problem contains the feasible set for the maximum expected shortfall problem, we have that  $\text{CMES}_\alpha(L) \geq \text{MES}_\alpha(L)$ . In Figure 3, we compare the values of  $\text{CMES}_\alpha(L)$  and  $\text{MES}_\alpha(L)$  for the loss random variable  $L$  in the counterparty credit risk example described in the above subsection with different  $\rho_x$  and  $\rho_y$ .

Throughout the experiments, we fix  $\alpha = 0.9$ . Assume that the counterparty 1 has initial rating B and exposure  $X_1$  that follows binomial(40, 0.4) and that counterparty 2 has initial rating C and exposure  $X_2$  that follows binomial(25, 0.7). In Figure 3a, the probability  $\nu$  corresponds to the law of joint rating  $Y = (Y_1, Y_2)$ , which can be calculated using bivariate Gaussian distribution with correlation factor  $\rho_y = 0.25$ ; similarly, the probability  $\mu$  corresponds to the law of counterparty exposures  $X = (X_1, X_2)$ , which can be determined using bivariate Gaussian distributions with correlation factor  $\rho_x$  varying from  $-1$  to  $1$ . We plot both risk measures over different correlation factor  $\rho_x$ . When fixing  $\rho_x = 0.35$  and allowing  $\rho_y$  change from  $-1$  to  $1$ , we obtain Figure 3b.

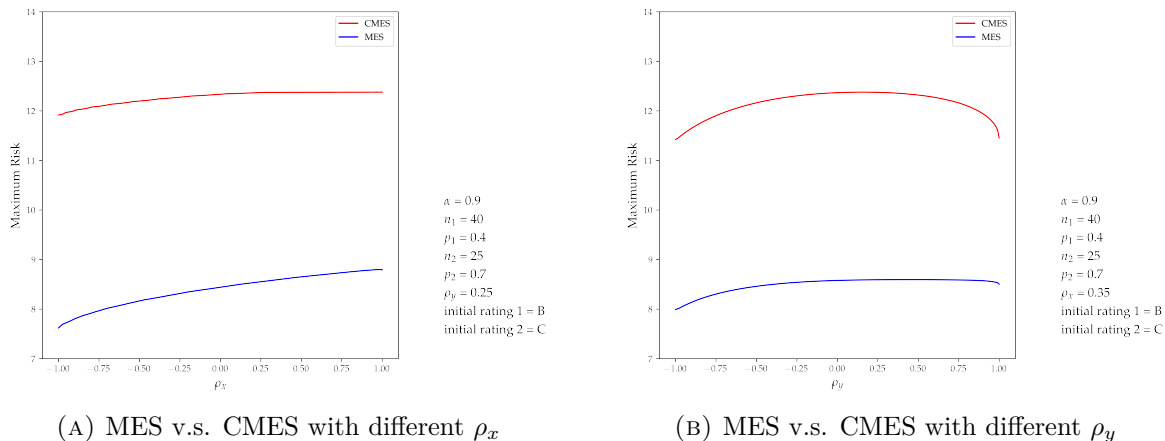


FIGURE 3. Comparison between the Maximum Expected Shortfall (MES) and the Choquet Maximum Expected Shortfall (CMES) using different levels of the correlation parameters  $\rho_x$  and  $\rho_y$ .

From the figures, one can also observe that the ratio of  $\text{CMES}_{0.9}(L)$  over  $\text{MES}_{0.9}(L)$  is between 130% to 160%. This ratio depends on the parameter  $\alpha$  and the two given distributions,  $\mu$  and  $\nu$ , which are eventually determined by the parameters  $n_1, p_1, n_2, p_2, \rho_x, \rho_y$ , and the values in Table 1.

## 7. CONCLUSION

This paper investigates the problem of bounding a Choquet risk measure of a nonlinear function of two risk factors. Specifically, we assume given (marginal) capacities on the marginal spaces, representing the ambiguous distributions of the risk factors, and we consider the problem of finding the joint capacity on the product space with these given marginals, which maximizes or minimizes the Choquet integral of a given portfolio loss function.

We treat this problem as a generalization of the optimal transport problem to the setting of nonadditive measures. We provide explicit characterizations of the optimal solutions for finite marginal spaces, and we investigate some of their properties. Furthermore, we investigate the relationship between properties of the marginal capacities and those of the optimizers (and, more generally, capacities in the feasible set). In particular, we show that the minimizing capacity  $\pi_*$  is balanced if and only if both marginal capacities are balanced, and we describe its core explicitly in that case. In contrast, in all but the most trivial cases, the maximizing capacity  $\pi^*$  is not balanced.

We further discuss the connections with linear programming, showing that the optimal transport problems for capacities are linear programs, and we also characterize their duals explicitly. We investigate a series of numerical examples, including a comparison with the classical optimal transport problem, and applications to counterparty credit risk.

## REFERENCES

- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7), 1505–1518.
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.
- Bauer, C. (2012). Products of non-additive measures: a Fubini-like theorem. *Theory and Decision*, 73, 621–647.
- Brigo, D., Morini, M., & Pallavicini, A. (2013). *Counterparty credit risk, collateral and funding: With pricing cases for all asset classes*. Hoboken: Wiley.
- Carlier, G., & Zhang, K. S. (2020). Existence of solutions to principal–agent problems with adverse selection under minimal assumptions. *Journal of Mathematical Economics*, 88, 64–71.
- Chen, Z., Hu, H., & Jiang, J. (2022). Convergence analysis on data-driven fortet-mourier metrics with applications in stochastic optimization. *Sustainability*, 14(8), 4501.
- Delbaen, F. (2002). Coherent risk measures on general probability spaces. In K. Sandmann & P. Schönbucher (Eds.), *Advances in finance and stochastics. essays in honour of dieter sondermann* (pp. 1–37). Springer.
- Denneberg, D. (1994). *Non-additive measure and integral* (Vol. 27). Springer Science & Business Media.
- Destercke, S. (2013). Independence and 2-monotonicity: Nice to have, hard to keep. *International Journal of Approximate Reasoning*, 54, 478–490.
- Dunford, N., & Schwartz, J. (1958). *Linear Operators, Part 1: General Theory*. Wiley-Interscience.
- Dyckerhoff, R. (2022). Personal communication.
- Eckstein, S., Guo, G., Lim, T., & Obłój, J. (2021). Robust pricing and hedging of options on multiple assets and its numerics. *SIAM Journal on Financial Mathematics*, 12(1), 158–188.
- Föllmer, H., & Schied, A. (2016). *Stochastic finance: An introduction in discrete time* (Fourth ed.). Berlin: de Gruyter.
- Galichon, A. (2016). *Optimal transport methods in economics*. Princeton: Princeton University Press.
- Garcia-Cespedes, J., de Juan Herrero, J., Rosen, D., & Saunders, D. (2010). Effective modelling of wrong-way risk, CCR capital and alpha in Basel II. *Journal of Risk Model Validation*, 4(1), 71–98.
- Ghirardato, P. (1997). On independence for non-additive measures, with a Fubini theorem. *Journal of Economic Theory*, 73(2), 261–291.
- Ghossoub, M., Hall, J., & Saunders, D. (2023). Maximum Spectral Measures of Risk with Given Risk Factor Marginal Distributions. *Mathematics of Operations Research*, 48(2), 1158–1182.
- Glasserman, P., & Yang, L. (2018). Bounding wrong-way risk in CVA calculation. *Mathematical Finance*, 28, 268–305.
- Grabisch, M. (2016). *Set functions, games, and capacities in decision making*. Switzerland: Springer.

- Gregory, J. (2020). *The xva challenge: Counterparty risk, funding, collateral, capital and initial margin*. John Wiley & Sons.
- Guillen, N., & Kitagawa, J. (2017). Pointwise estimates and regularity in geometric optics and other generated jacobian equations. *Communications on Pure and Applied Mathematics*, 70(6), 1146–1220.
- Hardy, M., & Saunders, D. (2022). *Quantitative enterprise risk management*. Cambridge, U.K.: Cambridge University Press.
- Hendon, E., Jacobsen, H., Sloth, B., & Tran, T. (1991). *The product of capacities and lower probabilities* (Tech. Rep.). mimeo, University of Copenhagen.
- Henry-Labordère, P. (2017). *Model-free hedging: A martingale optimal transport viewpoint*. Boca Raton: CRC Press.
- Kantorovich, L. (1942). On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37, 199–201.
- Kantorovich, L. (1948). On a problem of Monge (in russian). *Uspekhi Math. Nauk.*, 3, 225–226.
- Korman, J., & McCann, R. (2015). Optimal transportation with capacity constraints. *Transactions of the American Mathematical Society*, 367(3), 1501–1521.
- Korman, J., McCann, R., & Seis, C. (2015). Dual potentials for capacity constrained optimal transport. *Calculus of Variations and Partial Differential Equations*, 54(1), 573–584.
- Koshevoy, G. (1998). Distributive lattices and products of capacities. *Journal of Mathematical Analysis and Applications*, 219, 427–441.
- Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3, 83–95.
- Marinacci, M., & Montrucchio, L. (2004). Introduction to the mathematics of ambiguity. *Uncertainty in Economic Theory, I. Gilboa (ed.)*, 46–107.
- McCann, R. (2020). Displacement convexity of boltzmann’s entropy characterizes the strong energy condition from general relativity. *Camb. J. Math.*, 8(3), 609–681.
- McCann, R. J., & Zhang, K. S. (2019). On concavity of the monopolist’s problem facing consumers with nonlinear price preferences. *Communications on Pure and Applied Mathematics*, 72(7), 1386–1423.
- McNeil, A., Frey, R., & Embrechts, P. (2015). *Quantitative risk management* (Second ed.). Princeton: Princeton University Press.
- Memartoluie, A., Saunders, D., & Wirjanto, T. (2017). Wrong-way risk bounds in counterparty credit risk management. *Journal of Risk Management in Financial Institutions*, 10(2), 150–163.
- Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais (Dissertation on the theory of cuttings and embankments). *Mem. Math. Phys. Acad. Royale Sci.*, 666–704.
- Panaretos, V., & Zemel, Y. (2022). *An invitation to statistics in Wasserstein space*. Cham: Springer.
- Pennanen, T., & Perkkiö, A.-P. (2019). Convex duality in nonlinear optimal transport. *Journal of Functional Analysis*, 277(14), 1029–1060.
- Peyré, G., & Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6), 355–607.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior &*

- Organization*, 3(4), 323–343.
- Quiggin, J. (1993). *Generalized expected utility theory - the rank-dependent model*. Kluwer Academic Publishers.
- Rachev, S., & Rüschendorf, L. (1998). *Mass transportation problems*. New York: Springer.
- Rachev, S. T., & Rüschendorf, L. (1998). *Mass transportation problems: Volume i: Theory* (Vol. 1). Springer Science & Business Media.
- Rosen, D., & Saunders, D. (2010). Computing and stress testing counterparty credit risk capital. In E. Canabarro (Ed.), *Counterparty credit risk* (pp. 245–292). RiskBooks.
- Rosen, D., & Saunders, D. (2012). CVA the wrong way. *Journal of Risk Management in Financial Institutions*, 5(3), 252–272.
- Rüschendorf, L. (2013). *Mathematical risk analysis*. Berlin: Springer.
- Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63), 94.
- Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, 255 – 261.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 571–587.
- Shapiro, A. (2013). On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38(1), 142–152.
- Torra, V. (2023). The transport problem for non-additive measures. *European Journal of Operational Research*. doi: <https://doi.org/10.1016/j.ejor.2023.03.016>
- Torres, L., Pereira, L., & Amini, M. (2021). A survey on optimal transport for machine learning: Theory and applications. *arXiv preprint arXiv:2106.01963*.
- Villani, C. (2003). *Topics in optimal transportation*. Providence: American Mathematical Society.
- Villani, C. (2008). *Optimal transport, old and new*. Berlin: Springer.
- Walley, P., & Fine, T. (1982). Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3), 741–761.
- Yaari, M. (1987). The dual theory of choice under risk. *Econometrica*, 95–115.
- Zhang, K., Peyré, G., Fadili, J., & Pereyra, M. (2020). Wasserstein control of mirror langevin monte carlo. In *Conference on learning theory* (pp. 3814–3841).

**Mario Ghossoub:** UNIVERSITY OF WATERLOO – DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE  
 – 200 UNIVERSITY AVE. W. – WATERLOO, ON, N2L 3G1 – CANADA  
*Email address:* [mario.ghossoub@uwaterloo.ca](mailto:mario.ghossoub@uwaterloo.ca)

**David Saunders:** UNIVERSITY OF WATERLOO – DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE  
 – 200 UNIVERSITY AVE. W. – WATERLOO, ON, N2L 3G1 – CANADA  
*Email address:* [dsaunders@uwaterloo.ca](mailto:dsaunders@uwaterloo.ca)

**Kelvin Shuangjian Zhang:** UNIVERSITY OF WATERLOO – DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE  
 – 200 UNIVERSITY AVE. W. – WATERLOO, ON, N2L 3G1 – CANADA  
*Email address:* [ks3zhang@uwaterloo.ca](mailto:ks3zhang@uwaterloo.ca)