

Distorted optimal transport

Haiyan Liu^{*} Bin Wang[†] Ruodu Wang[‡] Sheng Chao Zhuang[§]

May 17, 2025

Abstract

Classic optimal transport theory is formulated through minimizing the expected transport cost between two given distributions. We propose the framework of distorted optimal transport by minimizing a distorted expected cost, which is the cost under a non-linear expectation. This new formulation is motivated by concrete problems in decision theory, robust optimization, and risk management, and it has many distinct features compared to the classic theory. We choose simple cost functions and study different distortion functions and their implications on the optimal transport plan. We show that on the real line, the comonotonic coupling is optimal for the distorted optimal transport problem when the distortion function is convex and the cost function is submodular and monotone. Some forms of duality and uniqueness results are provided. For inverse-S-shaped distortion functions and linear cost, we obtain the unique form of optimal coupling for all marginal distributions, which turns out to have an interesting “first comonotonic, then counter-monotonic” dependence structure; for S-shaped distortion functions a similar structure is obtained. Our results highlight several challenges and features in distorted optimal transport, offering a new mathematical bridge between the fields of probability, decision theory, and risk management.

Keywords: comonotonicity, distorted expectation, copulas, Choquet integral, risk aggregation.

^{*}Department of Mathematics and Department of Statistics and Probability, Michigan State University, United States. E-mail: liuhaiy1@msu.edu.

[†]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. E-mail: wangbin@amss.ac.cn.

[‡]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada. E-mail: wang@uwaterloo.ca.

[§]Department of Finance, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588, United States. E-mail: szhuang3@unl.edu.

1 Introduction

Optimal transport theory, originally developed by Monge and Kantorovich (see Villani (2009) for a history), has been an important field of study over the past several decades, by connecting various scientific fields, including probability, economic theory, operations research, statistics, machine learning, and quantitative finance. For a mathematical background on optimal transport and its applications, we refer to the monographs of Santambrogio (2015) and Villani (2009). A specialized treatment of optimal transport in economics is given by Galichon (2016).

Classic optimal transport theory concerns the minimization of the expected cost of transporting between two marginal distributions, that is, using a probabilistic formulation,

$$\text{to minimize } \mathbb{E}[c(X, Y)] \text{ subject to } X \stackrel{d}{\sim} \mu \text{ and } Y \stackrel{d}{\sim} \nu,$$

where μ and ν are two probability measures on suitable spaces, $X \stackrel{d}{\sim} \mu$ means that X has the distribution μ under a fixed probability measure \mathbb{P} with respect to which the expectation \mathbb{E} is evaluated, and c is a real-valued cost function. The joint distribution of (X, Y) is called a transport plan or a coupling in the literature.

In this paper, we consider a novel setting of minimizing *non-linear expectations* (e.g., Peng (2019)) of the transport cost. With a non-linear expectation \mathcal{E} replacing \mathbb{E} in the classic formulation, the corresponding optimal transport problem is

$$\text{to minimize } \mathcal{E}[c(X, Y)] \text{ subject to } X \stackrel{d}{\sim} \mu \text{ and } Y \stackrel{d}{\sim} \nu. \tag{1}$$

We will mainly focus on a popular and tractable class of non-linear expectations, the *distorted expectations*, denoted by \mathcal{E}^h , where h is a *distortion function* (see Section 2 for its definition). The *distorted optimal transport* problem is (1) with $\mathcal{E} = \mathcal{E}^h$. The non-linearity of \mathcal{E}^h leads to many mathematical challenges that are not treated in the classic theory.

In addition to its mathematical novelty, we explain in Section 2 three different motivations for using non-linear and distorted expectations in optimal transport, which arise from the fields of decision theory (e.g., Quiggin (1982); Yaari (1987)), robust optimization (e.g., Gilboa and Schmeidler (1989); Goh and Sim (2010)) and risk management (e.g., Ruschendorf (2013); Embrechts et al. (2014)), respectively. In particular, non-linear and distorted expectations are useful tools to model preferences under risk, regulatory risk measures, and transport cost under uncertainty. Thus, the study of distorted transport brings together ideas and techniques from these fields.

As the first study on distorted optimal transport, we will focus on the basic setting of transport

between two distributions on the real line. We work mainly with simple forms of cost functions, and analyze different distortion functions and their implications on the optimal transport problem. The optimal couplings will be described by using copulas (Nelsen (2006)), which are convenient in our setting. As a well-known result in classic optimal transport theory, if the cost function on the real plane is submodular, then an optimal transport plan is the comonotonic coupling. Moreover, such a transport plan is *universally optimal*, in the sense that its optimality does not depend on the marginal distributions. In the setting of distorted transport, we show that if the distortion function is convex and the cost function is submodular and monotone (i.e., increasing component-wise or decreasing component-wise), then a universally optimal transport plan is the comonotonic coupling (Theorem 1). This result highlights two distinct features of distorted optimal transport compared to the classic setting. First, convexity of the distortion function h is essential for the optimal transport plan. The classic setting corresponds to a linear h , thus both convex and concave. Second, monotonicity of the cost function is important for the optimal transport plan. Note that monotonicity is irrelevant to the classic transport problem, since the optimal coupling is invariant if we add to the cost function some terms that depend only on one of the variables. This is not true in the distorted transport setting, because the distorted expectation is not linear, and additive terms that only depend on one variable generally affect the optimal coupling. Moreover, we establish a weak version of duality for convex h , but strong duality holds only in case of an affine cost (Theorem 2). These results are discussed in Section 4.

We proceed to concentrate on the linear cost $c(x, y) = x + y$, which is submodular, supermodular, and monotone, in Sections 5-7. The linear cost represents an additive structure, and it is the standard model in decision making with background risk (see e.g., Pratt (1988)) and risk aggregation problems (e.g., Embrechts et al. (2013)). This cost function is trivial in the classic transport theory, but it yields some clean and nontrivial results in the framework of distorted optimal transport. With the linear cost, we show uniqueness of the universally optimal coupling based on strict convexity or concavity of h (Theorem 3). This and some other general results on the linear cost are presented in Section 5. In decision theory, abundant research has elicited distortion functions from experimental data of human behavior. A special form of distortion functions that is empirically plausible is the inverse S-shaped distortion functions (Edwards (1954)), which plays a central role in cumulative prospect theory (Tversky and Kahneman (1992)).¹ We pay special attention to this class of distortion functions in Section 6. Such distortion functions are neither convex nor concave,

¹Kahneman was awarded the Nobel Memorial Prize in Economic Sciences in 2002 “for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty.”

leading to considerable technical challenges. In this setting, we obtain equivalent conditions for the existence of universally optimal couplings, which turn out to have an interesting “first comonotonic then counter-monotonic” structure (Theorem 4). In contrast to inverse-S-shaped distortion functions in decision theory, S-shaped distortion functions are closely related to popular risk measures in risk management. Universally optimal couplings for S-shaped distortion functions (Theorem 5) and its connection to robust risk aggregation are discussed in Section 7.

The framework of distorted optimal transport is new and many questions remain unanswered. Section 9 concludes the paper with discussions on several open questions and future research directions.

2 Non-linear and distorted expectations

We first define non-linear and distorted expectations. Fix an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and let \mathcal{X} be the set of bounded random variables. A *non-linear expectation* \mathcal{E} is a mapping from \mathcal{X} to \mathbb{R} satisfying

- (i) Monotonicity: $\mathcal{E}[X] \leq \mathcal{E}[Y]$ if $X \leq Y$;
- (ii) Constant preserving: $\mathcal{E}[c] = c$ for $c \in \mathbb{R}$;
- (iii) Law invariance: $\mathcal{E}[X] = \mathcal{E}[Y]$ if X and Y are identically distributed.

In Peng (2019), only the first two properties are required in the definition. In optimal transport problems, only the distribution of $c(X, Y)$ matters, because the constraints on X and Y are specified by marginal distributions; therefore, it is without loss of generality to assume law invariance.² A *sub-linear expectation* \mathcal{E} is a non-linear expectation that further satisfies

- (iv) Subadditivity: $\mathcal{E}[X + Y] \leq \mathcal{E}[X] + \mathcal{E}[Y]$ for all $X, Y \in \mathcal{X}$;
- (v) Positive homogeneity: $\mathcal{E}[\lambda X] = \lambda \mathcal{E}[X]$ for $\lambda \geq 0$.

Similarly, a *super-linear expectation* \mathcal{E} is a non-linear expectation that satisfies positive homogeneity and $-\mathcal{E}$ is subadditive.

We will focus on the popular class of non-linear expectations, the distorted expectations. A *distortion function* is an increasing (in the non-strict sense) function $h : [0, 1] \rightarrow [0, 1]$ satisfying

²Even if we use a function $\mathcal{E} : \mathcal{X} \rightarrow \mathbb{R}$ that is not law invariant, the optimization in (1) is equivalent to using a law-invariant function $\hat{\mathcal{E}}$ given by $\hat{\mathcal{E}}[Z] = \inf\{\mathcal{E}[Z'] : Z' \text{ is identically distributed to } Z\}$. Hence, requiring law invariance does not lose generality.

$h(0) = 1 - h(1) = 0$. The set of distortion functions is denoted by \mathcal{H} . The *distorted expectation* associated with distortion function h is $\mathcal{E}^h : \mathcal{X} \rightarrow \mathbb{R}$ defined by

$$\mathcal{E}^h[X] = \int_0^\infty h(\mathbb{P}(X > x))dx + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - 1)dx. \quad (2)$$

For an unbounded random variable X , a distorted expectation can also be defined via (2), which may be infinite or undefined. We focus on bounded random variables in this paper to illustrate the main ideas, while keeping in mind that most results also hold on more general spaces, provided that $\mathcal{E}^h[X]$ is well defined. A distorted expectation is a non-linear expectation, and it is sub-linear if and only if h is concave (see Lemma 2 below). Moreover, any sub-linear expectation on \mathcal{X} can be written as the supremum of sub-linear distorted expectations, as shown by Kusuoka (2001).

Distorted expectations are used in many different fields under different names, and they belong to the general class of Choquet integrals characterized by Schmeidler (1986, 1989). We refer to the monographs of Denneberg (1994), Wakker (2010) and Föllmer and Schied (2016) for general theories of the Choquet integral and distorted expectations in economics and finance.

In decision theory, the distortion function h represents a subjective inflation/deflation of the true probability. If the random variable X is nonnegative, then (2) reduces to

$$\mathcal{E}^h[X] = \int_0^\infty h(\mathbb{P}(X > x))dx.$$

If the distortion function h is left-continuous, through a change of variable, \mathcal{E}^h can be written as (see e.g., Wang et al. (2020, Lemma 3))

$$\mathcal{E}^h[X] = \int_0^1 Q_{1-t}(X)dh(t), \quad (3)$$

where

$$Q_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq p\}, \quad p \in (0, 1];$$

i.e., $Q_p(X)$ is the left-quantile of X at $p \in (0, 1]$. From (3), it is clear that if h is the identity, then $\mathcal{E}^h[X] = \mathbb{E}[X]$, which is the expectation with respect to \mathbb{P} . The formula (3) will be frequently used in the paper, and it has the natural interpretation that a distorted expectation puts different weights on different quantile levels. The left-quantile Q_p itself can be expressed as a distorted expectation, $Q_p = \mathcal{E}^h$ where $h(t) = \mathbf{1}_{\{t > 1-p\}}$ for $t \in [0, 1]$. Another common example of \mathcal{E}^h , popular in finance,

is the Expected Shortfall (ES) at level $p \in [0, 1)$, defined as

$$\text{ES}_p(X) = \frac{1}{1-p} \int_p^1 Q_r(X) dr \quad \text{for } X \in \mathcal{X}. \quad (4)$$

It holds that $\text{ES}_p = \mathcal{E}^h$ where $h(t) = \min\{t/(1-p), 1\}$ for $t \in [0, 1]$. Moreover, ES admits a dual representation (Föllmer and Schied (2016, Theorem 4.52)):

$$\text{ES}_p(X) = \sup \left\{ \mathbb{E}^P[X] : P \ll \mathbb{P}, \frac{dP}{d\mathbb{P}} \leq \frac{1}{1-p} \right\} \quad \text{for all } X \in \mathcal{X}, \quad (5)$$

where $P \ll \mathbb{P}$ means that P is absolutely continuous with respect to \mathbb{P} .

Next, we describe three different motivations for considering distorted expectations in optimal transport from decision theory, robust optimization, and risk management, respectively.

1. **Behavioral decision making.** Distorted expectations are most commonly used in decision theory, where decision makers maximize distorted expected utility of their random payoffs. The use of distorted expectations was developed by Quiggin (1982), Yaari (1987), Schmeidler (1989) and Tversky and Kahneman (1992) to address paradoxes that the classic expected utility theory (EUT) fails to accommodate. By now, theories based on distorted expectation, in particular, the rank-dependent utility theory (Quiggin (1982); Yaari (1987)) and the cumulative prospect theory (Tversky and Kahneman (1992)), are the standard alternatives to EUT in decision making under risk; see Wakker (2010) for a general treatment. The portfolio selection problem with cumulative prospect theory is studied by He and Zhou (2011).

To give a concrete example, consider a decision maker whose preference is modelled by rank-dependent utility theory; that is, her welfare of a random payoff Z is evaluated by $\mathcal{E}^h[u(Z)]$, where \mathcal{E}^h is a distorted expectation and u is a utility function. Suppose that the decision maker can choose assignment between two known distributions μ and ν , for instance, pairing suppliers $X \stackrel{d}{\sim} \mu$ and consumers $Y \stackrel{d}{\sim} \nu$ of a certain commodity. The welfare of the decision maker for the assignment is $\mathcal{E}^h[u(f(X, Y))]$, where $f(x, y)$ represents the profit of matching supplier x with consumer y . The use of the rank-dependent utility in this context means that the decision maker puts different weights on large and small profits (for instance, she may be more concerned about pairs with small profit, for a fairness consideration). Therefore, the optimization problem of the decision maker is to maximize $\mathcal{E}^h[u(f(X, Y))]$ subject to $X \stackrel{d}{\sim} \mu$ and $Y \stackrel{d}{\sim} \nu$, which is a distorted transport problem.

2. **Robust optimization.** As a useful property of distorted expectations, any convex (resp, con-

cave) distorted expectation can be written as the supremum (resp. infimum) over a class of usual expectations under different probability measures (see Section 3 for details). Taking the convex case as an example, the distorted transport problem is equivalent to solving for the optimal transport plan under the worst-case scenario within a class of probability measures. This approach can be traced back to Wald’s minimax criterion (Wald (1950)) and was axiomatized by Gilboa and Schmeidler (1989) in decision theory. This idea also aligns well with the idea of distributionally robust optimization, studied by e.g., Goh and Sim (2010), Wiesemann et al. (2014) and Blanchet and Murthy (2019).

Consider a classic transport problem to minimize $\mathbb{E}[c(X, Y)]$ subject to $X \stackrel{d}{\sim} \mu$ and $Y \stackrel{d}{\sim} \nu$. Suppose that the decision maker is uncertain about the probability measure \mathbb{P} she used to assess the transport cost (for instance, \mathbb{P} may be estimated from data). Instead, she considers the robust objective

$$\text{to minimize } \sup_{P \in \mathcal{Q}} \mathbb{E}^P[c(X, Y)],$$

where \mathcal{Q} is a class of probability measures not far away from \mathbb{P} , called the uncertainty set. For many choices of \mathcal{Q} , the robust optimization problem can be converted into a distorted optimal transport problem. As a concrete example, consider the uncertainty set $\mathcal{Q} = \{P \ll \mathbb{P} : dP/d\mathbb{P} \leq 1/(1 - \beta)\}$ where $\beta \in [0, 1)$, that is, the set of all probability measures with a likelihood ratio with respect to \mathbb{P} bounded by $1/(1 - \beta)$, with $\beta = 0$ corresponding to $\mathcal{Q} = \{\mathbb{P}\}$, the case of no uncertainty. By (5),

$$\sup_{P \in \mathcal{Q}} \mathbb{E}^P[Z] = \text{ES}_\beta(Z) \text{ for all random variables } Z,$$

and hence, we arrive at a distorted transport problem. In general, any $\mathcal{E} : Z \mapsto \sup_{P \in \mathcal{Q}} \mathbb{E}^P[Z]$ is a sub-linear expectation, provided that it satisfies law invariance.

3. **Robust risk aggregation.** In financial risk management, distorted expectations are known as distortion risk measures; see McNeil et al. (2015). A robust risk aggregation problem typically concerns

$$\sup\{\mathcal{E}^h(X_1 + \dots + X_n) : X_1 \stackrel{d}{\sim} \mu_1, \dots, X_n \stackrel{d}{\sim} \mu_n\},$$

where μ_1, \dots, μ_n are n probability measures on \mathbb{R} . For robust risk aggregation problems in risk management, see the survey of Embrechts et al. (2014) and the monograph Rüschemdorf (2013). A robust risk aggregation problem for \mathcal{E}^h corresponds to a multi-marginal distorted

transport problem, and we focus on $n = 2$ in this paper. Robust risk aggregation problems are motivated by the fact that the dependence structure of multiple risks is typically difficult to accurately specify. This type of uncertainty is referred to *dependence uncertainty*, as studied by Embrechts et al. (2013, 2015) and Blanchet et al. (2024). Distorted expectations \mathcal{E}^h arise in this context because the standard regulatory risk measures used in banking and insurance, such as Basel IV and Solvency II, are the Value-at-Risk (VaR) and the ES, both of which are special cases of distorted expectations. Moreover, distortion risk measures are popular in the literature of risk management. In the special case that h is concave, robust risk aggregation can be reformulated as a multi-marginal classic optimal transport problem, as studied by Emaji et al. (2024).

In the above settings, distorted expectations \mathcal{E}^h appear for three different reasons: as preference models in the first setting, as a result of statistical uncertainty in the second, and as externally specified regulatory risk measures in the third. Moreover, although both the second and the third settings deal with uncertainty, their mathematical structures and interpretations are quite different. In robust optimization, uncertainty is reflected by the probabilities representing \mathcal{E}^h , and there is no uncertainty associated with the transport plan; in robust risk aggregation, uncertainty is reflected by the transport plan representing dependence among risks, and there is no uncertainty associated with the risk measure \mathcal{E}^h .

3 The framework and some preliminaries

We describe our main setting in this section. Throughout, equalities and inequalities between random variables hold in the almost sure sense. For real numbers u and v , we write $u \wedge v = \min\{u, v\}$, $u \vee v = \max\{u, v\}$ and $u_+ = \max\{u, 0\}$.

3.1 Non-linear and distorted optimal transport

We first briefly explain the classic Monge-Kantorovich optimal transport problem, and then we focus on non-linear and distorted optimal transport on the real line.

Denote by $\Pi(\mathfrak{X})$ the set of all Borel probability measures on a Polish space \mathfrak{X} equipped with the Borel σ -field $\mathcal{B}(\mathfrak{X})$. Consider Polish spaces $\mathfrak{X}, \mathfrak{Y}$, and probability measures $\mu \in \Pi(\mathfrak{X})$ and $\nu \in \Pi(\mathfrak{Y})$. We will always equip $\mathfrak{X} \times \mathfrak{Y}$ with the product σ -field. Given a function $c : \mathfrak{X} \times \mathfrak{Y} \rightarrow \mathbb{R}$, which models the transportation cost, the classic optimal transport problem raised by Monge asks for the minimum of the expected transport cost $\int_{\mathfrak{X}} c(x, T(x))\mu(dx)$ for T over the set $\mathcal{T}(\mu, \nu)$ of transport

maps from μ to ν , i.e., measurable functions $T : \mathfrak{X} \rightarrow \mathfrak{Y}$ such that $\mu \circ T^{-1} = \nu$. The Kantorovich problem is a relaxation of Monge's problem, which solves for the minimum of

$$\int_{\mathfrak{X} \times \mathfrak{Y}} c(x, y) \pi(\mathrm{d}x, \mathrm{d}y) \quad (6)$$

for π over the set $\Pi(\mu, \nu)$ that is the set of transport plans from μ to ν , i.e., distributions on $\mathfrak{X} \times \mathfrak{Y}$ with marginal μ on \mathfrak{X} and marginal ν on \mathfrak{Y} . Expressing the cost using random variables, the problem of (6) can be written as

$$\text{to minimize } \mathbb{E}[c(X, Y)] \text{ subject to } X \stackrel{\mathrm{d}}{\sim} \mu \text{ and } Y \stackrel{\mathrm{d}}{\sim} \nu. \quad (7)$$

Replacing \mathbb{E} in (7) by a non-linear expectation \mathcal{E} or \mathcal{E}^h , we formulate the problem of *non-linear optimal transport* mentioned in the introduction:

$$\text{to minimize } \mathcal{E}[c(X, Y)] \text{ subject to } X \stackrel{\mathrm{d}}{\sim} \mu \text{ and } Y \stackrel{\mathrm{d}}{\sim} \nu, \quad (8)$$

as well as the special case of *distorted optimal transport* :

$$\text{to minimize } \mathcal{E}^h[c(X, Y)] \text{ subject to } X \stackrel{\mathrm{d}}{\sim} \mu \text{ and } Y \stackrel{\mathrm{d}}{\sim} \nu. \quad (9)$$

The problems (8) and (9) are technically quite different from (7). In case c is a linear combination of separable terms, that is, $c(x + y) = f(x) + g(y)$ for some functions $f : \mathfrak{X} \rightarrow \mathbb{R}$ and $g : \mathfrak{Y} \rightarrow \mathbb{R}$, the problem (7) is trivial since $\mathbb{E}[c(X, Y)] = \mathbb{E}[f(X)] + \mathbb{E}[g(Y)]$, which does not depend on the dependence structure of (X, Y) and thus any transport plan is optimal. However, even with a linear cost c , the problems (8) and (9) are highly nontrivial because the non-linear expectation is not an additive operator. In the case of additive cost, we may, without loss of generality, replace $f(X)$ and $g(Y)$ by X and Y , respectively, and consider the problem

$$\text{to minimize } \mathcal{E}[X + Y] \text{ subject to } X \stackrel{\mathrm{d}}{\sim} \mu \text{ and } Y \stackrel{\mathrm{d}}{\sim} \nu, \quad (10)$$

where μ and ν are probability measures on \mathbb{R} . From now on, we will assume $\mathfrak{X} = \mathfrak{Y} = \mathbb{R}$.

In some contexts, such as robust risk aggregation (see Section 7), the maximization of $\mathcal{E}[c(X, Y)]$ is relevant, which represents the worst-case risk under unknown dependence structure. The maximization problem of $\mathcal{E}[c(X, Y)]$ can be converted to a minimization problem by taking the cost $-c$ and a different non-linear expectation (see Lemma 3 below for the explicit case of \mathcal{E}^h). Therefore,

we can, without loss of generality, only study the minimization problems (8)-(10).

3.2 Copulas

To address problem (7) in the classic setting and our problems (8)-(10), a convenient tool is the concept of copulas. For general references on copulas in risk management, see Nelsen (2006) and McNeil et al. (2015).

Let $\mathcal{I} = [0, 1]$, and thus \mathcal{I}^2 is the unit square $\mathcal{I} \times \mathcal{I}$. A *2-copula* (or briefly, a copula) is a distribution function on \mathcal{I}^2 with $U[0, 1]$ marginals. Equivalently, $C : \mathcal{I}^2 \rightarrow \mathcal{I}$ is a function with the following properties: For every $u, v \in \mathcal{I}$, $C(u, 0) = 0 = C(0, v)$, $C(u, 1) = u$ and $C(1, v) = v$, and for every $u_1, u_2, v_1, v_2 \in \mathcal{I}$ with $u_1 \leq u_2$ and $v_1 \leq v_2$, $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$. Throughout the paper, we denote by \mathcal{C} the set of all 2-copulas.

Sklar's theorem states that for any joint distribution function H with margins F and G , there exists a copula C such that

$$H(x, y) = C(F(x), G(y)) \text{ for all } x, y \in \mathbb{R}. \quad (11)$$

If F and G are continuous, then C is unique. The copula C in (11) is called a copula of $(X, Y) \stackrel{d}{\sim} H$, and it is denoted by $C_{X,Y}$ if it is unique.

Transport plans, including transport maps, can be represented by copulas; for this reason, transport plans are also called couplings in the literature. The optimization problem (8) can be equivalently formulated via copulas

$$\text{to minimize } \mathcal{E}[c(F_\mu^{-1}(U), F_\nu^{-1}(V))] \text{ subject to } C \in \mathcal{C} \text{ where } (U, V) \stackrel{d}{\sim} C, \quad (12)$$

where F_μ is the distribution function of μ with the corresponding left-quantile function F_μ^{-1} , that is, $F_\mu(x) = \mu((-\infty, x])$ for $x \in \mathbb{R}$ and $F_\mu^{-1}(t) = Q_t(X)$ for $t \in (0, 1]$ with $X \stackrel{d}{\sim} \mu$.

Two important copulas that we will frequently encounter are the comonotonic copula (or the Fréchet-Hoeffding upper bound)

$$C^+(u, v) = u \wedge v \text{ for } u, v \in [0, 1],$$

and the counter-monotonic copula (or the Fréchet-Hoeffding lower bound)

$$C^-(u, v) = (u + v - 1)_+ \text{ for } u, v \in [0, 1].$$

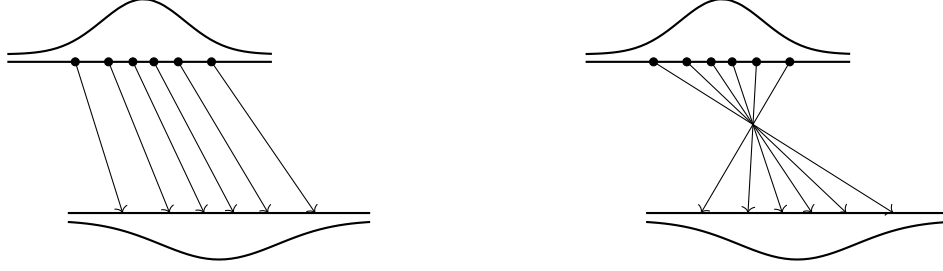


Figure 1: The transport maps between two normal distributions described by the copulas C^+ (left) and C^- (right)

Two random variables X and Y are *comonotonic* if there exist increasing functions f and g such that $X = f(X + Y)$ and $Y = g(X + Y)$; they are *counter-monotonic* if $X, -Y$ are comonotonic. It is well known that two random variables are comonotonic if and only if one of their copulas is C^+ , and two random variables are counter-monotonic if and only if one of their copulas is C^- . Moreover, $C^- \leq C \leq C^+$ for all $C \in \mathcal{C}$. This is known as the Fréchet-Hoeffding bounds; see [Nelsen \(2006, Theorem 2.2.3\)](#). The transport maps represented by C^- and C^+ are presented in [Figure 1](#).

Another class of copulas that we will use is the class of ordinal sum copulas ([Nelsen \(2006, Section 3.2.2\)](#)), indexed by $p \in (0, 1)$, defined by

$$\mathcal{C}_p = \{C \in \mathcal{C} : C(1 - p, 1 - p) = 1 - p\}. \quad (13)$$

One can verify $\bigcap_{p \in (0, 1)} \mathcal{C}_p = \{C^+\}$ (e.g., [Wang and Zitikis \(2021, Theorem 4\)](#)) and $C^- \notin \bigcup_{p \in (0, 1)} \mathcal{C}_p$.

The copulas C^+ and C^- play special roles in classic transport problems [\(7\)](#), equivalent to [\(9\)](#) with h being the identity. A function $c : \mathbb{R}^d \rightarrow \mathbb{R}$ is *submodular* if

$$c(\mathbf{u} \wedge \mathbf{v}) + c(\mathbf{u} \vee \mathbf{v}) \leq c(\mathbf{u}) + c(\mathbf{v}), \text{ for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^d,$$

where $\mathbf{u} \wedge \mathbf{v}$ is the component-wise minimum of \mathbf{u} and \mathbf{v} , and $\mathbf{u} \vee \mathbf{v}$ is the component-wise maximum of \mathbf{u} and \mathbf{v} . A function c is *supermodular* if $-c$ is submodular. A simple example of a submodular cost function is $c(x, y) = |x - y|^p$ for $p \geq 1$. For the classic transport problem, it is well known (see e.g., [Puccetti and Wang \(2015\)](#)) that if c is submodular, then C^+ minimizes [\(7\)](#), and if c is a supermodular function, then C^- minimizes [\(7\)](#). The class \mathcal{C}_p solves a special case of [\(9\)](#), which we will see below.

With given \mathcal{E} and c , optimality of a copula C for the problem [\(12\)](#) may depend on the choice of μ and ν . However, in some interesting cases, the choice of μ and ν is irrelevant for optimality. For instance, as we see above, for the identity h and submodular c , C^+ is optimal regardless of μ

and ν . This leads to our definition of universal optimality. In what follows, $X \stackrel{d}{=} X'$ means that X and X' have the same law.

Definition 1 (Universal optimality). Let \mathcal{E} be a non-linear expectation and $c : \mathbb{R}^2 \rightarrow \mathbb{R}$. The copula C is *universally optimal* for (c, \mathcal{E}) if for all $X, Y \in \mathcal{X}$ with copula C ,

$$\mathcal{E}[c(X, Y)] \leq \mathcal{E}[c(X', Y')] \quad \text{for all } X', Y' \in \mathcal{X} \text{ satisfying } X \stackrel{d}{=} X' \text{ and } Y \stackrel{d}{=} Y'. \quad (14)$$

If (14) holds with $\mathcal{E} = \mathcal{E}^h$ for $h \in \mathcal{H}$, we say that the copula C is universally optimal for (c, h) .

We sometimes drop “universally” in the text (but never in results) when the context is clear. We also say that a random vector (X, Y) is optimal if (14) holds. In the special case that $c(x, y) = x + y$ which we focus on in later sections, we omit c and say that the copula C is universally optimal for \mathcal{E} (or for h). As in the classic optimal transport literature, the term “optimality” refers to minimization. We will replace “optimal” with “maximal” if “ \leq ” in (14) is replaced with “ \geq ”. Maximality can be useful when computing the worst-case dependence structure in risk aggregation problems, treated in Section 7.

Universal optimality allows for the optimization problem (12) to be computed independent of the marginal distributions μ and ν . This is arguably a quite strong condition, and universal optimizers may not exist. However, universal optimizers do exist in some special and important cases. For instance, as we see above, comonotonicity is universally optimal for $\mathcal{E} = \mathbb{E}$ and c being submodular. For an example outside classic transport problems, take $\mathcal{E}^h(X) = -\text{ES}_p(-X)$ for $X \in \mathcal{X}$ and $p \in (0, 1)$; equivalently, $\tilde{\mathcal{E}}^h = \text{ES}_p$ (see Lemma 3). The recent result of Wang and Zitikis (2021, Theorem 5) on risk aggregation of ES implies that C is universally optimal for h if and only if $C \in \mathcal{C}_{1-p}$; in that result, the worst-case risk aggregation problem is studied by maximizing ES_p , which corresponds to minimizing \mathcal{E}^h .

3.3 Properties of distorted expectations

We present some well-known properties of distorted expectations in the following three lemmas, which will be useful in our analysis.

Lemma 1. *For $h \in \mathcal{H}$, \mathcal{E}^h is additive for comonotonic random variables.*

Lemma 1 follows from the fact that all Choquet integrals, including \mathcal{E}^h , are additive for comonotonic random variables (e.g., Schmeidler (1986)).

For random variables X and Y , we say that X is dominated by Y in *convex order*, denoted by $X \leq_{\text{cx}} Y$, if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for any convex function f provided that the expectations exist. A

function $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is *increasing in convex order* if $\rho(X) \leq \rho(Y)$ whenever $X \leq_{\text{cx}} Y$. In the above definition, if “convex function” is replaced with “increasing convex function”, “concave function”, or “increasing concave function”, then the three resulting orders are called increasing convex order, concave order, and increasing concave order, respectively.

Lemma 2. *For $h \in \mathcal{H}$, the following are equivalent:*

- (i) \mathcal{E}^h is subadditive (resp. superadditive);
- (ii) \mathcal{E}^h is increasing in convex order (resp. concave order);
- (iii) \mathcal{E}^h is increasing in increasing convex order (resp. increasing concave order);
- (iv) h is concave (resp. convex).

A more general version of Lemma 2 can be found in Theorem 3 of Wang et al. (2020), where several other properties are considered and h is not assumed monotone. Point (iii) is not included in Theorem 3 of Wang et al. (2020) but its equivalence to (ii) with increasing h can be easily shown using standard arguments in stochastic order. As an example, for $\mathcal{E}^h = \text{ES}_p$ where $p \in (0, 1)$, the function h is concave, and hence ES_p is subadditive; this is a key property that makes ES_p a coherent risk measure in the sense of Artzner et al. (1999).

The next lemma on sign changes of distorted expectations will also be useful. For $h \in \mathcal{H}$, let $\tilde{h} \in \mathcal{H}$ be given by $\tilde{h}(t) = 1 - h(1 - t)$ for $t \in [0, 1]$, which is called the *dual* of h . If h is continuous, then (3) can be rewritten as

$$\mathcal{E}^h[X] = \int_0^1 Q_{1-t}(X) dh(t) = \int_0^1 Q_t(X) d\tilde{h}(t).$$

Lemma 3. *For $h \in \mathcal{H}$ and $X \in \mathcal{X}$, $\mathcal{E}^h(X) = -\mathcal{E}^{\tilde{h}}(-X)$.*

Lemma 3 can be found in Dhaene et al. (2012) and Wang et al. (2020, Lemma 2), which also holds for non-monotone h in the latter reference. This lemma will be useful by allowing us to convert between a minimization problem and a corresponding maximization problem.

Finally, a distorted expectation may be written as the worst-case or best-case value of expectation over a collection of probability measures. In particular, h is concave if and only if there exists a collection \mathcal{Q} of probability measures on (Ω, \mathcal{A}) containing \mathbb{P} such that

$$\mathcal{E}^h[X] = \sup_{P \in \mathcal{Q}} \mathbb{E}^P[X] \quad \text{for all } X \in \mathcal{X}, \tag{15}$$

and h is convex if and only if there exists a collection \mathcal{Q} of probability measures on (Ω, \mathcal{A}) containing \mathbb{P} such that

$$\mathcal{E}^h[X] = \inf_{P \in \mathcal{Q}} \mathbb{E}^P[X] \quad \text{for all } X \in \mathcal{X}. \quad (16)$$

The representations (15) and (16) are direct consequences of the fact that \mathcal{E}^h with a concave h is a sub-linear expectation, and also a coherent risk measure (Artzner et al. (1999)). An example of such a representation is (5) in Section 2. Using (15) and (16), a special case of the distorted optimal transport (9) is to minimize the worst-case or best-case values of the expected cost over an uncertainty set, as explained in Section 2.

4 General results

In this section, we present some general results on non-linear and distorted optimal transport. In what follows, the cost function $c : \mathbb{R}^2 \rightarrow \mathbb{R}$ is monotone if it is increasing or decreasing in the component-wise sense.

Theorem 1. *The copula C^+ is universally optimal for any (c, \mathcal{E}) with super-linear \mathcal{E} and monotone submodular c . The copula C^- is universally optimal for any (c, \mathcal{E}) with sub-linear \mathcal{E} and monotone supermodular c .*

Proof. We first consider the case that $\mathcal{E} = \mathcal{E}^h$. Let h be concave and c be supermodular and monotone. Let X, Y be comonotonic. We next show that, for $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$,

$$\mathbb{E}[f \circ c(X, Y)] \geq \mathbb{E}[f \circ c(X', Y')] \quad \text{for all increasing convex } f. \quad (17)$$

That is, $c(X, Y)$ dominates $c(X', Y')$ in increasing convex order. Note that $\mathbb{E}[g(X, Y)] \geq \mathbb{E}[g(X', Y')]$ for all supermodular functions g ; see Müller and Stoyan (2002, Theorem 3.9.8). It remains to show that $f \circ c$ is supermodular. For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, let $t = c(\mathbf{u} \wedge \mathbf{v}) + c(\mathbf{u} \vee \mathbf{v}) - c(\mathbf{u}) - c(\mathbf{v})$. We have $t \geq 0$ because c is supermodular. Monotonicity of c implies either $c(\mathbf{u} \wedge \mathbf{v}) \leq c(\mathbf{u}) \leq c(\mathbf{u} \vee \mathbf{v})$ or $c(\mathbf{u} \wedge \mathbf{v}) \geq c(\mathbf{u}) \geq c(\mathbf{u} \vee \mathbf{v})$. Hence, there exists $\lambda \in [0, 1]$ such that $c(\mathbf{u}) = \lambda c(\mathbf{u} \wedge \mathbf{v}) + (1 - \lambda)c(\mathbf{u} \vee \mathbf{v})$, which implies $c(\mathbf{v}) + t = (1 - \lambda)c(\mathbf{u} \wedge \mathbf{v}) + \lambda c(\mathbf{u} \vee \mathbf{v})$. Convexity and increasing monotonicity of f yield

$$f(c(\mathbf{u} \wedge \mathbf{v})) + f(c(\mathbf{u} \vee \mathbf{v})) \geq f(c(\mathbf{u})) + f(c(\mathbf{v}) + t) \geq f(c(\mathbf{u})) + f(c(\mathbf{v})),$$

and thus $f \circ c$ is supermodular. Therefore, (17) holds. By Lemma 2, \mathcal{E}^h is increasing in increasing convex order. Hence, the comonotonic vector (X, Y) maximizes $\mathcal{E}^h[c(X, Y)]$. Using Lemma 3,

(X, Y) minimizes $\mathcal{E}^{\tilde{h}}[-c(X, Y)]$. Note that \tilde{h} is an arbitrary convex distortion function and $-c$ is an arbitrary monotone submodular cost function. Therefore, we conclude that C^+ is universally optimal for any (c, h) with c monotone and submodular and h convex.

Next, suppose that \mathcal{E} is a super-linear expectation. Using Theorem 4 of [Kusuoka \(2001\)](#),³ there exists a set $\mathcal{H}_0 \subseteq \mathcal{H}$ of convex distortion functions such that $\mathcal{E}[X] = \inf_{h \in \mathcal{H}_0} \mathcal{E}^h[X]$ for $X \in \mathbb{R}$. Since C^+ is universally optimal for (c, h) with any convex h , it is also universally optimal for (c, \mathcal{E}) .

The other statement is symmetric by using the fact that $\mathbb{E}[g(X, Y)] \geq \mathbb{E}[g(X', Y')]$ for all submodular functions g , where X, Y are counter-monotonic, $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$. \square

Different from the classic transport problem, the cost function c is assumed to be monotone in [Theorem 1](#). This excludes, in particular, the popular cost function $c(x, y) = |x - y|^p$. Unfortunately, this restriction is essential and cannot be dropped. For instance, take $c(x, y) = x + y$ which is both supermodular and submodular. With given marginal distributions, a counter-monotonic random vector (X, Y) minimizes $\text{ES}_p(X + Y)$, which has a simple interpretation that a maximally hedged portfolio has the smallest risk. This optimality is implied by the second statement in [Theorem 1](#). On the other hand, if we take $c(x, y) = x - y$ which is also both supermodular and submodular, then a *comonotonic* random vector (X, Y) minimizes $\text{ES}_p(X - Y)$, with the same interpretation as above. We can see that although c is submodular in both cases, the optimal transport plans (copulas) are completely different, due to the fact that c is not monotone in the second example. This explains why monotonicity is important in distorted optimal transport problems.

To explain why monotonicity is irrelevant to the classic transport problem, for the linear expectation, optimality of (X, Y) for the cost c is the same as that for the cost $c'(x, y) = c(x, y) + f(x) + g(y)$ with arbitrary functions f and g , because monotonicity of the cost c' can be amended by freely choosing f and g . However, this is not true for distorted transport problems, since $\mathcal{E}^h[c'(X, Y)] \neq \mathcal{E}^h[c(X, Y)] + \mathcal{E}^h[f(X)] + \mathcal{E}^h[g(Y)]$ in general, and monotonicity of the cost function cannot be adjusted freely by adding separable terms.

Next, we state a weak form of duality. For any cost function c and $\mu \in \Pi(\mathbb{R})$, let

$$S_c = \{(\phi, \psi) \mid \phi(x) + \psi(y) \leq c(x, y) \text{ for all } x, y \in \mathbb{R}\},$$

and let X^μ be a random variable with distribution μ .

³This result has a continuity assumption that is later removed by [Jouini et al. \(2006\)](#) and [Delbaen \(2012\)](#).

Theorem 2. Let \mathcal{E} be a super-linear expectation. For any cost function c , weak duality holds:

$$\inf_{X \stackrel{d}{\sim} \mu; Y \stackrel{d}{\sim} \nu} \mathcal{E}[c(X, Y)] \geq \sup_{(\phi, \psi) \in S_c} \{\mathcal{E}[\phi(X^\mu)] + \mathcal{E}[\psi(X^\nu)]\}. \quad (18)$$

If c is affine (i.e., $c(x, y) = \alpha + \beta x + \gamma y$ for some $\alpha, \beta, \gamma \in \mathbb{R}$) and $\mathcal{E} = \mathcal{E}^h$ for some $h \in \mathcal{H}$, then strong duality holds:

$$\inf_{X \stackrel{d}{\sim} \mu; Y \stackrel{d}{\sim} \nu} \mathcal{E}^h[c(X, Y)] = \sup_{(\phi, \psi) \in S_c} \left\{ \mathcal{E}^h[\phi(X^\mu)] + \mathcal{E}^h[\psi(X^\nu)] \right\}. \quad (19)$$

Proof. For any $(\phi, \psi) \in S_c$, by super-linearity, for any random variables $X \stackrel{d}{\sim} \mu$ and $Y \stackrel{d}{\sim} \nu$, we have

$$\mathcal{E}[c(X, Y)] \geq \mathcal{E}[\phi(X) + \psi(Y)] \geq \mathcal{E}[\phi(X)] + \mathcal{E}[\psi(Y)].$$

Taking an infimum over (X, Y) on the left-hand side and a supremum over $(\phi, \psi) \in S_c$ on the right-hand side yields (18). If $\mathcal{E} = \mathcal{E}^h$ and c is affine with $c(x, y) = \alpha + \beta x + \gamma y$, then we can take $\psi(x) = \alpha + \beta x$ and $\phi(y) = \gamma y$. If $\beta\gamma \geq 0$, then, for comonotonic X and Y , using comonotonic additivity of \mathcal{E}^h in Lemma 1, $\mathcal{E}^h[\alpha + \beta X + \gamma Y] = \mathcal{E}^h[\alpha + \beta X] + \mathcal{E}^h[\gamma Y]$. If $\beta\gamma < 0$, then, for counter-comonotonic X and Y , using the same property, $\mathcal{E}^h[\alpha + \beta X + \gamma Y] = \mathcal{E}^h[\alpha + \beta X] + \mathcal{E}^h[\gamma Y]$. In either case, (18) holds as an equality, and thus strong duality (19) holds. \square

The strong duality in Theorem 2 cannot be expected with more generality. On the right-hand side of (18), the non-linearity caused by \mathcal{E} is only on the marginal level, whereas it is on the level of joint distribution for the primal formulation. Strong duality in the form of (19) heavily relies on both comonotonic additivity of \mathcal{E}^h and affinity of c . The strong duality (19) in Theorem 2 is not very useful, since for an affine cost $c(x, y) = \alpha + \beta x + \gamma y$, the corresponding optimal transport can be explicitly obtained as either comonotonicity or counter-monotonicity, depending on the sign of β and γ . Even weak duality may fail if h is not convex, as shown in the next example.

Example 1. Fix $p \in (0, 1)$ and let $h(t) = \min\{t/(1-p), 1\}$ for $t \in [0, 1]$; that is, $\mathcal{E}^h = \text{ES}_p$. Note that h is concave. For $X \stackrel{d}{\sim} \mu = \nu = \text{U}[0, 1]$, and $Y = 1 - X$, we have

$$\text{ES}_p(X + Y) = 1 < 1 + p = \text{ES}_p(X) + \text{ES}_p(Y).$$

Therefore, with $c(x, y) = x + y$, $\phi(x) = x$, $\psi(y) = y$, we can see that (18) does not hold.

Generally, the computation of

$$\inf_{X \stackrel{d}{\sim} \mu; Y \stackrel{d}{\sim} \nu} \mathcal{E}^h[c(X, Y)]$$

for non-convex and non-concave h is a challenging task, and this holds true even for the linear cost $c(x, y) = x + y$. Wang et al. (2019) obtained the following simple lower bound

$$\inf_{X \stackrel{d}{\sim} \mu; Y \stackrel{d}{\sim} \nu} \mathcal{E}^h[X + Y] \geq \mathcal{E}^{h^*}[X^\mu] + \mathcal{E}^{h^*}[X^\nu], \quad (20)$$

where $h^* := \sup\{g \leq h \mid g \text{ is convex}\}$ is the convex envelope of h (see Figure 6 below), with equality holding if the marginal distributions μ and ν satisfy the so-called conditional joint mixability; see Theorem 4.3 of Wang et al. (2019). The latter property is nontrivial, and strict inequality in (20) often holds; this may explain the difficulty to establish strong duality.

From next section on, we will focus on distorted expectations and a linear cost.

5 Linear cost with a strictly convex or concave distortion

From now on, our main objective is to study universal optimality with linear cost $c(x, y) = x + y$. This setting also includes the more general case of separable cost, $c(x, y) = f(x) + g(y)$ for arbitrary measurable functions f and g because the latter can be converted into the simple linear case by marginal transformations, that is, replacing μ and ν with $\mu \circ f^{-1}$ and $\nu \circ g^{-1}$.

Recall that the copula C is universally optimal for h (omitting the cost function c) if for any $X, Y \in \mathcal{X}$ with copula C ,

$$\mathcal{E}^h[X + Y] \leq \mathcal{E}^h[X' + Y'] \quad \text{for all } X', Y' \in \mathcal{X} \text{ satisfying } X \stackrel{d}{=} X' \text{ and } Y \stackrel{d}{=} Y'.$$

The case that h is neither convex nor concave is the most mathematically challenging, but in this section we first completely understand the case with convexity or concavity.

Below we present a necessary condition for a copula to be universally optimal, which will be useful in proving several other results in the paper. For fixed $h \in \mathcal{H}$ and $u, v \in [0, 1]$, define the function $K_{u,v}^h : [C^-(u, v), C^+(u, v)] \rightarrow [0, 1]$ by

$$K_{u,v}^h(t) = h(1 - t) + h(1 - u - v + t). \quad (21)$$

Lemma 4. *Suppose that $C \in \mathcal{C}$ is universally optimal for $h \in \mathcal{H}$. We have $K_{u,v}^h(C(u, v)) \leq K_{u,v}^h(t)$ for all $u, v \in [0, 1]$ and $t \in [C^-(u, v), C^+(u, v)]$.*

Proof. Given $u, v \in [0, 1]$, for any $t \in [C^-(u, v), C^+(u, v)]$, there exists $\lambda \in [0, 1]$ such that

$$t = \lambda C^+(u, v) + (1 - \lambda)C^-(u, v).$$

Define by $C_0 = \lambda C^+ + (1 - \lambda)C^-$, which satisfies $C_0(u, v) = t$, and let $s = C(u, v)$.

Take μ as a Bernoulli distribution with $\mu(\{0\}) = u$ and ν as a Bernoulli distribution with $\nu(\{0\}) = v$. Let $X \sim \mu$ and $Y \sim \nu$ be such that C_0 is a copula of (X, Y) . The joint distribution of (X, Y) is given by Sklar's theorem; see (11). Similarly, let $X^* \sim \mu$ and $Y^* \sim \nu$ be such that C is a copula of (X^*, Y^*) . Write $Z = X + Y$ and $Z^* = X^* + Y^*$. It is straightforward to compute

$$\begin{aligned} \mathbb{P}(Z = 0) &= C_0(u, v) = t, \\ \mathbb{P}(Z = 1) &= u + v - 2C_0(u, v) = u + v - 2t, \\ \mathbb{P}(Z = 2) &= 1 - u - v + C_0(u, v) = 1 - u - v + t. \end{aligned}$$

Consequently, we have

$$\mathcal{E}^h[Z] = \int_0^\infty h \circ \mathbb{P}(Z > z) dz = h(1 - t) + h(1 - u - v + t) = K_{u,v}^h(t).$$

Similarly, $\mathcal{E}^h[Z^*] = h(1 - s) + h(1 - u - v + s) = K_{u,v}^h(s)$. By the optimality of C , we obtain $\mathcal{E}^h[Z^*] \leq \mathcal{E}^h[Z]$, and hence $K_{u,v}^h(s) \leq K_{u,v}^h(t)$ for all $t \in [C^-(u, v), C^+(u, v)]$. \square

Lemma 4 says that the function $K_{u,v}^h(t)$ attains its minimum on the interval $[C^-(u, v), C^+(u, v)]$ at the point $C(u, v)$ for any C that is optimal.

Remark 1. The proof of Lemma 4 only involves mixtures of Bernoulli distributions μ and ν . Hence, the conclusion on $K_{u,v}^h$ in Lemma 4 holds true if optimality of C is required only for Bernoulli marginal distributions, instead of requiring it for all marginal distributions.

The next result characterizes universal optimality in case h is strictly concave or strictly convex. Note that Theorem 1 does not address the issue of uniqueness.

Theorem 3. *For a strictly convex $h \in \mathcal{H}$, C^+ is the unique universally optimal copula; for a strictly concave $h \in \mathcal{H}$, C^- is the unique universally optimal copula.*

Proof. Both optimality statements follow from Theorem 1 by taking the linear cost $c(x, y) = x + y$. Below, we show the uniqueness statements.

First, let h be strictly convex. For $r \in (0, 1]$ and $t \in [(2r - 1)_+, r]$, note that $0 \leq 1 - 2r + t \leq 1 - t \leq 1$. Strict convexity of h implies that the function $K_{r,r}^h$, that is, $K_{r,r}^h(t) = h(1 - t) + h(1 - 2r + t)$,

is strictly decreasing on $[(2r-1)_+, r]$. Hence, $K_{r,r}^h(r) < K_{r,r}^h(t)$ for any $t \in [(2r-1)_+, r)$, and $K_{r,r}^h(t)$ has the unique minimum point $t = r$ on the interval $[(2r-1)_+, r]$. By Lemma 4, the optimality of C yields $C(r, r) = r$ for all $r \in (0, 1]$, and this implies $C \geq C^+$. Since $C \leq C^+$ also holds, we conclude that $C = C^+$.

Next, let h be strictly concave. Take $u, v \in [0, 1]$ with $u + v \leq 1$, and write $r = (u + v)/2$. For any $t \in [0, r]$, we have $0 \leq 1 - 2r + t \leq 1 - t \leq 1$. Strict concavity of h implies that the function $K_{u,v}^h$, that is, $K_{u,v}^h(t) = h(1-t) + h(1-u-v+t)$, is strictly increasing on $[0, r]$. Hence, $K_{u,v}^h(0) < K_{u,v}^h(t)$ for any $t \in (0, r]$, and $K_{r,r}^h(t)$ has the unique minimum point $t = 0$ on the interval $[0, r]$. By Lemma 4, the optimality of C yields $C(u, v) = 0$ for all $u, v \in [0, 1]$ with $u + v \leq 1$. This together with the 1-Lipschitz continuity of C (e.g., Theorem 2.2.4 of Nelsen (2006)) implies $C \leq C^-$. Since $C \geq C^-$ also holds, we conclude that $C = C^-$. \square

Theorem 3 yields that C^+ and C^- are the unique universally optimal copulas for the two classes of distorted optimal transport problems with concave or convex h . In the next section, we study the more intriguing case that h is not convex nor concave.

Remark 2. In a different context, a weaker version of the statement on C^+ in Theorem 3 can be found in Cheung (2010, Theorem 7), where h is assumed continuously differentiable with $h'(0) < \infty$.

Theorem 1 implies that the copula C^- is universally optimal for a convex h and the copula C^+ is universally optimal for a concave h . Below, we present a corresponding version of this result for maximizers instead of minimizers. This fact will be used later in the proof of another result.

Proposition 1. *The copula C^+ is universally maximal for a concave h and the copula C^- is universally maximal for a convex h .*

Proof. This proposition is a consequence of two facts. First, $X + Y$ has the largest convex order when X, Y are comonotonic and $X + Y$ has the smallest convex order when X, Y are countermonotonic; see Corollary 3.28 of Rüschendorf (2013). Second, h is concave (resp. convex) if and only if \mathcal{E}^h is increasing (resp. decreasing) in convex order by Lemma 2. These two facts yield the desired statement. \square

A similar uniqueness statement to Theorem 3 can be made for the maximizers, and it is omitted.

6 Linear cost with an inverse S-shaped distortion

In this section we focus on inverse S-shaped distortion functions for their plausibility in decision theory and their intriguing mathematical properties regarding distorted optimal transport. We

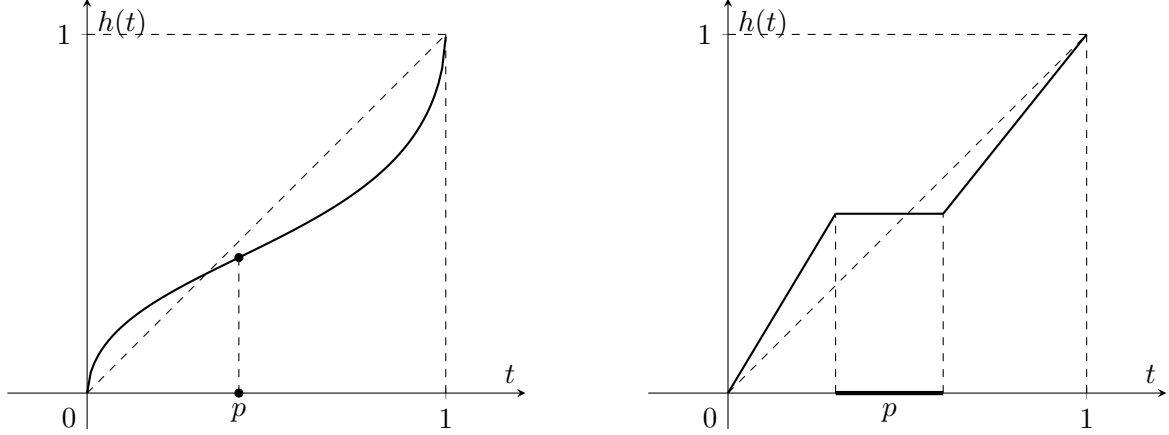


Figure 2: ISS distortion functions that are strictly ISS (left, with a unique inflection point) and not strictly ISS (right, with many inflection points)

continue to consider the linear cost $c(x, y) = x + y$.

6.1 Inverse S-shaped distortion function and two classes of copulas

First, we formally define this class of distortion functions.

Definition 2 (ISS distortion function). A distortion function $h \in \mathcal{H}$ is *inverse S-shaped* (ISS) if it is continuous and there exists a point $p \in (0, 1)$ (called an *inflection point*) such that h is concave on $(0, p)$ and convex on $(p, 1)$. It is *strictly ISS* if it is strictly concave on $(0, p)$ and strictly convex on $(p, 1)$.

ISS distortion functions are also called *cavev* (meaning “concave-convex”); see Wakker (2001). Figure 2 presents two examples of ISS distortion functions. The ISS distortion functions reflect on the empirical observation that small probabilities of both very good and very bad events are both overweighed in decision making (Tversky and Kahneman (1992); Gonzalez and Wu (1999); Abdellaoui (2000)). Clearly, a strictly ISS distortion function is strictly increasing and has a unique inflection point.

Next, we introduce two special classes of copulas that are candidates for solving optimal transport in the context of ISS distortion functions. For any $p \in (0, 1)$, define

$$C_p^\pm(u, v) = \begin{cases} u \wedge v & u \wedge v \in [0, 1 - p] \\ (u + v - 1) \vee (1 - p) & \text{otherwise;} \end{cases} \quad (22)$$

$$C_p^\mp(u, v) = \begin{cases} (u + v - 1 + p)_+ & u \vee v \in [0, 1 - p] \\ u \wedge v & \text{otherwise.} \end{cases} \quad (23)$$

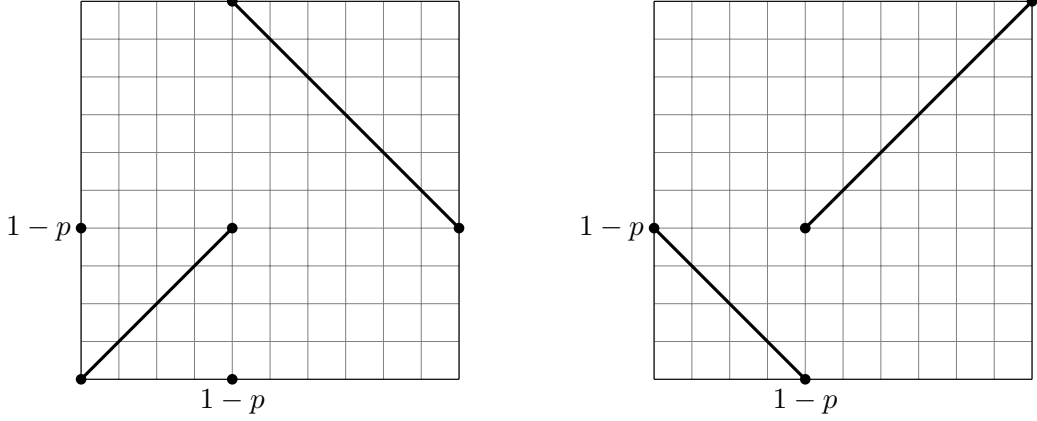


Figure 3: Supports of the copulas C_p^\pm (left) and C_p^\mp (right)

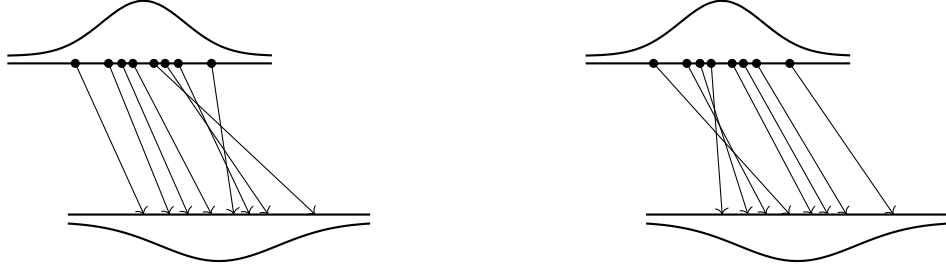


Figure 4: The transport maps between two normal distributions described by the copulas C_p^\pm (left) and C_p^\mp (right) with $p = 0.5$

Figure 3 presents the supports of C_p^\pm and C_p^\mp , which are represented by some line segments in $[0, 1]^2$.

Since $C_p^\pm(1-p, 1-p) = C_p^\mp(1-p, 1-p) = 1-p$, C_p^\pm and C_p^\mp are both in the set \mathcal{C}_p defined in (13). Indeed, C_p^\pm and C_p^\mp are ordinal sum copulas generated by C^+ and C^- in different orders; this explains our choice of notation, with “ \pm ” representing “first + then -” in the ordinal sum copula construction. The corresponding transport maps are illustrated in Figure 4. A stochastic representation of C_p^\pm , that is, a random vector $(U, V) \stackrel{d}{\sim} C_p^\pm$, is given by

$$U \stackrel{d}{\sim} U[0, 1] \text{ and } V = U\mathbf{1}_{\{U \leq 1-p\}} + (2-p-U)\mathbf{1}_{\{U \geq 1-p\}}. \quad (24)$$

A stochastic representation of C_p^\mp is given by

$$U \stackrel{d}{\sim} U[0, 1] \text{ and } V = (1-p-U)\mathbf{1}_{\{U \leq 1-p\}} + U\mathbf{1}_{\{U \geq 1-p\}}. \quad (25)$$

6.2 Optimality within a subset

To address the optimal transport problem for ISS distortion functions, we need a modification of universal optimality constrained within a subset of copulas. Let $\mathcal{C}' \subseteq \mathcal{C}$ and $h \in \mathcal{H}$. The copula $C \in \mathcal{C}'$ is *universally optimal within \mathcal{C}'* for h if for any $X, Y \in \mathcal{X}$ with copula C , we have

$$\mathcal{E}^h[X + Y] \leq \mathcal{E}^h[X' + Y']$$

for all $X', Y' \in \mathcal{X}$ with a copula in \mathcal{C}' satisfying $X \stackrel{d}{=} X'$ and $Y \stackrel{d}{=} Y'$.

The next proposition illustrates the symmetric roles of C_p^\pm and C_p^\mp in (22)-(23) within the set \mathcal{C}_p . Recall that \tilde{h} is the dual of h , and we speak of maximality, instead of optimality, for a copula maximizing the distorted cost.

Proposition 2. *Let $h \in \mathcal{H}$. The copula C_p^\pm is universally optimal (resp. within \mathcal{C}_p) for h if and only if the copula C_{1-p}^\mp is universally maximal (resp. within \mathcal{C}_{1-p}) for \tilde{h} .*

Proof. We only show the “only if” statement, and the “if” statement is symmetric. For a random vector (X', Y') , let (X, Y) have copula C_{1-p}^\mp and the same marginal distributions as (X', Y') . Note that $(-X, -Y)$ has copula C_p^\pm . Optimality of C_p^\pm implies $\mathcal{E}^h[-X - Y] \leq \mathcal{E}^h[-X' - Y']$. By using Lemma 3,

$$\mathcal{E}^{\tilde{h}}[X + Y] = -\mathcal{E}^h[-X - Y] \geq -\mathcal{E}^h[-X' - Y'] = \mathcal{E}^{\tilde{h}}[X' + Y'],$$

and this gives $\mathcal{E}^{\tilde{h}}[X + Y] \geq \mathcal{E}^{\tilde{h}}[X' + Y']$. Hence, C_{1-p}^\mp is universally maximal for \tilde{h} . Note that in the above argument, if (X', Y') has a copula in \mathcal{C}_{1-p} , then $(-X', -Y')$ has a copula in \mathcal{C}_p , and hence the optimality statements within \mathcal{C}_p and \mathcal{C}_{1-p} hold true. \square

In the next result, we show that C_p^\pm is optimal within \mathcal{C}_p for an ISS distortion function with an inflection point $p \in (0, 1)$. This, combined with Proposition 2, also shows that C_p^\mp is maximal within \mathcal{C}_p for such an ISS distortion function. The maximization problem is relevant in the context of risk aggregation treated in Section 7.

Proposition 3. *Let $h \in \mathcal{H}$ be an ISS distortion function with an inflection point $p \in (0, 1)$. The copula C_p^\pm is universally optimal within \mathcal{C}_p for h , and it is the unique optimizer if h is strictly ISS.*

Proof. Let (X, Y) be a random vector with a copula in \mathcal{C}_p . Using the quantile representation of \mathcal{E}^h in (3), we have

$$\mathcal{E}^h[X + Y] = \int_0^p Q_{1-t}(X + Y)dh(t) + \int_p^1 Q_{1-t}(X + Y)dh(t). \quad (26)$$

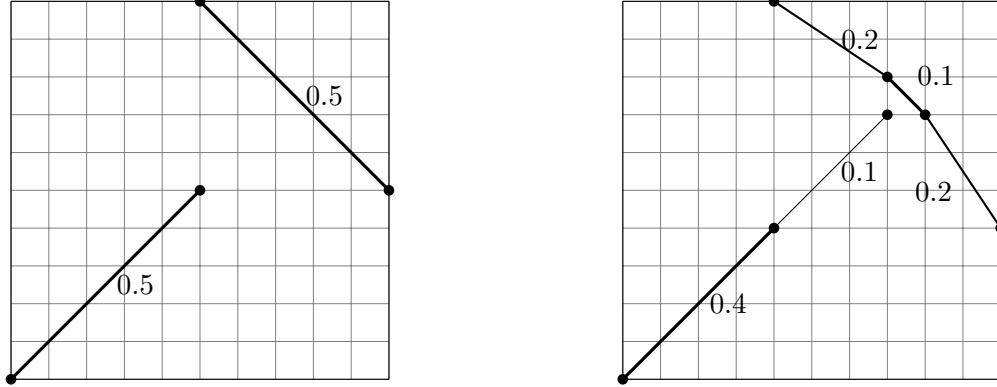


Figure 5: Supports of the copulas $C_{0.5}^{\pm}$ (left) and C^* (right), where the numbers represent the corresponding weight of each line segment

The key observation here is that for copulas in the set \mathcal{C}_p , the quantile of $X + Y$ at level $t \in [0, p)$ is solely determined by the copula of (X, Y) on $[0, p]^2$, and the quantile of $X + Y$ at level $t \in (p, 1]$ is solely determined by the copula of (X, Y) on $[p, 1]^2$. Therefore, to minimize (26), it suffices to consider the two regions separately. The term $\int_0^p Q_{1-t}(X + Y)dh(t)$ is minimized by counter-monotonicity since h is concave on $[0, p]$, and $\int_p^1 Q_{1-t}(X + Y)dh(t)$ is minimized by comonotonicity since h is convex on $[p, 1]$. Hence, C_p^{\pm} is universally optimal within \mathcal{C}_p for h . For a strictly ISS h , the uniqueness statement of the optimal copula follows from the corresponding uniqueness statements for comonotonicity and counter-monotonicity with strictly convex and strictly concave distortion functions in Theorem 3. \square

We have seen that C_p^{\pm} is universally optimal within \mathcal{C}_p for an ISS function h with inflection point p . One may naturally wonder whether C_p^{\pm} is universally optimal also within \mathcal{C} for such an h . It turns out that this is not always the case, as shown by the following example.

Example 2. We consider the following ISS distortion function

$$h(t) = \begin{cases} 2t - 2t^2 & t \in [0, 0.5] \\ 1 - 2t + 2t^2 & t \in [0.5, 1], \end{cases}$$

where the inflection point $p = 0.5$. For standard uniform random variables X, Y, X', Y' where (X, Y) has the copula $C_{0.5}^{\pm}$ and (X', Y') has the copula C^* depicted in Figure 5 (right panel), we can compute $\mathcal{E}^h[X + Y] \approx 0.9167 > \mathcal{E}^h[X' + Y'] \approx 0.9087$; we omit the detailed computation. Therefore, $C_{0.5}^{\pm}$ is not universally optimal for h .

In what follows, we search for conditions on h under which C_p^{\pm} is universally optimal.

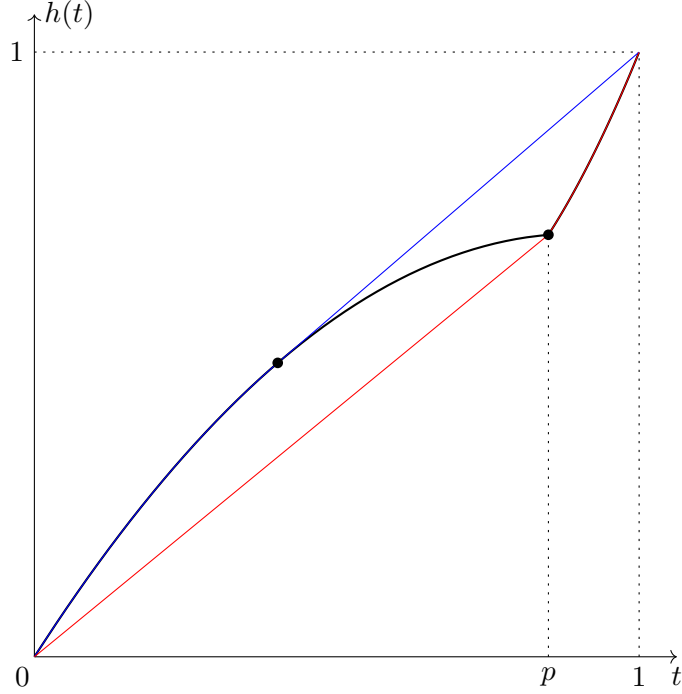


Figure 6: A strictly ISS distortion function satisfying $h'(p^+) \geq h'(0^+)$ with its convex envelope in red and concave envelope in blue

6.3 A characterization of universal optimality

Next, we present the most technically advanced result in the paper, which establishes a necessary and sufficient condition on h for the existence of a universally optimal copula, and identifies its form. In what follows, $h'(t^+)$ represents the right derivative of h at $t \in [0, 1)$, and similarly $h'(t^-)$ represents the left derivative at $t \in (0, 1]$. For an ISS distortion function h , the left and right derivatives always exist.

The key condition on h here is $h'(p^+) \geq h'(0^+)$, where p is the inflection point of h . This condition is rather restrictive. Since h is ISS, this means that the slope of h on $[0, p)$ is always smaller than or equal to the slope of h on $(p, 1]$; see Figure 6. This condition imposes a kink at p , i.e., $h'(p^+) > h'(p^-)$, since the slope is decreasing on $[0, p)$ and increasing on $(p, 1]$. Admittedly, such a condition is rarely observed for human behaviour as extreme losses and gains are typically both weighted more than the middle part of the distribution (Wakker (2010)). Nevertheless, we will see that this condition is necessary and sufficient for a universally optimal copula to exist if h is strictly ISS.

We first show universal optimality of C_p^\pm under the condition $h'(p^+) \geq h'(0^+)$, a stronger optimality than the one obtained in Proposition 3. The proof of this result relies on some separate

results on distorted optimal transport for conditional distributions. For this, let us introduce some notation. For any $Z \in \mathcal{X}$, let F_Z be its distribution function, and take a uniformly distributed random variable U_Z on $[0, 1]$ such that $F_Z^{-1}(U_Z) = Z$ almost surely; the existence of U_Z is guaranteed by, e.g., Lemma A.32 of [Föllmer and Schied \(2016\)](#). Let $Z_p^\vee = F_Z^{-1}((1-p) + pU_Z)$; the random variable Z_p^\vee is called the tail risk of Z beyond its $(1-p)$ -quantile by [Liu and Wang \(2021\)](#), and the distribution of Z_p^\vee is called the p -tail distribution of F . Similarly, let $Z_p^\wedge = F_Z^{-1}((1-p)U_Z)$, which represents the left tail risk. Here, the superscripts \vee and \wedge reflect on the large and the small regions of the distribution. With the notation above, we summarize some results needed for the proof of [Theorem 4](#) in the following two lemmas.

Lemma 5. *Let h be an ISS distortion function with an inflection point $p \in (0, 1)$ satisfying $h'(p^+) \geq h'(0^+)$. A decomposition of \mathcal{E}^h holds: for some $\theta, \theta_1, \theta_2 \geq 0$ and convex $h_1, h_2 \in \mathcal{H}$,*

$$\mathcal{E}^h[Z] = \theta \mathbb{E}[Z] - \theta_1 \mathcal{E}^{h_1}[Z_p^\vee] + \theta_2 \mathcal{E}^{h_2}[Z_p^\wedge], \text{ for all } Z \in \mathcal{X}. \quad (27)$$

Proof. Take $\theta \in [h'(0^+), h'(p^+)]$. Let $g(t) = h(t) - \theta t$ for $t \in [0, 1]$. Note that g is concave on $[0, p]$ and convex on $[p, 1]$. By definition, we have the relations $Q_t(Z_p^\vee) = Q_{(1-p)+pt}(Z)$ and $Q_t(Z_p^\wedge) = Q_{(1-p)t}(Z)$ for $t \in (0, 1)$. Using these relations, we have

$$\begin{aligned} \mathcal{E}^h[Z] &= \theta \mathbb{E}[Z] + \int_0^p Q_{1-t}(Z) dg(t) + \int_p^1 Q_{1-t}(Z) dg(t) \\ &= \theta \mathbb{E}[Z] + \int_0^p Q_{1-t/p}(Z_p^\vee) dg(t) + \int_p^1 Q_{(1-t)/(1-p)}(Z_p^\wedge) dg(t) \\ &= \theta \mathbb{E}[Z] + \int_0^1 Q_{1-s}(Z_p^\vee) dg_1(s) + \int_0^1 Q_{1-s}(Z_p^\wedge) dg_2(s), \end{aligned}$$

where $g_1(s) = g(ps) = h(ps) - \theta ps$ and $g_2(s) = g(p + (1-p)s) = h(p + (1-p)s) - \theta(p + (1-p)s)$ for $s \in (0, 1)$. Let $g_3(s) = -g_1(s)$ and $g_4(s) = g_2(s) - h(p) + \theta p$. Since $h'(t^+) \leq h'(0^+) \leq \theta$ for $t \in (0, p)$ and $h'(t^+) \geq h'(p^+) \geq \theta$ for $t \in (p, 1)$, we know that g_3 and g_4 are increasing. Moreover, it is easy to see that g_3 and g_4 are convex and $g_3(0) = g_4(0) = 0$. Therefore,

$$\int_0^1 Q_{1-s}(Z_p^\vee) dg_1(s) + \int_0^1 Q_{1-s}(Z_p^\wedge) dg_2(s) = - \int_0^1 Q_{1-s}(Z_p^\vee) dg_3(s) + \int_0^1 Q_{1-s}(Z_p^\wedge) dg_4(s).$$

Letting $\theta_1 = g_3(1)$ and $\theta_2 = g_4(1)$ yields [\(27\)](#). □

Lemma 6. *Let g be a convex distortion function and μ and ν be probability measures on \mathbb{R} . The*

random vector (X, Y) with copula C_p^\pm solves both the optimization problems

$$\text{to maximize } \mathcal{E}^g[(X + Y)_p^\vee] \text{ subject to } X \stackrel{d}{\sim} \mu \text{ and } Y \stackrel{d}{\sim} \nu, \quad (28)$$

$$\text{and to minimize } \mathcal{E}^g[(X + Y)_p^\wedge] \text{ subject to } X \stackrel{d}{\sim} \mu \text{ and } Y \stackrel{d}{\sim} \nu. \quad (29)$$

Proof. We first show the statement for the case of (28). Note that the mapping $X \mapsto \mathcal{E}^g[X_p^\vee]$ is a p -tail risk measure in the sense of Liu and Wang (2021), and we can use Theorem 3 of Liu and Wang (2021), which yields that the problem (28) has the same maximum as

$$\text{to maximize } \mathcal{E}^g[X' + Y'] \text{ subject to } X' \stackrel{d}{\sim} \mu_p \text{ and } Y' \stackrel{d}{\sim} \nu_p, \quad (30)$$

where μ_p is the p -tail distribution of μ and ν_p is the p -tail distribution of ν . By Proposition 1, (30) is maximized by counter-monotonicity of X' and Y' . By construction, (X, Y) with copula C_p^\pm and marginal distributions μ and ν satisfies $\mathcal{E}^g[(X + Y)_p^\vee] = \mathcal{E}^g[X' + Y']$. Therefore, (X, Y) with copula C_p^\pm solves (28).

The proof for the case of (29) is similar, by noting that comonotonicity minimizes \mathcal{E}^g of the sum of two random variables for a convex g by Theorem 1. \square

The next result establishes that $h'(p^+) \geq h'(0^+)$ is necessary and sufficient for the existence of a universally optimal copula when h is strictly ISS.

Theorem 4. *Let $h \in \mathcal{H}$ be an ISS distortion function with an inflection point $p \in (0, 1)$ and C be a copula. Statement (i) implies statement (ii) below:*

$$(i) \ C = C_p^\pm \text{ and } h'(p^+) \geq h'(0^+);$$

$$(ii) \ C \text{ is universally optimal for } h.$$

If h is strictly ISS, then (i) and (ii) are equivalent.

Proof. We first show the (i) \Rightarrow (ii) implication. For $X, Y \in \mathcal{X}$, by Lemma 5,

$$\mathcal{E}^h[X + Y] = \theta \mathbb{E}[X + Y] - \theta_1 \mathcal{E}^{h_1}[(X + Y)_p^\vee] + \theta_2 \mathcal{E}^{h_2}[(X + Y)_p^\wedge], \quad (31)$$

where $\theta, \theta_1, \theta_2 \geq 0$ and $h_1, h_2 \in \mathcal{H}$ are convex. By Lemma 6 and the fact that $\mathbb{E}[X + Y]$ does not depend on the copula of (X, Y) , the value of (31) is minimized by C_p^\pm under the constraints $X \sim \mu$ and $Y \sim \nu$. Therefore, (ii) holds.

Next, for a strictly ISS h , we prove the (ii) \Rightarrow (i) implication in three steps.

(a) First, we determine $C(u, v)$ in case $u \wedge v \in [0, 1 - p]$. If $0 < u = v \leq (1 - p)/2$, let $r = u$. Then, $p + 2r - 1 < 0$, and for $t \in (0, r)$, we have $1 - t > p$, $1 - 2r + t > p$ and $1 - 2r + t < 1 - t$. Strict convexity of h on the interval $(p, 1)$ implies that $K_{r,r}^h(t) = h(1 - t) + h(1 - 2r + t)$ is strictly decreasing on the interval $[0, r]$ and thus strictly decreasing on $[(2r - 1)_+, r]$. Hence, $K_{r,r}^h$ has the unique minimum point $t = r$ on the interval $[(2r - 1)_+, r]$. By Lemma 4, we have

$$C(r, r) = r \quad \text{for all } r \in (0, (1 - p)/2).$$

If $(1 - p)/2 \leq u = v \leq 1 - p$, again let $r = u$. Then, $p + 2r - 1 \geq 0$, and for $t \in (p + 2r - 1, r)$, we have $1 - t > p$ and $1 - 2r + t > p$. Strict convexity of h on the interval $(p, 1)$ implies that $K_{r,r}^h(t) = h(1 - t) + h(1 - 2r + t)$ is strictly decreasing on the interval $[p + 2r - 1, r] \subseteq [(2r - 1)_+, r]$, and hence the minimum point of $K_{r,r}^h$ on the interval $[(2r - 1)_+, r]$ cannot be in $(p + 2r - 1, r)$. Lemma 4 yields either $C(r, r) = r$ or $C(r, r) \leq p + 2r - 1$. By continuity of C , this implies

$$C(r, r) = r \quad \text{for all } r \in ((1 - p)/2, 1 - p).$$

If $0 \leq u \leq 1 - p$ and $u \leq v$, then $u = C(u, u) \leq C(u, v) \leq C^+(u, v) = u$, which yields $C(u, v) = u = u \wedge v$. Similarly, $C(u, v) = v = u \wedge v$ if $0 \leq v \leq 1 - p$ and $v \leq u$. To summarize,

$$C(u, v) = u \wedge v, \quad \text{if } u \wedge v \leq 1 - p.$$

(b) Next, we determine $C(u, v)$ in case $u, v \geq 1 - p$. If $u + v \leq 2 - p$, then $0 \leq 2 - p - (u + v)$. Moreover, $u \wedge v \geq 1 - p$ implies $p \geq 2 - p - (u + v)$ and $(u + v)/2 \leq u + v + p - 1$. Strict concavity of h on the interval $[2 - p - (u + v), p]$ implies that $K_{u,v}^h$ is strictly increasing on the interval $[1 - p, (u + v)/2]$. Thus, the unique minimum point of $K_{u,v}^h$ on the interval $[(u + v - 1)_+, u \wedge v]$ cannot be in $(1 - p, (u + v)/2]$. By Lemma 4, $C(u, v) \leq 1 - p$. Noting that $C(u, v) \geq C(1 - p, 1 - p) = 1 - p$, we have

$$C(u, v) = 1 - p, \quad \text{if } u, v \in (1 - p, 1) \text{ and } u + v \leq 2 - p. \quad (32)$$

If $u + v \geq 2 - p$, then $v \geq 2 - p - u$. Since C induces a measure, which is nonnegative on $[u, 1] \times [2 - p - u, v]$, we have $C(1, v) - C(1, 2 - p - u) - C(u, v) + C(u, 2 - p - u) \geq 0$. Using (32), this implies $v - (2 - p - u) + 1 - p = u + v - 1 \geq C(u, v)$. On the other hand, since $C \leq C^-$,

we get $C(u, v) = u + v + 1$. Therefore, putting the above two cases together, we conclude

$$C(u, v) = (u + v - 1) \vee (1 - p) \quad \text{if } u \wedge v \geq 1 - p. \quad (33)$$

(c) The points (a) and (b) above shows that the optimal copula is $C = C_p^\pm$. It remains to show $h'(p^+) \geq h'(0^+)$. Let $u = 1 - p$ and $v = 1 - \varepsilon$, where ε is a small positive number satisfying $\varepsilon < p \wedge (1 - p)$. By (33), we have $C(1 - p, 1 - \varepsilon) = 1 - p$. By Lemma 4, the function $t \mapsto K_{1-p, 1-\varepsilon}^h(t) = h(1 - t) + h(p + \varepsilon - 1 + t)$ has a minimum point $t = 1 - p$ on the interval $[1 - p - \varepsilon, 1 - p]$. Hence,

$$K_{1-p, 1-\varepsilon}^h(1 - p) = h(p) + h(\varepsilon) \leq K_{1-p, 1-\varepsilon}^h(1 - p - \varepsilon) = h(p + \varepsilon) + h(0),$$

which implies

$$\frac{h(p + \varepsilon) - h(p)}{\varepsilon} \geq \frac{h(\varepsilon) - h(0)}{\varepsilon}.$$

Letting $\varepsilon \rightarrow 0+$ yields $h'(p^+) \geq h'(0^+)$.

The above three steps complete the proof of the implication (ii) \Rightarrow (i). \square

The condition $h'(p^+) \geq h'(0^+)$ is quite strong in Theorem 4. This result has two main implications that are useful for the optimal transport theory.

- (a) If $h'(p^+) \geq h'(0^+)$ holds, then we can explicitly solve the distorted optimal transport problem regardless of the marginal distributions, just as in the case of comonotonicity in the classic transport theory, which optimizes submodular costs regardless of the marginal distributions.
- (b) If $h'(p^+) \geq h'(0^+)$ does not hold, then there is no universal solution to the distorted optimal transport problem; that is, different marginal conditions lead to different solutions, which can be difficult to obtain in general.

Point (a) can be used to find solutions for some optimal transport problems, and an example is illustrated in Section 8. Point (b) helps to clarify to which extent beyond the usual framework, we can still hope for universal optimizers. As a general message from Theorem 4, without convexity or concavity of h , such general optimizers can exist only in very restrictive cases.

An example of h that does not satisfy the condition $h'(p^+) \geq h'(0^+)$ is presented in Example 2, where we have seen that C_p^\pm is not universally optimal for h . In that example, $h'(p^+) = 0 < h'(0^+) = 2$.

If the condition $h'(p^+) \geq h'(0^+)$ is not satisfied, then C_p^\pm is not universally optimal for h . Nevertheless, C_p^\pm may still be an optimizer for some specific choices of μ and ν . For instance, if both μ and ν are uniform on $[0, 1]$, and the convex envelope of h is linear on $(0, p)$, then the copula C_p^\pm minimizes $\mathcal{E}^h[X + Y]$ for $X \stackrel{d}{\sim} \nu$ and $Y \stackrel{d}{\sim} \nu$, as shown by Wang et al. (2019, Theorem 4.3).

Remark 3. If we consider the maximization problem instead of minimization in Theorem 4, then by Proposition 2, the corresponding condition is $h'(p^-) \geq h'(1^-)$ and the corresponding maximizer is C_p^\mp . Based on this observation, the minimal or maximal cost can be directly computed from the stochastic representation in (24)-(25).

Remark 4. For a proof of $C = C_p^\pm$ in the \Rightarrow direction of Theorem 4, we rely on Lemma 4, which only involves mixtures of Bernoulli distributions (see Remark 1). Therefore, to pin down C_p^\pm as the only possible form for an optimal copula, it suffices to require optimality only for Bernoulli marginal distributions.

7 Risk measures, S-shaped distortion, and risk aggregation

As mentioned in Section 3, the framework of distorted optimal transport includes robust risk aggregation problems with distortion risk measures, and we only consider the case of aggregating two risks in this paper.

Although belonging to the same class of distorted expectations, the motivation and applications of distortion risk measures in risk management are quite different from the distorted expectations used for modeling preferences in decision theory. There are two main differences. First, the most widely used distortion risk measures are those that are convex (i.e., with a concave h , such as ES) and those that have a simple distortion function, such as the VaR and the Range-Value-at-Risk (RVaR) defined below. Second, maximum for risk measures corresponds the worst-case, where as maximum for dual utilities corresponds to the best-case; therefore, the interpretations of maximization and minimization are quite different for the two settings. In particular, in robust risk aggregation, one often focuses on the maximum of the risk measure, instead of the minimum.

The two popular classes of distortion risk measures, VaR and RVaR, are defined below. VaR can be defined either as a left-quantile or a right-quantile in the literature. We define VaR at level $p \in (0, 1)$ as the corresponding right-quantile, that is,

$$\text{VaR}_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) > p\}, \text{ for } X \in \mathcal{X},$$

and this choice is because the right-quantile admits a maximizing copula (and the left-quantile Q_p

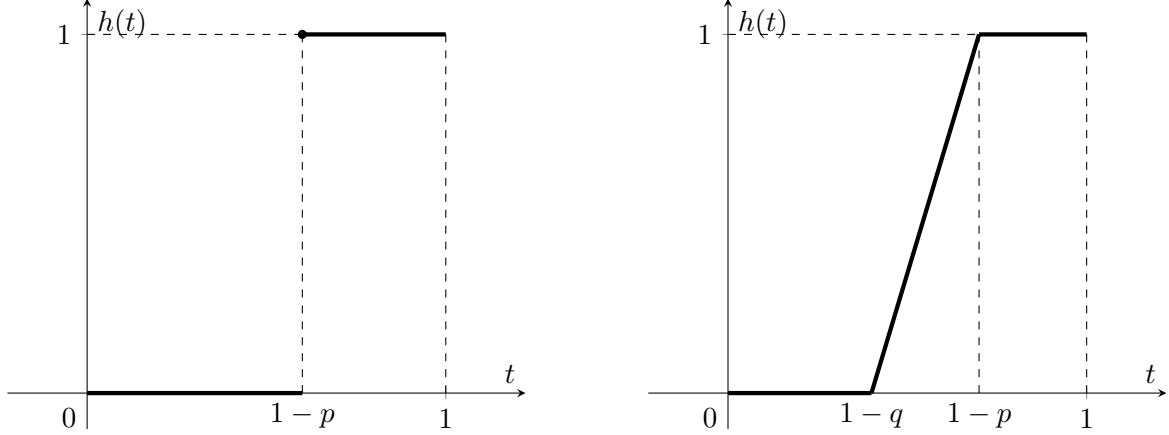


Figure 7: The distortion functions corresponding to VaR_p (left) and $\text{RVaR}_{p,q}$ (right)

admits a minimizing copula); see [Bernard et al. \(2014, Lemma 4.2\)](#). For levels $p, q \in (0, 1)$ with $p < q$, RVaR at level (p, q) is defined as

$$\text{RVaR}_{p,q}(X) = \frac{1}{q-p} \int_p^q Q_r(X) dr, \text{ for } X \in \mathcal{X}.$$

The distortion function for VaR_p is given by $h(t) = \mathbf{1}_{\{t \geq 1-p\}}$ for $t \in [0, 1]$ and the distortion function for $\text{RVaR}_{p,q}$ is given by $h(t) = (t \wedge (1-p) - (1-q))_+ / (q-p)$ for $t \in [0, 1]$. [Figure 7](#) depicts the distortion functions of VaR_p and $\text{RVaR}_{p,q}$. The VaR has been a standard risk measure in banking and insurance since the 1990s (see [Jorion \(2006\)](#)), and RVaR was introduced by [Cont et al. \(2010\)](#) as a robust risk measure. Both VaR and ES are limiting cases of RVaR , and RVaR can be expressed as the inf-convolution of VaR and ES as shown by [Embrechts et al. \(2018\)](#).

An interesting observation from [Figure 7](#) is RVaR has an S-shaped (instead of ISS) distortion function, and the distortion function of VaR is the limit of S-shaped ones. Since results for a concave h are available from [Theorems 1 and 3](#) and [Proposition 1](#), below we will focus on the S-shaped distortion functions, and present a parallel result to [Theorem 4](#).

Definition 3 (SS distortion function). A distortion function $h \in \mathcal{H}$ is *S-shaped* (SS) if it is continuous and there exists a point $p \in (0, 1)$ (called an *inflection point*) such that h is convex on $(0, p)$ and concave on $(p, 1)$. It is *strictly SS* if it is strictly convex on $(0, p)$ and strictly concave on $(p, 1)$.

The following result characterizes universal optimality for SS distortion functions.

Theorem 5. *Let $h \in \mathcal{H}$ be an SS distortion function with an inflection point $p \in (0, 1)$ and C be a copula. Statement (i) implies statement (ii) below:*

- (i) $C = C_p^\pm$ and $h'(p^+) \leq h'(0^+)$;

(ii) C is universally maximal for h .

If h is strictly SS, then (i) and (ii) are equivalent.

We omit the proof of Theorem 5 as it is symmetric to Theorem 4. Nevertheless, in case h has a bounded derivative, we can directly obtain Theorem 5 from Theorem 4, and we explain this argument below. Suppose that h satisfies $h'(p^+) \leq h'(0^+)$. For $\lambda \in (0, 1)$, define the distortion function g_λ by $g_\lambda(t) = (t - \lambda h(t))/(1 - \lambda)$. The assumption that h has a bounded derivative implies that g_λ is increasing for $\lambda > 0$ small enough. Since h is SS, we have that g_λ is an ISS distortion function, with $g'_\lambda(p^+) \geq g'_\lambda(0^+)$. By Theorem 4, the distorted transport problem for g_λ is minimized by C_p^\pm . Since $\mathcal{E}^{g_\lambda} = (\mathbb{E} - \lambda \mathcal{E}^h)/(1 - \lambda)$ and the term \mathbb{E} does not depend on the transport plan, we know that the transport problem for \mathcal{E}^h is maximized by C_p^\pm . The uniqueness statement in Theorem 5 also follows from the same observation.

Corresponding statements on minimizers can be obtained similarly as discussed in Remark 3.

For $p, q \in (0, 1)$ with $p < q$, any point in $[1 - q, 1 - p]$ is an inflection point of the distortion function h of $\text{RVaR}_{p,q}$. Taking the inflection point $r = 1 - p$ of h , the condition $h'(r^+) = 0 = h'(0^+)$ holds. Therefore, the condition in Theorem 5 holds, and a universally maximal copula is C_r^\pm ; thus, we obtain a maximizing dependence structure in the $\text{RVaR}_{p,q}$ risk aggregation problem; a related result on this problem is Theorem 3 of Liu and Wang (2021). This copula is not unique since the distortion function of RVaR is not strictly SS; indeed, by definition, $\text{RVaR}_{p,q}$ neglects the region of the distribution beyond its q -quantile and below its p -quantile, leaving some flexibility of the optimal copula.

As the limiting case of $\text{RVaR}_{p,q}$ as $p \uparrow q$, the VaR_q risk aggregation problem is also maximized by C_{1-q}^\pm . This fact was shown by Makarov (1981) and Rüschendorf (1982). As the other limiting case of $\text{RVaR}_{p,q}$ as $q \rightarrow 1$, the ES_p risk aggregation problem is maximized by C^+ , which is also implied by Theorem 1. All optimizers above are not unique; for instance, \mathcal{C}_{1-p} is the set of all maximizers for ES_p (Wang and Zitikis (2021)).

8 An application

In this section, we illustrate the results in Theorems 4 and 5 in an example of economic production. Suppose that a manufacturer produces a certain type of products in a given period of time. Each individual product requires a worker (e.g., an engineer) and a co-worker (e.g., a designer) to collaborate. The company has an equal number of workers and co-workers, and needs to match each worker $x \in \mathfrak{X}$ with a co-worker $y \in \mathfrak{Y}$, where \mathfrak{X} and \mathfrak{Y} are two arbitrary index sets. For simplicity,

suppose that the quality of a product has the additive form $f(x) + g(y)$, where $f : \mathfrak{X} \rightarrow \mathbb{R}$ is the production function of worker x and $g : \mathfrak{Y} \rightarrow \mathbb{R}$ is the production function of co-worker y . We assume measurability of f and g , and that the workers x and the co-workers y are modelled by two distributions μ and ν . The company assesses that the 20% products with the worst quality can only be sold at the minimum price, and the 30% products with the best quality are more than good enough to be sold at a maximum price. For this reason, the company's main interest is the average quality of their products between 20% and 70% quantiles.⁴

The company's optimization problem is to maximize the range-average quality, with $p = 0.2$ and $q = 0.7$,

$$\text{RVaR}_{p,q}(f(X) + g(Y)) \quad \text{subject to } X \stackrel{\text{d}}{\sim} \mu \text{ and } Y \stackrel{\text{d}}{\sim} \nu, \quad (34)$$

where $\text{RVaR}_{p,q}$ is defined in Section 7. The problem in (34) can be equivalently formulated as

$$\text{RVaR}_{p,q}(Z + W) \quad \text{subject to } Z \stackrel{\text{d}}{\sim} \tilde{\mu} \text{ and } W \stackrel{\text{d}}{\sim} \tilde{\nu}, \quad (35)$$

where $\tilde{\mu} = \mu \circ f^{-1}$ and $\tilde{\nu} = \nu \circ g^{-1}$.

To solve this problem, as discussed in Section 7, note that $\text{RVaR}_{p,q} = \mathcal{E}^h$, where h is an SS-distortion function with inflection point $r = 1 - p = 0.8$. Indeed, any point in $[1 - q, 1 - p]$ is an inflection point (see Figure 7), but only $r = 1 - p$ satisfies the condition in Theorem 5. By using Theorem 5, a maximizer to (35) is $C_{0.8}^\pm$. The interpretation of this transport plan (see Figure 3) is that, the worst 20% workers and the worst 20% co-workers are matched together following a comonotonic pattern to produce the low-quality products, and the best 80% workers and the best 80% co-workers are matched together following a counter-monotonic pattern, so that the medium-range products between 20% and 70% quantiles have good quality. Our universal optimality guarantees this solution regardless of the skill distributions $\tilde{\mu}$ and $\tilde{\nu}$.

Some numerical examples are reported in Table 1 below, where we compute the value of $\text{RVaR}_{p,q}(Z + W)$ in (35) for five different couplings, and some continuous marginal distributions. Because we have explicit methods to simulate from each copula, we calculate each number in Table 1 by the average of the 20% and 70% sample quantiles from a Monte Carlo simulation of sample size 10,000,000. From the results, we can see that the optimal copula leads to a substantially larger range-average quality.

⁴We assume the number of workers and coworkers is very large so that it does not hurt to use continuous distributions and precise probability levels.

Coupling	$\tilde{\mu}$ and $\tilde{\nu}$	
	Beta(5, 2), uniform[0, 1]	Expo(1), Log-normal(0, 1)
Comonotonicity	1.1585	1.5717
Counter-monotonicity	1.2017	2.0210
Independence	1.1644	1.9123
Sub-optimal copula $C_{0.5}^{\pm}$	1.2123	2.0304
Optimal copula $C_{0.8}^{\pm}$	1.2931	2.3824

Table 1: The range-average quality $\text{RVaR}_{0.2,0.7}(Z + W)$ where $Z \stackrel{d}{\sim} \tilde{\mu}$ and $W \stackrel{d}{\sim} \tilde{\nu}$

9 Concluding remarks

The framework of non-linear and distorted optimal transport is proposed, and we obtain many results in this framework. As our main findings, we established universal optimality and duality for some classes of non-linear expectations and cost functions, and fully characterized universal optimality for ISS and SS distortion functions with linear cost. The optimal copula is also illustrated with an application of economic production. Since the framework is new, there are many open questions, and we discuss some of them below.

First, we have focused on transport on the real line, and the structure of the real line is used extensively. For the applications that motivated the paper, behavioral decision making, robust optimization and robust risk aggregation, the real line is the most natural space, but for other applications, more general spaces are needed. To extend our results to \mathbb{R}^d with $d \geq 2$ or general Polish spaces requires thorough separate study.

Second, we established weak duality in Theorem 2 only in case \mathcal{E} is super-linear. The case of a general non-linear expectation is difficult to analyze, and we only have a simple bound (20) for $\mathcal{E} = \mathcal{E}^h$. Similarly, strong duality is only established for convex h and affine cost, and we wonder whether it also holds in more general situations.

Third, universally optimal couplings are obtained under the assumption that the distortion functions are convex, concave, ISS, or SS. Although these distortion functions are the most commonly used in the application domains, we wonder whether universally optimal couplings can be obtained, or proved to be non-existent, for more general forms of h . Moreover, optimal couplings for specific marginal distributions, other than all marginal distributions, are left for future research.

Fourth, most results in our paper are obtained for distorted expectations or sub-linear/super-linear expectations. General non-linear expectations that are neither sub-linear nor super-linear

will require substantially different techniques.

Fifth, the distorted optimal transport can be generalized to other settings other than the standard one we considered. Motivated by different applications, classic transport problems have been generalized to, among many others, the multi-marginal setting (see [Rachev and Rüschendorf \(1998\)](#); [Pass \(2015\)](#)), transport between vector-measures (see [Wang and Zhang \(2024\)](#)), transport between capacities (see [Ghossoub et al. \(2023\)](#)), and the constrained settings, such as those of martingale transport ([Beiglböck et al. \(2013\)](#); [Galichon et al. \(2014\)](#); [Beiglböck and Juillet \(2016\)](#)), supermartingale transport ([Nutz and Stebegg \(2018\)](#)) or directional transport ([Nutz and Wang \(2022\)](#)). Each of these generalizations can be combined with our framework of distorted transport. We expect them to lead to great additional challenges and new mathematics on optimal transport in the future.

Acknowledgements

RW acknowledge financial support from Canada Research Chairs (CRC-2022-00141) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

References

- Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, **46**(11), 1497–1512.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.
- Beiglböck, M., Henry-Labordère, P. and Penkner, F. (2013) Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, **17**(3), 477–501.
- Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Annals of Probability*, **44**(1), 42–106.
- Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.
- Blanchet, J., Lam, H., Liu, Y. and Wang, R. (2024). Convolution bounds on quantile aggregation. *Operations Research*, forthcoming.
- Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, **44**(2), 565–600.

- Cheung, K. C. (2010). Characterizing a comonotonic random vector by the distribution of the sum of its components. *Insurance: Mathematics and Economics*, **47**(2), 130–136.
- Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.
- Delbaen, F. (2012). *Monetary Utility Functions*. Osaka University Press, Osaka.
- Denneberg, D. (1994). *Non-additive Measure and Integral*. Springer Science & Business Media.
- Dhaene, J. and Kukush, A., Linders, D. and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, **2**(2), 319–328.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, **51**(4), 380–417.
- Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.
- Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A., (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.
- Embrechts, P., Wang, B. and Wang, R., (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.
- Ennaji, H., Mérigot, Q., Nenna, L. and Pass, B. (2024). Robust risk management via multi-marginal optimal transport. *Journal of Optimization Theory and Applications*, **202**, 554–581.
- Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Fourth Edition. Walter de Gruyter, Berlin.
- Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press.
- Galichon, A., Henry-Labordère, P. and Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Annals of Applied Probability*, **24**(1), 312–336.
- Ghossoub, M., Saunders, D. and Zhang, S. (2023). Bounds on Choquet risk measures in finite product spaces with ambiguous marginals. *Statistics and Risk Modeling*, **41**, 49–72.
- Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**(2), 141–153.
- Goh, J. and Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, **58**(4), 902–917.
- Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, **38**(1), 129–166.

- He, X. D. and Zhou, X. Y. (2011). Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science*, **57**(2), 315–331.
- Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. Third edition. McGraw-Hill.
- Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, **9**, 49–71.
- Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.
- Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.
- Makarov, G. D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and its Applications*, **26**, 803–806.
- McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised edition, New Jersey: Princeton University Press.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley, England.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. New York, Springer.
- Nutz, M. and Stebegg, F. (2018). Canonical supermartingale couplings. *Annals of Probability*, **46**(6), 3351–3398.
- Nutz, M. and Wang, R. (2022). The directional optimal transport. *Annals of Applied Probability*, **32**(2), 1400–1420.
- Pratt, J. W. (1988). Aversion to one risk in the presence of others. *Journal of Risk and Uncertainty*, **1**(4), 395–413.
- Pass, B. (2015). Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, **49**(6), 1771–1790.
- Peng, S. (2019). *Nonlinear Expectations and Stochastic Calculus under Uncertainty: With Robust CLT and G-Brownian Motion*. Springer, Berlin.
- Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, **3**(4), 323–343.
- Rachev, S. T., and Rüschendorf, L. (1998). *Mass Transportation Problems: Volume I: Theory* (Vol. 1). Springer, Berlin.

- Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, **14**(3), 623–632.
- Rüschendorf, L. (2013). *Mathematical Risk Analysis*. Springer, Heidelberg.
- Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*. Springer, New York.
- Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**(2), 255–261.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.
- Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4), 297–323.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer-Verlag, Berlin.
- Wakker, P. P. (2001). Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle. *Econometrica*, **69**(4), 1039–1059.
- Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.
- Wald, A. (1950). *Statistical Decision Functions*. John Wiley & Sons.
- Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.
- Wang, R., Xu, Z. Q. and Zhou, X. Y. (2019). Dual utilities on risk aggregation under dependence uncertainty. *Finance and Stochastics*, **23**(4), 1025–1048.
- Wang, R. and Zhang, Z. (2024). Simultaneous optimal transport. *Transactions of the American Mathematical Society*, forthcoming.
- Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**(3), 1413–1429.
- Wiesemann, W., Kuhn, D. and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1203–1466.
- Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.