Annals of Operations Research
https://doi.org/10.1007/s10479-024-06401-x

ORIGINAL RESEARCH

Distributionally robust multiobjective optimization with application to risk measure theory

Davide La Torre[1] · **Matteo Rocca**[2]

Received: 28 March 2024 / Accepted: 15 November 2024

Abstract
We introduce the concept of a distributionally robust multiobjective optimization problem, which offers a comprehensive framework for addressing issues related to the statistical estimation of unknown probabilities. By employing scalarization methods, we establish optimality conditions, followed by the exploration of applications in financial portfolio management and risk assessment.

Keywords Robustness · Probability measure · Multiobjective optimization · Stochastic optimization · Portfolio optimization · Risk measures

1 Introduction

In recent years, both stochastic optimization and multiobjective optimization have experienced substantial growth in research and the development of new applications. For an overview of stochastic multiobjective optimization, interested readers may refer to sources such as Ben Abdelaziz (2012) and Gutjahr and Pichler (2016). One of the most significant applications of stochastic multiobjective optimization undoubtedly lies in portfolio optimization. In fact, financial decision-makers frequently encounter investment strategies that incorporate multiple criteria while also being susceptible to uncertainty and randomness.

In any decision-making scenario, randomness assumes a pivotal role, underscoring the necessity of generating solutions that not only optimize outcomes but also demonstrate resilience against noise and perturbations.

This paper analyzes the concept of Distributionally Robust Multiobjective Optimization. A robust optimization model seeks solutions that remain effective even when faced with unforeseen changes or disturbances. We will establish scalarization results and optimality conditions in the convex case. Additionally, we will demonstrate how the concept of robustness is intricately connected to solving a minmax problem across the range of feasible distributions.

✉ Matteo Rocca
matteo.rocca@uninsubria.it

1 SKEMA Business School, Université Côte d'Azur, Sophia Antipolis Campus, Nice, France

2 Department of Economics, Universitá degli Studi dell'Insubria, Via Monte Generoso 71, 21100 Varese, Italy

When contemplating robust solutions to a multiobjective optimization problem, a phenomenon known as “loss of efficiency” arises compared to the solution obtained in the nominal problem. For instance, in the scalar case and assuming a minimization problem, this entails that the robust optimal value is greater than or equal to the optimal value of the nominal problem, and robust solutions represent ε solutions of the nominal problem. Another significant result presented in this paper pertains to the estimation of the efficient frontiers of both the nominal problem and the robust problem, along with the associated efficiency loss. This is accomplished through appropriate nonlinear scalarizations and set distances, contingent upon the configuration of the uncertainty set. Within this framework, we will conduct a sensitivity analysis of the robust solutions concerning variations in the uncertainty set. The Monge-Kantorovich distance between distributions will emerge as a pivotal metric for estimating the loss of efficiency.

The research in this paper presents techniques that can be applied to a variety of financial portfolio problems involving risk measures. Risk measures are critical tools used in finance to quantify and manage the uncertainty and potential losses associated with investments. By incorporating these risk measures, the methodologies described can be adapted to tackle diverse portfolio optimization and management challenges. The versatility of the research findings suggests their broad relevance and potential impact, as finance, investment, and risk management professionals may find these techniques valuable for improving portfolio decision-making and risk mitigation practices (La Torre & Mendivil, 2018, 2022; La Torre et al., 2021).

The paper is organized as follows. Next Sect. 2 presents some relevant problems in Portfolio Theory and Risk Measures that can be addressed by using the theory developed in this paper. Section 3 presents some preliminary results and properties. Section 4 introduces the basic concepts and definitions in multiobjective optimization. Section 5 presents the main results, while Sect. 6 discusses some optimality conditions for efficient solutions and also provides stability results. The applications to financial portfolio optimization and risk management are presented in Sect. 7. Finally, Sect. 9, concludes the paper.

2 Motivating examples

Risk measures are essential tools used in various fields such as finance, economics, engineering, and insurance to quantify and manage uncertainty and potential losses associated with different activities or investments. They provide valuable insights into the level of risk inherent in a particular decision, portfolio, or investment strategy.

Risk measures are applied across various domains, including investment management, portfolio construction, asset allocation, risk assessment, derivatives pricing, and regulatory compliance. They assist policymakers in making informed decisions, managing risks effectively, and achieving their financial objectives.

In the literature, there are different types of risk measures:

- Statistical Measures: Statistical risk measures employ mathematical and statistical methods to assess risk. They encompass metrics such as standard deviation, variance, and coefficient of variation, which offer insights into the dispersion of returns or outcomes around the mean (McNeil et al., 2015).
- Value-at-Risk (VaR): VaR is a widely used risk measure that quantifies the maximum potential loss of a portfolio or investment over a specified time horizon at a given confi-

dence level. VaR helps investors understand the worst-case scenario they may encounter under normal market conditions (Jorion, 2006).

- Conditional Value-at-Risk (CVaR): CVaR, also known as expected shortfall, extends the concept of VaR by providing information about the expected loss in the tail of the distribution beyond the VaR threshold. It offers insights into the severity of losses beyond the VaR level (Acerbi & Tasche, 2002).
- Sharpe Ratio: The Sharpe Ratio measures the risk-adjusted return of an investment or portfolio by comparing the excess return (return above the risk-free rate) to the standard deviation of returns. It helps investors evaluate the return earned per unit of risk taken (Sharpe, 1994).
- Drawdowns: Drawdowns quantify the peak-to-trough decrease in the value of an investment or portfolio over a designated period. By conducting drawdown analysis, investors gain insight into both the scale and duration of possible losses (Lo, 2002).

Among risk measures, a relevant role is played by the notion of *coherent risk measure*. This is a definition of risk that possesses desirable properties for risk measures (Acerbi & Tasche, 2002; Artzner et al., 1999). Coherent risk measures were introduced to address limitations and inconsistencies associated with traditional risk measures, such as value-at-risk (VaR), which may not fully capture the underlying risk characteristics of a portfolio or investment.

They have become increasingly important in contemporary risk management and financial theory thanks to their robust mathematical characteristics and adherence to risk management principles. These measures offer a consistent framework for evaluating and controlling risk in diverse financial and economic contexts, ensuring accuracy and dependability in risk assessment and decision-making procedures.

More precisely, the definition of coherent risk measure reads as follows. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$, $\mathcal{X}$ the set of random vectors $X : \Omega \to \mathbb{R}^d$ representing the profit of financial positions. We denote by $\mathcal{P}$ the set of probability distributions over Ω. A coherent vector risk measure (see e.g. Jouini et al. (2004)) is a function $\mathcal{R} : \mathcal{X} \to \mathbb{R}^d$ such that

- $\mathcal{R}(X + y) = \mathcal{R}(X) - y$ for all $y \in \mathbb{R}^d$.
- $\mathcal{R}(X) \geq_{\mathbb{R}^d_+} \mathcal{R}(Y)$ whenever $Y \leq_{\mathbb{R}^d_+} X$.
- $\mathcal{R}(tX) = t\mathcal{R}(X)$ for all $t > 0$ and $X \in \mathcal{X}$,
- $\mathcal{R}(X) + \mathcal{R}(Y) \leq_{\mathbb{R}^d_+} \mathcal{R}(X + Y)$ for all $X, Y \in \mathcal{X}$.
- Finally, $\mathcal{R}$ is $\mathbb{R}^d_+$-convex if $t\mathcal{R}(X) + (1 - t)\mathcal{R}(Y) \geq_{\mathbb{R}^d_+} \mathcal{R}(tX + (1 - t)Y)$ for any $X, Y \in \mathcal{X}$.

Here, for vectors $a, b \in \mathbb{R}^d$, $a \leq_{\mathbb{R}^d} b$ means $a \in b - \mathbb{R}^d$. We now recall the dual representation of coherent vector risk measure (see e.g. Cascos and Molchanov (2007)).

Theorem 2.1 *Let $\mathcal{R} : \mathcal{X} \to \mathbb{R}^d$ be a vector-valued risk measure. The following claims are equivalent:*

1. *$\mathcal{R}$ is coherent*
2. *there exist $\mathcal{P}_i \subseteq \mathcal{P}$ $i = 1, \ldots, d$ such that for all $X \in \mathcal{X}$*

$$\mathcal{R}(X) = \begin{bmatrix} \sup_{P_1 \in \mathcal{P}_1} \mathbb{E}_{P_1}[-X_1] \\ \vdots \\ \sup_{P_d \in \mathcal{P}_d} \mathbb{E}_{P_d}[-X_d] \end{bmatrix} \tag{1}$$

If $\psi : \mathbb{R}^n \to \mathcal{X}$ is a portfolio aggregator (see e.g. Shapiro et al. (2009)), then $\mathcal{R}(\psi(x))$ is a composite vector risk measure. Concavity of ψ ensures convexity of $\mathcal{R}$ (see Mastrogiacomo and Rocca (2021)). Similarly to Theorem 2.1 we have the following result.

Theorem 2.2 *Let $\mathcal{R} : \mathcal{X} \to \mathbb{R}^d$ be a vector-valued risk measure. The following claims are equivalent:*

1. *$\mathcal{R}$ is coherent*
2. *there exist $\mathcal{P}_i \subseteq \mathcal{P}$ $i = 1, \dots, d$ such that for all $X \in \mathcal{X}$*

$$\mathcal{R}(\psi(x)) = \begin{bmatrix} \sup_{P_1 \in \mathcal{P}_1} \mathbb{E}_{P_1}[-\psi_1(x)] \\ \vdots \\ \sup_{P_d \in \mathcal{P}_d} \mathbb{E}_{P_d}[-\psi_d(x)] \end{bmatrix} \tag{2}$$

In the sequel of this paper we will deal with the multiobjective optimization problem:

$$\min_{x \in X} \mathcal{R}(\psi(x)) \tag{3}$$

where

$$X = \left\{ x \in \mathbb{R}^n : x_i \geq 0,\ i = 1, \dots, n;\ \sum_{i=1}^{n} x_i \leq 1 \right\}$$

In the following we denote

$$\mathcal{P}_i(x) = \left\{ \bar{P}_i \in \mathcal{P}_i : \sup_{P_i \in \mathcal{P}_i} \mathbb{E}_{P_i \in \mathcal{P}_i}[-\psi_i(x)] = \mathbb{E}_{\bar{P}_i}[-\psi_i(x)] \right\} \tag{4}$$

The theory we are going to develop in the next sections will help to provide optimality conditions and robustness conditions for this problem.

3 Mathematical preliminaries

Let $(\Omega, \mathcal{B}, \mu)$ be a sample space, where Ω is the space of events, $\mathcal{B}$ is the Borel σ-algebra on Ω, and μ is a probability measure. Let us denote by $\omega \in \Omega$ a simple event of the space Ω and by X a subset of $\mathbb{R}^n$. Following Gutjahr and Pichler (2016), we consider stochastic multiobjective problems of the form

$$\min_{x \in X} (\phi_1(x, \omega), \dots, \phi_m(x, \omega)), \tag{5}$$

where $\phi_j : X \times \Omega \to \mathbb{R}$, $j = 1, \dots, m$. Associated with every function ϕ_j there is a vector

$$(\mathcal{F}_j^{(1)}(\phi_j(x, \omega), \mu), \dots, \mathcal{F}_j^{(r_j)}(\phi_j(x, \omega), \mu)) \tag{6}$$

composed by functionals $\mathcal{F}_j^{(s)}(\phi_j(x, \omega), \mu)$ of $\phi_j(x, \omega)$, which can be the expectation, the variance, a quantile, some risk measure, or some other summary statistics of the random variable $\phi_j(x, \omega)$. In this way, we associate with the problem (5) the deterministic multiobjective optimization problem defined as

$$\min_{x \in X} (z_1^{(1)}(x, \mu), \dots, z_1^{(r_1)}(x, \mu), \dots, z_m^{(1)}(x, \mu), \dots, z_m^{(r_m)}(x, \mu)) \tag{7}$$

where $z_j^{(s)}(x, \mu) = \mathcal{F}_j^{(s)}(\phi_j(x, \omega), \mu)$, $j = 1, \dots, m$, $s = 1, \dots, r_j$. The multiobjective approach can be considered as a generalization of traditional single-valued mean-risk models which correspond to the take $m = 1$. In Caballero et al. (2001, 2004) the following particular cases are considered:

1. $r_j = 1$ for every j and we have that $\mathcal{F}_j^{(1)}(\phi_j(x,\omega),\mu) = \mathbb{E}_\mu(\phi_j(x,\omega))$
2. $r_j = 1$ for every j and we have that $\mathcal{F}_j^{(1)}(\phi_j(x,\omega),\mu) = \sigma_\mu^2(\phi_j(x,\omega))$
3. $r_j = 2$ for every j and we have that $\mathcal{F}_j^{(1)}(\phi_j(x,\omega),\mu) = \mathbb{E}_\mu(\phi_j(x,\omega))$, $r_j = 1$, $\mathcal{F}_j^{(2)}(\phi_j(x,\omega),\mu) = \sigma_\mu^2(\phi_j(x,\omega))$

In the preceding model, we assume that the probability distribution μ on Ω is known. However, in practical scenarios, our understanding of the statistical properties of the model parameters is often incomplete, and the associated probability distributions are never fully determined (i.e., we encounter Knightian ambiguity regarding probability distributions). Specifically, the probability distribution characterizing the uncertainty of model parameters is known ambiguously. A common approach to address this ambiguity, from a statistical perspective, involves estimating the probability distribution using empirical samples. Consequently, the decision-making process can be conducted based on the estimated probability distribution. However, this statistical approach may yield imprecise conclusions, either due to the presence of outliers or errors in the sampling procedure.

Ambiguous stochastic optimization presents a modeling approach designed to shield the decision-maker from ambiguity inherent in the underlying probability distribution. Several solutions have been suggested in the literature to address this issue, drawing on the concepts of imprecise probability and set-valued probability (refer to La Torre and Mendivil (2018, 2022), La Torre et al. (2021) for more details).

Another approach to handle ambiguity involves assuming that the underlying probability distribution is unknown and resides within an ambiguity set of probability distributions. In this paper, we will explore this methodology, which, akin to robust optimization, mitigates ambiguity in probability distribution through a worst-case (min-max) strategy. This approach is what we define as Distributionally Robust Multiobjective Optimization.

When the single objective case is considered, i.e. $j = 1$, such an approach was investigated in Shapiro and Kleywegt (2010). In Shapiro and Kleywegt (2010) the authors considered the particular case in which $r_1 = 1$ and $\mathcal{F}_1^{(1)}(\phi_1(x,\omega),\mu) = \mathbb{E}_\mu(\phi_1(x,\omega))$.

Now let us suppose that a family $\mathcal{A}$ of probability distributions (measures) is provided. We also suppose that $\mathcal{A}$ is a compact set with respect to the Monge-Kantorovich distance between proability measures. $\mathcal{A}$ models the set of all possible distributions that are relevant for the considered problem. Then one may hedge against the worst expected value resulting from the distributions in the set $\mathcal{A}$ by solving the distributionally robust optimization problem (DRP):

$$\min_{x\in X} \max_{\mu\in\mathcal{A}} \mathbb{E}_\mu(\phi_1(x,\omega)) \tag{8}$$

Next sections are devoted to the analysis of the mathematical properties of this model in vector form as well as its potential application to portfolio optimization and risk management.

4 Multiobjective optimization

In Multiobjective Optimization (MOP), the goal is to optimize a system or process considering multiple conflicting criteria simultaneously. Unlike single-objective optimization, where there is only one objective function to be optimized, MOP involves optimizing several objectives concurrently.

The primary challenge in MOP lies in finding solutions that balance trade-offs among the conflicting objectives. These solutions are known as Pareto-optimal solutions or Pareto fronts, where improving one objective typically leads to deterioration in another Ehrgott (2005).

Various techniques have been developed to solve MOP, including evolutionary algorithms, mathematical programming methods, and heuristic approaches. Evolutionary algorithms, such as NSGA-II (Non-dominated Sorting Genetic Algorithm II), are particularly popular due to their ability to efficiently explore the solution space and identify Pareto-optimal solutions (Coello Coello et al., 2007; Deb et al., 2002).

In MOP, decision-makers must carefully analyze the trade-offs between different objectives to make informed decisions. This often involves visualizing the Pareto front to understand the relationships among the objectives and identify the most suitable solutions based on decision-maker preferences and constraints.

MOP provides a powerful framework for addressing complex optimization problems with multiple conflicting objectives, offering insights into the inherent trade-offs and enabling decision-makers to make well-informed decisions.

In mathematical details, given a compact subset Ξ of $\mathbb{R}^s$ and a vector-valued map $J : \Xi \subset \mathbb{R}^s \to \mathbb{R}^p$, $J = (J_1, \ldots, J_p)$ with $J_i : \Xi \subset \mathbb{R}^s \to \mathbb{R}$, any finite-dimensional MOP problem can be written:

$$\min_{\xi \in \Xi} J(\xi). \tag{MP}$$

We assume that an ordering on $\mathbb{R}^p$ is induced by the Pareto cone $\mathbb{R}^p_+$. A point $\xi \in \Xi$ is said to be Pareto optimal or efficient if it is feasible and, for any possible $\xi' \in X$, $J(\xi) \leq_{\mathbb{R}^p_+} J(\xi')$ implies $J(\xi) = J(\xi')$. Equivalently, a point $\xi \in \Xi$ is said to be Pareto efficient if $(J(\xi) - \mathbb{R}^p_+) \cap J(\Xi) = \{J(\xi)\}$. We denote by $\mathrm{Eff}(J)$ the set of efficient points for function J. We say, instead, that $\xi \in \Xi$ is weakly Pareto efficient when $(J(\xi) - \mathrm{int}\, \mathbb{R}^p_+) \cap J(\Xi) = \emptyset$. We denote by $\mathrm{WEff}(J)$ the set of weakly efficient points for function J. Finally, we mention also that the point ξ is properly Pareto efficient (with respect to C) when there exists a cone C with $\mathbb{R}^p_+ \subseteq \mathrm{int}\, C$ such that ξ is Pareto efficient with respect to the cone C, i.e.

$$(J(\xi) - C) \cap J(\Xi) = \{J(\xi)\} \tag{9}$$

In the following $\mathrm{PEff}_C(J)$ denotes the set of Pareto properly efficient points. Obviously, every properly Pareto efficient point is also Pareto efficient. For a deeper exposition of the notions of Pareto efficiency, one can see (Sawaragi et al., 1985).

Scalarization techniques allow reducing a MOP problem to a single criterion one (see (Sawaragi et al., 1985)). Linear scalarization is the most classical scalarization approach but other nonlinear techniques are available under convexity assumptions. In this context, a MOP model is reduced to a single criterion one by summing up all criteria with different weights, which gives the relative importance of each criterion for the DM. Hence, a scalarized version of a MOP model is given by:

$$\min_{\xi \in \Xi} \beta \cdot J(\xi) = \sum_{i=1}^{p} \beta_i J_i(\xi), \tag{10}$$

where $\beta = (\beta_1, \ldots, \beta_p)$ is a vector taking values in $\mathbb{R}^p_+$.

Nonlinear scalarization techniques can also be used for addressing MOP problems. In nonlinear scalarization, a nonlinear combination of the objective functions is used to construct a scalar objective function. This scalar function allows for the representation of the trade-offs and preferences among the multiple objectives. Nonlinear scalarization techniques offer flexibility in capturing complex relationships among objectives and they enable decision-makers to express nonlinear preferences and capture non-convex trade-offs in the optimization process.

5 Main results

In this section we provide some scalarization results for different types of Pareto efficient solutions. We assume X is a compact and convex subset of $\mathbb{R}^n$, Ω is a convex compact subset of a topological space. We assume the space $\mathcal{P}$ of probability distributions over Ω is endowed with a suitable topology and $\mathcal{A}$ is a convex and compact subset of $\mathcal{P}$. We set (we assume maxima are attained)

$$Z_j^{(s)}(x) = \max_{\mu \in \mathcal{A}} z_j^{(s)}(x, \mu) \tag{11}$$

and consider the Distributionally Robust Multiobjective Optimization Problem (DRMP) defined as follows

$$\min_{x \in X} (Z_1^{(1)}(x), \ldots, Z_1^{(r_1)}(x), \ldots, Z_m^{(1)}(x), \ldots, Z_m^{(r_m)}(x)) \tag{DRMP}$$

We set also $\bar{\mathcal{A}}_j^{(s)}(x) = \{\mu \in \mathcal{A} : Z_j^{(s)}(x) = z_j^{(s)}(x, \mu)\}$

Remark 5.1 Assume $\mathcal{A} = \{\delta_\omega, \ \omega \in \Omega\}$ (the set of Dirac measures over Ω), $r_j = 1$ for every $j = 1, \ldots, m$ and $\mathcal{F}_j^1(\phi_j(x, \omega), \mu) = \mathbb{E}_\mu(\phi_j(x, \omega))$. Then problem (DRMP) reduces to

$$\min_{x \in X} (\max_{\omega \in \Omega} \phi_1(x, \omega), \ldots, \max_{\omega \in \Omega} \phi_1(x, \omega)) \tag{12}$$

i.e. to a robust multiobjective optimization problem (see e.g. Ben-tal et al. (2009)).

We prove the following result.

Theorem 5.1 *In problem (DRMP) assume $z_j^{(s)}(x, \mu)$ are convex with respect to $x \in X$ and concave with respect to $\mu \in \mathcal{A}$.*

(i) $\bar{x} \in X$ is a properly efficient solution for for problem (DRMP) if and only if there exist numbers $\bar{\lambda}_j^{(s)} > 0$, and probability distributions $\bar{\mu}_j^{(s)} \in \mathcal{A}$, $j = 1, \ldots, m$, $s = 1, \ldots, r_j$ with $\bar{\mu}_j^{(s)} \in \bar{\mathcal{A}}_j^{(s)}(\bar{x})$ such that

$$\sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} z_j^{(s)}(\bar{x}, \bar{\mu}_j^{(s)}) \leq \sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} z_j^{(s)}(x, \bar{\mu}_j^{(s)}), \quad \forall x \in X \tag{13}$$

i.e. $\bar{x}$ minimizes function $\sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} z_j(x, \bar{\mu}_j^{(s)})$ over X. This is equivalent to say that $\bar{x}$ is properly efficient for the problem of minimizing

$$(z_1^{(1)}(x, \bar{\mu}_1^{(1)}), \ldots, z_1^{(1)}(x, \bar{\mu}_1^{(r_1)}), \ldots, z_m^{(1)}(x, \bar{\mu}_m^{(r_m)}), \ldots, z_m^{(r_m)}(x, \bar{\mu}_m^{(r_m)})) \tag{14}$$

(ii) $\bar{x} \in X$ is a weakly efficient solution for problem (DRMP) if and only if there exist numbers $\bar{\lambda}_j^{(s)} \geq 0$, and probability distributions $\bar{\mu}_j^{(s)} \in \mathcal{A}$ $j = 1, \ldots, m$, $s = 1, \ldots, r_j$ with $\bar{\mu}_j^{(s)} \in \bar{\mathcal{A}}_j^{(s)}(x)$ when $\bar{\lambda}_j^{(s)} > 0$ such that

$$\sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} z_j^{(s)}(\bar{x}, \bar{\mu}_j^{(s)}) \leq \sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} z_j^{(s)}(x, \bar{\mu}_j^{(s)}), \quad \forall x \in X \tag{15}$$

i.e. $\bar{x}$ minimizes function $\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j(x,\bar{\mu}_j^{(s)})$ over X. This is equivalent to say that $\bar{x}$ is weakly efficient for the problem of minimizing the multiobjective function (14)

Proof (i) Since functions $z_j^{(s)}(x,\mu)$ are convex w.r.t. x, functions $Z_j^{(s)}(x)$ are convex, it is known (see e.g. Geoffrion (1968)) that $\bar{x}\in X$ is a properly efficient solution for problem (DRMP) if and only if there exist $\bar{\lambda}_j^{(s)}>0,\ j=1,\ldots,m,\ s=1,\ldots,r_j$ such that $\forall x\in X$ it holds:

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} Z_j^{(s)}(\bar{x}) \le \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} Z_j^{(s)}(x),\quad \forall x\in X \tag{16}$$

i.e.

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\mu\in\mathcal{A}} z_j^{s}(\bar{x},\mu) \le \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\mu\in\mathcal{A}} z_j^{(s)}(x,\mu),\quad \forall x\in X \tag{17}$$

i.e.

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\mu\in\mathcal{A}} z_j^{s}(\bar{x},\mu) \le \min_{x\in X}\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\mu\in\mathcal{A}} z_j^{(s)}(x,\mu) \tag{18}$$

Now let $p=r_1+\cdots+r_m$ and consider the set $\mathcal{A}^p=\mathcal{A}\times\cdots\times\mathcal{A}$ (p times) and let $\phi=(\mu_1^{(1)},\ldots,\mu_1^{(r_1)},\ldots,\mu_m^{(1)},\ldots,\mu_m^{(r_m)})\in\mathcal{A}^p$. It holds

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\mu\in\mathcal{A}} z_j^{(s)}(x,\mu_j^{(s)}) = \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\phi\in\mathcal{A}^p} z_j^{(s)}(x,\mu_j^{(s)}) = \tag{19}$$

$$= \max_{\phi\in\mathcal{A}^p}\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(x,\mu_j^{(s)}) \tag{20}$$

Since functions $z_j^{(s)}$ are convex w.r.t. x and concave w.r.t. μ, function

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(x,\mu_j^{(s)}) \tag{21}$$

is convex w.r.t. x and concave w.r.t. $\mu_j^{(s)}$, $j=1,\ldots,m,\ s=1,\ldots,r_j$. By Ky-Fan minimax Theorem it holds

$$\max_{\phi\in\mathcal{A}^p}\min\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(x,\mu_j^{(s)}) \ge \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\phi\in\mathcal{A}^p} z_j^{(s)}(\bar{x},\mu_j^{(s)}) \tag{22}$$

Hence there exist $\bar{\mu}_j^{(s)}\in\mathcal{A},\ j=1,\ldots,m,\ s=1,\ldots,r_j$ such that for every $x\in X$

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(x,\bar{\mu}_j^{(s)}) \ge \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(\bar{x},\bar{\mu}_j^{(s)}) \tag{23}$$

and

$$\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)} z_j^{(s)}(\bar{x},\bar{\mu}_j^{(s)}) = \sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}\max_{\phi\in\mathcal{A}^p} z_j^{(s)}(\bar{x},\mu_j^{(s)}) \tag{24}$$

Since $\lambda_j^{(s)} > 0,\ j = 1, \ldots, m,\ s = 1, \ldots, r_j$, the last equality gives

$$z_j^{(s)}(\bar{x}, \bar{\mu}_j^{(s)}) = \max_{\phi \in \mathcal{A}^p} z_j^{(s)}(\bar{x}, \mu_j^{(s)}),\ j = 1, \ldots, m,\ s = 1, \ldots, r_j \tag{25}$$

i.e. $\bar{\mu}_j^{(s)} \in \bar{\mathcal{A}}_j^{(s)}(\bar{x})$

(ii) Since functions $z_j^{(s)}(x, \mu)$ are convex w.r.t. x, functions $Z_j^{(s)}(x)$ are convex, it is known (see e.g. Luc (1989)) that $\bar{x} \in X$ is a weakly efficient solution for problem (DRMP) if and only if there exist $\bar{\lambda}_j^{(s)} \geq 0,\ j = 1, \ldots, m,\ s = 1, \ldots, r_j$ such that $\forall x \in X$ it holds:

$$\sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} Z_j^{(s)}(\bar{x}) \leq \sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} Z_j^{(s)}(x), \quad \forall x \in X \tag{26}$$

The proof follows now in a fashion similar to that of point i).

□

Remark 5.2 (i) Assume $\phi_(x, \mu)$ are convex w.r.t. $x \in X$, $r_j = 1$ and $z_j^{(1)}(x, \mu) = \mathbb{E}_\mu(\phi_j(x, \mu)$. Then Problem (DRMP) satisfies the assumptions of Theorem 5.1.

(ii) Assume $\phi_(x, \mu)$ are linear w.r.t. $x \in X, r_j = 2$ for every j and $z_j^{(1)}(x, \mu) = \mathbb{E}_\mu(\phi_j(x, \mu)$, $z_j^{(2)}(x, \mu) = \sigma_\mu^2(\phi_j(x, \omega))$. Then Problem (DRMP) satisfies the assumptions of Theorem 5.1. the same holds if $r_j = 1$ for every j and $z_j^{(1)}(x, \mu) = \sigma_\mu^2(\phi_j(x, \omega))$.

6 Optimality conditions and stability results

This section is devoted to the presentation of optimality conditions and stability results. Optimality conditions establish criteria for identifying optimal solutions, delineating conditions under which a given solution is optimal. Stability results, on the other hand, assess the robustness and sensitivity of optimization algorithms to perturbations in problem data or algorithmic parameters. They ensure that computed solutions remain valid under varying conditions. The results in this section are proved by means of linear and nonlinear scalarization techniques.

Theorem 6.1 *In problem (DRMP) assume $z_j^{(s)}(x, \mu)$ are convex with respect to $x \in X$ and concave with respect to $\mu \in \mathcal{A}$.*

i) Let $\bar{x} \in X$ be a properly efficient solution for problem (DRMP). Then there exist $\bar{\lambda}_j^{(s)} > 0$, $\bar{\mu}_j^{(s)} \in \bar{\mathcal{A}}^{(s)}(\bar{x}),\ j = 1, \ldots, m,\ s = 1, \ldots, r_j$ such that

$$0 \in \sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} \partial_1 z_j^{(s)}(\bar{x}, \bar{\mu}_j^{(s)}) + N_X(\bar{x}) \tag{27}$$

where $N_X(\bar{x}) = \{v \in \mathbb{R}^n : \langle v, x \rangle \leq 0\}$ denotes the normal cone to X at $\bar{x}$.

ii) Let $\bar{x}$ be a weakly efficient solution for problem (DRMP). Then there exist $\bar{\lambda}_j^{(s)} \geq 0$, $\bar{\mu}_j^{(s)} \in \mathcal{A}$ with $\bar{\mu}_j^{(s)} \in \bar{\mathcal{A}}^{(s)}(\bar{x}),\ j = 1, \ldots, m,\ s = 1, \ldots, r_j$ such that

$$0 \in \sum_{j=1}^{m} \sum_{s=1}^{r_j} \bar{\lambda}_j^{(s)} \partial_1 z_j^{(s)}(\bar{x}, \bar{\mu}_j^{(s)}) + N_X(\bar{x}) \tag{28}$$

Proof It is immediate since conditions 27 and 28 are necessary for $\bar{x}$ to be a minimizer of function $\sum_{j=1}^{m}\sum_{s=1}^{r_j}\bar{\lambda}_j^{(s)}z_j(x,\bar{\mu}_j^{(s)})$ over X (see e.g. Rockafellar and Wets (1998)). □

Now we consider stability conditions in the case $r_j = 1$ for every j and

$$\mathcal{F}_j^{(1)}(\phi_j(x,\omega),\mu) = \mathbb{E}_\mu(\phi_j(x,\omega))$$

Define the Monge-Kantorovich distance between probability distributions as:

$$d_M(\mu_1,\mu_2) = \sup_{f\in Lip_1(\Omega)}\left(\int_\Omega f(\omega)d\mu_1(\omega) - \int_\Omega f(\omega)d\mu_2(\omega)\right) \tag{29}$$

Let $\mathcal{A}^0 \subseteq \mathcal{A}$ and $\mathcal{A}^\lambda = (1-\lambda)\mathcal{A}^0 + \lambda\mathcal{A}$. The parameter λ describes the magnitude of ambiguity about the probability distribution. We set (we assume maxima are attained)

$$Z_j^\lambda(x) = \max_{\mu\in\mathcal{A}^\lambda}\mathbb{E}_\mu(\phi_j(x,\omega)) \tag{30}$$

and consider the problem

$$\min_{x\in X}\ (Z_1^\lambda(x),\ldots,Z_m^\lambda(x)) \tag{DRMP$^\lambda$}$$

We assume $Z_j^0 = \mathbb{E}_{\mu^0}\phi_j(x,\omega)$ is lower bounded on X. Remember that $A\subset\mathbb{R}^m$ is said to be $\mathbb{R}^m_+$-closed if $A+\mathbb{R}^m_+$ is closed and A is said to be minorized if there exists $\underline{a}\in A$ such that $\underline{a}+\mathbb{R}^m_+\supseteq A$. When $\mathrm{Im}\,f^\lambda$ is $\mathbb{R}^m_+$-closed for $\lambda\in[0,1]$, we can see that Min (RMP$^\lambda$) is non-empty and the domination property for $\mathrm{Im}\,f^\lambda$ holds, that is,

$$\mathrm{Im}Z^\lambda \subset \mathrm{Min}\,(\mathrm{DRMP}^\lambda) + \mathbb{R}^m_+ \tag{31}$$

holds because $\mathrm{Im}Z^\lambda$ is minorized by the lower boundedness assumption of $Z_j^0(x)$ on X (see for example Lemma 3.5, Theorems 3.3 and 4.3 in Luc (1989)).

We need the following relation between sets introduced in Kuroiwa (2001). For $A,B\subseteq\mathbb{R}^m$ we denote

$$A \leq^l B \text{ if } A+\mathbb{R}^m_+ \supset B,$$

or equivalently, if for every $b\in B$ there exists $a\in A$ such that $a\leq_{\mathbb{R}^m_+} b$ holds. The relation is reflexive and transitive, but not antisymmetric. By using the notation, the inclusion of domination property (31) can be written as follows:

$$\mathrm{Min}\,\mathrm{DRMP}^\lambda \leq^l \mathrm{Im}Z^\lambda.$$

Proposition 6.1 *Assume that* $\mathrm{Im}(Z^0)$ *is* $\mathbb{R}^m_+$*-closed. Then set relation*

$$\mathrm{Min}\,\mathrm{DRMP}^0 \leq^l \mathrm{Im}(Z^\lambda)$$

holds for every $\lambda\in[0,1]$, *and hence, in particular*

$$\mathrm{Min}\,\mathrm{DRMP}^0 \leq^l \mathrm{Min}(\mathrm{DRMP}^\lambda) \textit{ and } \mathrm{WMin}\,\mathrm{DRMP}^0 \leq^l \mathrm{WMin}(\mathrm{DRMP}^\lambda).$$

Proof For every $x\in X$, by using the domination property for $\mathrm{Im}(Z^0)$, there exists $\hat{x}\in X$ such that $Z^0(\hat{x})\in\mathrm{Min}(\mathrm{DRMP}^0)$ and $Z^0(\hat{x})\leq Z^0(x)$. From $Z^0(x)\leq Z^\lambda(x)$, then we have

$$Z^\lambda(x) = Z^0(\hat{x}) + (Z^\lambda(x)-Z^0(x)) + (Z^0(x)-Z^0(\hat{x})) \in \mathrm{Min}(\mathrm{DRMP}^0)+\mathbb{R}^m_+.$$

This shows $\mathrm{Min}\,\mathrm{RMP}^0 \leq^l \mathrm{Im}(Z^\lambda)$. □

In order to prove the main result about location of the optimal values in problem DRMP$^\lambda$, we need the following proposition.

Proposition 6.2 *Assume that*

$$|\phi_j(x,\omega_1)-\phi_j(x,\omega_2)| \le k_j d(\omega_1,\omega_2),\ \forall x \in X \tag{32}$$

Then we have

$$|Z_j^\lambda(x)-Z_j^0(x)| \le \lambda k_j D(\mathcal{A}) \tag{33}$$

where $z_j(x,\mu)=\mathbb{E}_\mu(\phi_j(x,\omega))$, $D(\mathcal{A})=\sup_{\mu_1,\mu_2\in\mathcal{A}} d_M(\mu_1,\mu_2)$ *is the diameter of the set* $\mathcal{A}$, $Z_j^\lambda(x)=\sup_{\mu\in\mathcal{A}^\lambda} z_j(x,\mu)$

Proof We have $\mu\in\mathcal{A}^\lambda$ if and only if there exists $\mu^1\in\mathcal{A}$, $\mu^0\in\mathcal{A}^0$ such that $\mu=(1-\lambda)\mu^0+\lambda\mu^1$.

$$\begin{aligned}
Z_j^\lambda(x) &= \max_{\mu\in\mathcal{A}^\lambda} z_j(x,\mu) = \max_{\mu^0\in\mathcal{A}^0\mu^1\in\mathcal{A}} z_j(x,(1-\lambda)\mu^0+\lambda\mu^1) && (34)\\
&= \max_{\mu^0\in\mathcal{A}^0\mu^1\in\mathcal{A}} [(1-\lambda)z_j(x,\mu^0)+\lambda z_j(x,\mu^1)] && (35)\\
&= (1-\lambda)\max_{\mu^0\in\mathcal{A}^0} z_j(x,\mu^0)+\lambda\max_{\mu^1\in\mathcal{A}} z_j(x,\mu^1) && (36)\\
&= \max_{\mu^0\in\mathcal{A}^0} z_j(x,\mu^0)+\lambda[\max_{\mu^1\in\mathcal{A}} z_j(x,\mu^1)-\max_{\mu^0\in\mathcal{A}^0} z_j(x,\mu^0)] && (37)\\
&= \max_{\mu^0\in\mathcal{A}^0} z_j(x,\mu^0)+\lambda\max_{\mu^0\in\mathcal{A}^0\mu^1\in\mathcal{A}} [z_j(x,\mu^1)-z_j(x,\mu^0)] && (38)
\end{aligned}$$

Hence

$$\begin{aligned}
\left|Z_j^\lambda(x)-Z_j^0(x)\right| &= \lambda\left|\max_{\mu^1\in\mathcal{A},\mu^0\in\mathcal{A}^0}[z_j(x,\mu^1)-z_j(x,\mu^0)]\right|\\
&= \lambda\max_{\mu^1\in\mathcal{A},\mu^0\in\mathcal{A}^0}[z_j(x,\mu^1)-z_j(x,\mu^0)] && (39)
\end{aligned}$$

since $\mathcal{A}^0\subseteq\mathcal{A}$. For $\mu^1\in\mathcal{A}$, $\mu^0\in\mathcal{A}^0$, $\omega^0\in\Omega$ we have

$$\begin{aligned}
z_j(x,\mu^1)-z_j(x,\mu^0) &= \int_\Omega \phi_j(x,\omega)d\mu^1-\int_\Omega\phi_j(x,\omega)d\mu^0 && (40)\\
&= \int_\Omega \phi_j(x,\omega)d\mu^1-\int_\Omega\phi_j(x,\omega^0)d\mu^1\\
&\quad +\int_\Omega \phi_j(x,\omega^0)d\mu^0-\int_\Omega\phi_j(x,\omega)d\mu^0\\
&= \int_\Omega(\phi_j(x,\omega)-\phi_j(x,\omega^0))d\mu^1-\int_\Omega(\phi_j(x,\omega)-\phi_j(x,\omega^0))d\mu^0
\end{aligned}$$

Hence

$$\begin{aligned}
|z_j(x,\mu^1)-z_j(x,\mu^0)| &= \left|\int_\Omega(\phi_j(x,\omega)-\phi_j(x,\omega^0))d\mu^1-\int_\Omega(\phi_j(x,\omega)-\phi_j(x,\omega^0))d\mu^0\right|\\
&= \le\left|\int_\Omega(\phi_j(x,\omega)-\phi_j(x,\omega^0))d(\mu^1-\mu^0)\right| \le k_j\int_\Omega d(\omega,\omega^0)d(\mu^1-\mu^0)\\
&= k_j\left(\int_\Omega d(\omega,\omega^0)d\mu^1-\int_\Omega d(\omega,\omega^0)d\mu^0\right)\le k_j d_M(\mu^1,\mu^0)
\end{aligned}$$

We obtain

$$|Z_j^\lambda(x) - Z_j^0(x)| \le \lambda k_j \max_{\mu \in \mathcal{A}} d_M(\mu, \mu^0) \le \lambda k_j D(\mathcal{A})$$

which concludes the proof. □

The next proposition concerns location of the optimal values and estimates the efficiency loss due to ambiguity about the probability distribution.

Proposition 6.3 *Assume that for all $j = \ldots, n$, $\phi_j(x, \omega)$ is convex w.r.t. $x \in X$ and $\mathrm{Im} Z^\lambda$ is $\mathbb{R}^m_+$ -closed for any $\lambda \in [0, 1]$. Assume further $|\phi_j(x, \omega_1) - \phi_j(x, \omega_2)| \le k_j d(\omega_1, \omega_2)$ for every $x \in X$. Then for $k = (k_1, \ldots, k_m)$ the set relation*

$$\mathrm{WMin}(RMSOP^\lambda) \le^l \mathrm{Im}(Z^0) + \lambda D(\mathcal{A})k \tag{41}$$

holds. In particular

$$\mathrm{WMin}(RMSOP^\lambda) \le^l \mathrm{WMin}(RMSOP^0) + \lambda D(\mathcal{A})k \tag{42}$$

Proof By contradiction, if the relation (41) does not hold, there exists $\bar{x} \in X$ such that

$$Z^0(\bar{x}) + \lambda D(\mathcal{A})k \notin \mathrm{WMin}(RMSOP^\lambda) + \mathbb{R}^m_+ \tag{43}$$

From the convexity of $\phi_j(x, \omega)$ w.r.t. x and the domination property for $\mathrm{Im}(Z^\lambda)$ it holds

$$\mathrm{Im}(Z^\lambda) \subseteq \mathrm{Min}(RMSOP^\lambda) + \mathbb{R}^m_+ \subseteq \mathrm{WMin}(RMSOP^\lambda) + \mathbb{R}^m_+ \subseteq \mathrm{Im}(Z^\lambda) + \mathbb{R}^m_+ \tag{44}$$

This shows

$$\mathrm{WMin}(RMSOP^\lambda) + \mathbb{R}^m_+ = \mathrm{Im}(Z^\lambda) + \mathbb{R}^m_+ \tag{45}$$

Convexity of $\phi_j(\cdot, \omega)$ implies that Z^λ is $\mathbb{R}^m_+$- convex. Therefore $\mathrm{Im}(Z^\lambda) + \mathbb{R}^m_+$ is convex. Also, from the $\mathbb{R}^m_+$-closedness of $\mathrm{Im}(Z^\lambda)$ we get that $\mathrm{Im}(Z^\lambda) + \mathbb{R}^m_+$ is closed. By using the strong separation theorem, there exists $\bar{\alpha} = (\bar{\alpha}_1, \ldots, \bar{\alpha}_m) \in \mathbb{R}^m \backslash \{0\}$ and $\beta \in \mathbb{R}$ such that

$$\langle \bar{\alpha}, Z^0(\bar{x}) + \lambda D(\mathcal{A})k \rangle < \beta \le \langle \bar{\alpha}, y \rangle \tag{46}$$

for all $y \in \mathrm{WMin}(RMSOP^\lambda) + \mathbb{R}^m_+$. It is easy to show that $\bar{\alpha} \in \mathbb{R}^m_+$. Also, by using the domination property for $\mathrm{Im}(Z^\lambda)$ we have

$$\mathrm{Im}(Z^\lambda) \subseteq \mathrm{Min}(RMSOP^\lambda) + \mathbb{R}^m_+ \tag{47}$$

Putting $y = Z^\lambda(\bar{x})$ we get

$$\beta \le \langle \bar{\alpha}, Z^\lambda(\bar{x}) \rangle = \sum_{j=1}^m \bar{\alpha}_j \sup_{\mu \in \mathcal{A}} z_j((\bar{x}, (1 - \lambda)\mu^0 + \lambda\mu)) \tag{48}$$

$$= \sum_{j=1}^m \bar{\alpha}_j \sup_{\mu \in \mathcal{A}} \left[(1 - \lambda) z_j(\bar{x}, \mu^0) + \lambda z_j(\bar{x}, \mu)\right] \tag{49}$$

$$= \sum_{j=1}^m \bar{\alpha}_j \sup_{\mu \in \mathcal{A}} \left(z_j(\bar{x}, \mu^0) + \lambda(z_j(\bar{x}, \mu) - z_j(\bar{x}, \mu^0))\right] \tag{50}$$

$$= \sum_{j=1}^m \bar{\alpha}_j \left[Z_j^0 + \lambda \sup_{\mu \in \mathcal{A}} ((z_j(\bar{x}, \mu) - z_j(\bar{x}, \mu^0)) \right] \le \langle \bar{\alpha}, Z^0(\bar{x}) + \lambda D(\mathcal{A})k \rangle < \beta \tag{51}$$

This is a contradiction and hence the set relation (41) holds. □

It is worth mentioning a counterpart of Proposition 6.3 in the case Ω is finite, i.e. $\Omega = \{\omega_1, \dots, \omega_m\}$ and, In this case, let $\mathcal{A}$ and $\mathcal{A}^0$ be two sets of probability distributions over Ω, $P = (p_1, \dots p_m)$. Assume $\phi_j(x, \omega)$ are continuous with respect to x and let

$$\tilde{k}_j = \max_{x \in X,\ \omega^1, \omega^2 \in \Omega} \left| \phi_j(x, \omega^1) - \phi_j(x, \omega^2) \right| \tag{52}$$

and $\tilde{k} = (\tilde{k}_1, \dots, \tilde{k}_n)$.

Proposition 6.4 *If $\Omega = (\omega_1, \dots, \omega_m)$ and $\phi_j(x, \omega)$ are continuous and convex with respect to $x \in X$ and $\mathrm{Im} Z^\lambda$ is $\mathbb{R}^m_+$-closed for any $\lambda \in [0, 1]$. then it holds*

$$\mathrm{WMin}(RMSOP^\lambda) \leq^l \mathrm{WMin}(RMSOP^0) + \lambda \tilde{D}(\mathcal{A})\tilde{k} \tag{53}$$

with $\tilde{D}(\mathcal{A}) = \max_{P, Q \in \mathcal{A}} \sum_{i=1}^m |p_i - q_i|$.

The next result concerns stability of the optimal solutions. We need the following definition (see e.g. Li and Xu (2010)).

Definition 6.1 Let $f : X \to \mathbb{R}$. We say that $x^0 \in X$ is an isolated minimizer of order $\alpha > 0$ and constant $h > 0$ when for every $x \in X$ it holds

$$f(x) - f(x^0) \geq h \|x - x^0\|^\alpha \tag{54}$$

Theorem 6.2 *Let $\mu = (\mu_1 \dots, \mu_m) \in \mathcal{A}^m$, $\mu^0 = (\mu^0_1, \dots, \mu^0_m)$ and let*

$$L_0(x, \mu) = \sum_{j=1}^m \beta_j z_j(x, \mu_j) \tag{55}$$

with $\beta_j \in (0, 1]$, $j = 1, \dots, m$. Assume $x^0 \in X$ is an isolated minimizer of order α and constant h for function $l(x, \mu^0)$. Then if x^λ is a minimizer of

$$L_\lambda(x) = \sum_{j=1}^m \beta_j \max_{\mu_j \in \mathcal{A}_\lambda} z_j(x, \mu_j) \tag{56}$$

it holds

$$\|x^\lambda - x^0\| \leq \left(\frac{2\lambda}{h} \right)^{1/\alpha} D(\mathcal{A})^{1/\alpha} \tag{57}$$

Proof We have $L_0(x\mu^0) - L_0(x^0, \mu^0) \geq h\|x - x^0\|^\alpha$. Let $x \in S_\lambda$. Following the argument of Theorem 5.1, there exist $\mu_j \in \mathcal{A}_\lambda$ such that x minimizes function

$$L(x, \mu) = \sum_{j=1}^m \beta_j z_j(x, \mu_j) \tag{58}$$

It holds

$$L(x^0, \mu) - L(x, \mu) = L(x^0, \mu^0) - L(x, \mu^0) + w \tag{59}$$

where

$$w = \left[L(x^0, \mu) - L(x^0, \mu^0) \right] + \left[L(x, \mu^0) - L(x, \mu) \right] \tag{60}$$

We have

$$
\begin{aligned}
|w| &\leq |L(x^0, \mu) - L(x^0, \mu^0)| + |L(x, \mu^0) - L(x, \mu)| && (61)\\
&\leq \sum_{j=1}^{m} \beta_j |z_j(x, \mu_j) - z_j(x, \mu_j^0)| + \sum_{j=1}^{m} \beta_j |z_j(x, \mu_j^0) - z_j(x, \mu_j)| && (62)\\
&\leq 2\lambda \sum_{j=1}^{m} \beta_j D(\mathcal{A}) \leq 2\lambda D(\mathcal{A}) && (63)
\end{aligned}
$$

We claim that

$$L(x, \mu^0) - L(x^0, \mu^0) \leq |w| \tag{64}$$

Indeed, suppose to the contrary that $L(x, \mu^0) - L(x^0, \mu^0) - |w| > 0$. If $w = 0$, then

$$L(x^0, \mu) - L(x, \mu) > 0 \tag{65}$$

which contradicts to $x^0 \in S_\lambda$. If $w \neq 0$ then

$$L(x, \mu) - L(x^0, \mu) = L(x, \mu^0) - L(x^0, \mu^0) - w > 0 \tag{66}$$

which again contradicts to $x^0 \in S_\lambda$. Observe now that we have

$$h\|x - x^0\|^\alpha \leq L(x, \mu^0) - L(x^0, \mu^0) \tag{67}$$

and hence

$$h\|x - x^0\|^\alpha \leq 2\lambda D(\mathcal{A}) \tag{68}$$

So it holds

$$\|x - x^0\| \leq \left(\frac{2\lambda}{h}\right)^{1/\alpha} D(\mathcal{A})^{1/\alpha} \tag{69}$$

□

The following corollary follows from the previous result and, therefore, we omit its proof.

Corollary 6.1 *The following statements are true.*

(i) Let us denote by S_λ the set of minimizers of function L_λ (clearly S_0 is the set of minimizers of function $l(x, \mu^0)$). Then we have

$$e(S_\lambda, S_0) \leq \left(\frac{2\lambda}{h}\right)^{1/\alpha} D(\mathcal{A})^{1/\alpha} \tag{70}$$

(ii) Consequently there exists $x(\lambda) \in \mathrm{Eff}(DRMP^\lambda)$ such that

$$d(x(\lambda), \mathrm{Eff}(DRMP^0) \leq \left(\frac{2\lambda}{h}\right)^{1/\alpha} D(\mathcal{A})^{1/\alpha} \tag{71}$$

7 Applications

In this section, we explore the applications of the previous results to vector-valued risk measures, which are extensively discussed in the extant literature (see (Jouini et al. 2004 and Corazza et al. 2024)). Our focus lies on analyzing the multiobjective optimization problem

$$\min_{x \in X} \mathcal{R}(\psi(x)) \tag{72}$$

where

$$X = \{x \in \mathbb{R}^n : x_i \geq 0,\ i = 1, \dots, n;\ \sum_{i=1}^{n} x_i \leq 1\}$$

In the following we denote by $\mathcal{P}_i(x)$ the following set:

$$\mathcal{P}_i(x) = \{\bar{P}_i \in \mathcal{P}_i : \sup_{P_i \in \mathcal{P}_i} \mathbb{E}_{P_i \in \mathcal{P}_i}[-\psi_i(x)] = \mathbb{E}_{\bar{P}_i}[-\psi_i(x)]\} \tag{73}$$

Theorems 5.1 and 6.1 entail the following results.

Proposition 7.1 *Let $\psi = (\psi_1, \dots, \psi_d)$ and assume ψ_i is concave $i = 1, \dots, d$. Then $\bar{x} \in X$ is a weakly risk minimizing portfolio if and only if there exist a nonzero vector $\lambda \in \mathbb{R}^d_+$ and probability distributions $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that $\bar{x}$ minimizes function*

$$\sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i}[-\psi_i(x)] \tag{74}$$

Proposition 7.2 *Let $N_X(\bar{x})$ the normal cone to X at $\bar{x} \in X$. If $\psi_i(x)$ are concave and differentiable w.r.t. x, then $\bar{x}$ is a weakly risk minimizing portfolio if and only if there exist λ_i, $i = 1, \dots, d$, $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that*

$$\sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i} \nabla \psi_i(\bar{x}) \in N_X(\bar{x}) \tag{75}$$

that is if and only if $\bar{x}$ solves the variational inequality

$$\left\langle \sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i} \nabla \psi_i(\bar{x}), x - \bar{x} \right\rangle \geq 0, \forall x \in X \tag{76}$$

(here $\mathbb{E}_{\bar{P}_i} \nabla \psi_i(\bar{x})$ is a vector of expected values).

Observe that by Theorem 6.46 in Rockafellar and Wets (1998), we get

$$N_X(\bar{x}) = \left\{z \in \mathbb{R}^n : z = \alpha \mathbf{1} + (y_1, \dots, y_n), \alpha \geq 0, y_i \leq 0,\ y_i = 0 \text{ if } \bar{x}_i > 0\right\} \tag{77}$$

Condition (82) entails that $\bar{x} \in X$ is a weakly risk minimizing portfolio if and only if there exist a nonzero vector $\lambda \in \mathbb{R}^d_+$ and probability distributions $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that

$$\sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i} \frac{\partial \psi_i}{\partial x_{j_1}}(\bar{x}) = \sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i} \frac{\partial \psi_i}{\partial x_{j_m}}(\bar{x}) = y \tag{78}$$

for $j_1, \ldots, j_m \in I = \{i = 1, \ldots, n : x_i > 0\}$ and

$$\sum_{i=1}^{d} \lambda_i \mathbb{E}_{\bar{P}_i} \frac{\partial \psi_i}{\partial x_{k_s}}(\bar{x}) \leq y \tag{79}$$

for $k_s \in I_1 = \{i = 1, \ldots, n : x_i = 0\}$. Condition (78) states that at the optimal solution $\bar{x}$ the weighted expected marginal contribution to the portfolio aggregation must be equal for each asset entering the optimal solution. According to Condition (79) assets with a lower weighted expected marginal contribution to the portfolio aggregation will not enter the optimal solution.

Now we consider the case of a linear portfolio aggregator. Let A be a matrix with n columns and d rows and let

$$\psi(x) = AX(x) \tag{80}$$

where $X(x) = (x_1 X_1, \ldots, x_d X_d)^T$, $x = (x_1, \ldots, x_n) \in X$. In this case Theorem 7.1 entails that $\bar{x} \in X$ is a weakly risk minimizing portfolio if and only if there exist a nonzero vector $\lambda \in \mathbb{R}^d_+$ and probability distributions $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that $\bar{x}$ minimizes over X the linear function

$$\sum_{i=1}^{d} \lambda_i \langle A_i, X(x) \rangle \tag{81}$$

where A_i denotes the i-th row of A. Theorem 7.2 entails the existence of a nonzero vector $\lambda \in \mathbb{R}^d_+$ and $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that

$$\sum_{i=1}^{d} \lambda_i \langle A_i, \mu(\bar{P}_i) \rangle \in N_X(\bar{x}) \tag{82}$$

where $\mu(\bar{P}_i) = (\mu_1(\bar{P}_i), \ldots, \mu_d(\bar{P}_i))^T$ with $\mu_i(\bar{P}_i) = \mathbb{E}_{\bar{P}_i}(X_i)$.

Condition (82) entails that $\bar{x} \in X$ with for every i is a weakly risk minimizing portfolio if and only if there exist a nonzero vector $\lambda \in \mathbb{R}^d_+$ and probability distributions $\bar{P}_i \in \mathcal{P}_i(\bar{x})$ such that

$$\sum_{i=1}^{d} \lambda_i a_{ij_1} \mu_{j_1}(\bar{P}_i) = \sum_{i=1}^{d} \lambda_i a_{ij_m} \mu_{j_m}(\bar{P}_i) = y \tag{83}$$

for $j_1, \ldots, j_m \in I = \{i = 1, \ldots, n : x_i > 0\}$ and

$$\sum_{i=1}^{d} \lambda_i a_{ik_s} \mu_{k_1}(\bar{P}_i) \leq y \tag{84}$$

for $k_s \in I_1 = \{i = 1, \ldots, n : x_i = 0\}$, $s = 1, \ldots, n - m$. Condition (83) states that at the optimal solution $\bar{x}$ the weighted expected contribution to the portfolio aggregation must be equal for each asset entering the optimal solution. According to Condition (84) assets with a lower weighted expected contribution to the portfolio aggregation will not enter the optimal solution.

Finally, we consider the case of separate linear aggregation, i.e. when $a_{ij} \neq 0 \Rightarrow a_{sj} = 0, \forall s \neq i$. In this case conditions (83) and (84) become

$$\lambda_{j_1} a_{ij_1} \mu_{ij_1}(\bar{P}_i) = \lambda_{j_m} a_{ij_m} \mu_{ij_m}(\bar{P}_i) = y \tag{85}$$

for i and j_s such that $a_{ij_s} \neq 0$ and $\bar{x}_{j_s} > 0$ and

$$\lambda_{j_1} a_{ij_1} \mu_{ij_1}(\bar{P}_i) \leq y \tag{86}$$

for for i and j_s such that $a_{ij_s} \neq 0$ and $\bar{x}_{j_s} = 0$ Condition (85) states the existence of probability distributions such that at the optimal solution $\bar{x}$ the expected return of each asset entering the optimal solution is equal.

We now consider application of Proposition 6.3 to the optimization of vector risk measures. We consider the particular case of the vector CVaR, defined as follows Jouini et al. (2004). Associated to the vector of significance levels $\alpha = (\alpha_1, \ldots, \alpha_n)$, i.e., we use the composite risk measure defined as

$$CVaR_\alpha(\psi(\cdot)) = \begin{bmatrix} CVaR_{\alpha_1}(\psi_1(\cdot)) \\ \vdots \\ CVaR_{\alpha_n}(\psi_n(\cdot)) \end{bmatrix}.$$

It is already known that each component of the vector-valued CVAR can be written as

$$CVAR_{\alpha_i}(\psi(\cdot)) = \sup_{P \in \mathcal{P}_{\alpha_i}} \mathbb{E}_P[-\psi(\cdot)] \tag{87}$$

with

$$\mathcal{P}_{\alpha_i} := \left\{ \mathbf{q}^i = (q_1^i, \ldots, q_k^i) \in \mathbb{R}_+^n : \frac{q_j^i}{p_j} \leq \frac{1}{\alpha_i}, \sum_{j=1}^k q_j^i = 1 \right\}. \tag{88}$$

Let now $\beta = (\beta_\mathbf{1}, \ldots, \beta_\mathbf{d})$, $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $\alpha_i \leq \beta_i$ for every i and $\alpha_i < \beta_i$ for at least one i. Then $\mathcal{P}_{\alpha_i} \supseteq \mathcal{P}_{\beta_i}$ and it is easily seen that we have $\mathcal{P}_{(1-\lambda)\alpha_i + \lambda\beta_i} = (1-\lambda)\mathcal{P}_{\alpha_i} + \lambda\mathcal{P}_{\beta_i}$. Proposition 6.4 gives the following result. We denote by WMin_α the weakly efficient frontier for the problem of minimizing CVaR_α.

Proposition 7.3 *For $\beta = (\beta_\mathbf{1}, \ldots, \beta_\mathbf{d})$, $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $\alpha_i \leq \beta_i$ for every i and $\alpha_i < \beta_i$ for at least one i. Then it holds*

$$\text{WMin}_{(1-\lambda)\alpha+\lambda\beta} \leq^l \text{WMin}_\alpha + \lambda\mathcal{D}(\tilde{\mathcal{P}}_\alpha)\tilde{k} \tag{89}$$

and $\tilde{\mathcal{P}}_\alpha = (\mathcal{P}_{\alpha_1} \times, \cdots, \times\mathcal{P}_{\alpha_d})$ and $\tilde{k} = (\tilde{k}_1, \ldots, \tilde{k}_d)$, with $\tilde{k}_i$ defined by (52).

Remark 7.1 Proposition 7.3 estimates the efficiency loss due to a lower significance level.

7.1 Computational considerations and implementability

In this section we provide some final remarks and briefly discuss some computational issues related to the solution of the minimization problem Eq. (72). First, it is important to recognize that the mean of a random variable can be estimated by applying the strong law of large numbers. This offers a straightforward way to approximate the objective function mentioned in Eq. (72). Next, it is important to point out that the objective function outlined in Eq. (72) is convex thanks to the hypotheses on the functions ψ_i. By leveraging the convexity of the objective function in Eq. (72), various optimization algorithms can be employed effectively. These algorithms can efficiently converge towards the optimal solution without encountering issues such as getting trapped in local minima. Some of these algorithms include gradient descent, Newton's method, and convex programming techniques. These algorithms utilize the convex nature of the objective function to iteratively approach the optimal solution in a reliable and efficient manner.

However, when the convexity property of the functions ψ_i is no longer assumed, the minimization problem Eq. (72) is non-convex and it requires the implementation of altenative minimization algorithms. These include, for instance, metaheuristic approaches. Metaheuristics are a class of optimization algorithms that are inspired by natural phenomena or behavioral patterns. They rely on randomized search and learning strategies to efficiently explore the search space and find good, if not optimal, solutions. Metaheuristics can be particularly useful to solve complex, non-convex optimization problems that are difficult to solve using traditional mathematical programming techniques. Examples include genetic algorithms, simulated annealing, tabu search, and swarm intelligence algorithms like particle swarm optimization. Metaheuristics are often used in a wide range of applications, including scheduling, routing, design optimization, and machine learning.

8 Discussion

The findings presented in this paper make several contributions to the field of disributionally robust multi-objective optimization.

First, by establishing scalarization results and optimality conditions for the convex case, we have strengthened the theoretical foundations. These analytical insights can help researchers and practitioners better understand the properties and behavior of distributionally robust multi-objective optimization models. The connection between robustness and solving a min-max problem across feasible distributions further elucidates the underlying mechanisms at play.

A key result of our work is the characterization of the "loss of efficiency" phenomenon when transitioning from nominal to robust optimization. We have shown that robust optimal values are less than or equal to the nominal optimal values, and that robust solutions represent ϵ-optimal solutions of the original problem. This understanding can help decision-makers navigate the tradeoffs between optimality and robustness when applying these techniques, particularly in financial portfolio optimization and risk management applications.

Furthermore, our approach to estimating the efficient frontiers of both the nominal and robust problems, along with the associated efficiency loss, provides a valuable analytical framework. By leveraging appropriate nonlinear scalarizations and set distances, contingent on the uncertainty set configuration, practitioners can quantify the impacts of robustness on the Pareto-optimal solutions. The sensitivity analysis with respect to variations in the uncertainty set offers additional insights that can inform the selection of appropriate robustness parameters.

We also demonstrate the applications of the theory of distributionally robust multiobjective optimization to the field of risk theory. Risk measures are critical tools used across various industries, including finance, economics, engineering, and insurance. These measures are employed to quantify and manage the uncertainty and potential losses associated with different activities or investments.

While this paper presents significant advancements, there are also several limitations that warrant further investigation. For instance, the convex case examined may not capture the full complexity of real-world problems, which often exhibit nonlinear, non-convex, or discrete characteristics. Expanding the analysis to these more general settings would be an important direction for future research. Additionally, the integration of distributionally robust optimization with other multi-objective techniques, such as metaheuristics, could yield promising avenues for enhancing the applicability of these methods.

9 Conclusion

In this paper, we have introduced the concept of Distributionally Robust Multiobjective Optimization and emphasized its significance in addressing uncertainty within probability distributions, especially in situations where distributions lack clarity or definition. Additionally, we have presented scalarization results and optimality conditions, particularly focusing on the convex case, which lays the foundation for understanding solution structures in uncertain environments.

The paper discusses the estimation of efficient frontiers for both nominal and robust problems, providing insights into associated efficiency loss measures. We conduct sensitivity analysis of robust solutions concerning variations in the uncertainty set, identifying the Monge-Kantorovich distance as a key metric for quantifying efficiency loss under uncertainty.

The practical applicability of our approach in portfolio optimization and risk management is underscored, highlighting its relevance in real-world decision-making contexts. Overall, the methodologies and findings presented contribute significantly to the advancement of optimization theory and provide valuable insights for decision-makers grappling with uncertain portfolio management and risk assessment scenarios.

Decalarations

Funding This study has not been funded.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Conflict of interest Both authors declare that they have no conflict of interest.

References

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance, 26*(7), 1487–1503.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203–228.

Ben Abdelaziz, F. (2012). Solution approaches for the multiobjective stochastic programming. *European Journal of Operations Research, 216*, 1–16.

Ben-tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton: Princeton University Press.

Caballero, R., Cerda, E., Munos, M. M., & Rey, L. (2004). Stochastic approach versus multiobjective approach for obtaining efficient solutions in stochastic multiobjective programming problems. *European Journal of Operational Research, 158*, 633–648.

Caballero, R., Cerda, E., Munos, M. M., Rey, L., & Stancu-Minasian, I. M. (2001). Efficient solution concepts and their relations in stochastic multiobjective programming. *Journal of Optimization Theory and Applications, 110*, 53–74.

Cascos, I., & Molchanov, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and Stochastics, 11*(3), 373–397.

Coello Coello, C. A., Lamont, G. B., & Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems*. Cham: Springer.

Corazza, M., Garcia, R., Khan, F., La Torre, D., & Masri, H. (2024). Artificial intelligence and beyond for finance. In *Transformations in banking, finance and regulation* (Vol. 15). World Scientific.
Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197.
Ehrgott, M. (2005). *Multicriteria optimization* (Vol. 491). Berlin: Springer.
Geoffrion, A. M. (1968). Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications, 22*, 618–630.
Gutjahr, W. J., & Pichler, A. (2016). Stochastic multi-objective optimization: A survey on non-scalarizing methods. *Annals of Operation Research, 236*, 475–499.
Jorion, P. (2006). *Value at risk: The new benchmark for managing financial risk* (3rd ed.). New York: McGraw-Hill Education.
Jouini, E., Meddeb, M., & Touzi, N. (2004). Vector-valued coherent risk measures. *Finance and Stochastics, 8*, 531–552.
Kuroiwa, D. (2001). On set-valued optimization. *Nonlinear Analysis: Theory, Methods and Applications, 47*, 1395–1400.
La Torre, D., & Mendivil, F. (2018). Portfolio optimization under partial uncertainty and incomplete information: A probability multimeasure-based approach. *Annals of Operation Research, 267*, 267–279. https://doi.org/10.1007/s10479-016-2298-x
La Torre, D., & Mendivil, F. (2022). Stochastic efficiency and inefficiency in portfolio optimization with incomplete information: A set-valued probability approach. *Annals of Operation Research, 311*, 1085–1098. https://doi.org/10.1007/s10479-020-03886-0
La Torre, D., Mendivil, F., & Rocca, M. (2021). Modeling portfolio efficiency using stochastic optimization with incomplete information and partial uncertainty. *Annals of Operation Research*. https://doi.org/10.1007/s10479-021-04372-x
Li, S. J., & Xu, S. (2010). Sufficient conditions of isolated minimizers for constrained programming problems. *Numerical Functional Analysis and Optimization, 31*, 715–727.
Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal, 58*(4), 36–52.
Luc, D. T. (1989). *Theory of vector optimization*. Berlin: Springer.
Mastrogiaocomo, E., & Rocca, M. (2021). Set optimization of set-valued risk measures. *Annals of Operations Research, 296*, 291–314.
McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques, and tools*. Princeton: Princeton University Press.
Rockafellar, R. T., & Wets, R.J.-B. (1998). *Variational analysis*. Heidelberg: Springer.
Sawaragi, Y., Hirotaka, N., & Tetsuzo, T. (Eds.). (1985). *Theory of multiobjective optimization*. Amsterdam: Elsevier.
Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. Philadelphia: SIAM.
Shapiro, A., & Kleywegt, A. (2010). Minimax analysis of stochastic problems. *Optimization Methods and Software, 17*, 523–542.
Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management, 21*(1), 49–58.