

# Multi-objective risk-averse two-stage stochastic programming problems

Çağın Ararat\*<sup>†</sup>      Özlem Çavuş\*<sup>†</sup>      Ali İrfan Mahmutoğulları\*

November 15, 2017

## Abstract

We consider a multi-objective risk-averse two-stage stochastic programming problem with a multivariate convex risk measure. We suggest a convex vector optimization formulation with set-valued constraints and propose an extended version of Benson’s algorithm to solve this problem. Using Lagrangian duality, we develop scenario-wise decomposition methods to solve the two scalarization problems appearing in Benson’s algorithm. Then, we propose a procedure to recover the primal solutions of these scalarization problems from the solutions of their Lagrangian dual problems. Finally, we test our algorithms on a multi-asset portfolio optimization problem under transaction costs.

**Keywords and phrases:** multivariate risk measure, multi-objective risk-averse two-stage stochastic programming, risk-averse scalarization problems, convex Benson algorithm, nonsmooth optimization, bundle method, scenario-wise decomposition

**Mathematics Subject Classification (2010):** 49M27, 90C15, 90C25, 90C29, 91B30.

## 1 Introduction

We consider a *multi-objective risk-averse two-stage stochastic programming problem* of the general form

$$\begin{aligned} \min \quad & z \text{ w.r.t. } \mathbb{R}_+^J \\ \text{s.t.} \quad & z \in R(Cx + Qy) \\ & (x, y) \in \mathcal{X}, z \in \mathbb{R}^J. \end{aligned}$$

In this formulation,  $x$  is the *first-stage* decision variable,  $y$  is the *second-stage* decision variable and  $\mathcal{X}$  is a compact finite-dimensional set defined by linear constraints.  $C, Q$  are cost parameters which are matrices of appropriate dimension. We assume that  $C$  is deterministic and  $Q$  is random.  $R(\cdot)$  is a *multivariate convex risk measure*, which is a set-valued mapping from the space of  $J$ -dimensional random vectors into the power set of  $\mathbb{R}^J$  (see Hamel and Heyde (2010)). In other words,  $R(Cx + Qy)$  is the set of deterministic cost vectors  $z \in \mathbb{R}^J$  for which  $Cx + Qy - z$  becomes acceptable in a certain sense.

The above problem is a *vector optimization* problem and solving it is understood as computing the *upper image*  $\mathcal{P}$  of the problem defined by

$$\mathcal{P} = \text{cl} \{ z \in \mathbb{R}^J \mid z \in R(Cx + Qy), (x, y) \in \mathcal{X} \},$$

whose boundary is the so-called *efficient frontier*. Here,  $\text{cl}$  denotes the closure operator. One would be interested in finding a set  $\mathcal{Z}$  of *weakly efficient solutions*  $(x, y, z)$  with  $z \in R(Cx + Qy)$  for some  $(x, y) \in \mathcal{X}$  such that there is no  $z' \in R(Cx' + Qy')$  with  $(x', y') \in \mathcal{X}$  and  $z' < z$ . Here, “ $<$ ” denotes the componentwise strict order in  $\mathbb{R}^J$ . The  $z$  components of these solutions are on the efficient frontier. In addition, the set  $\mathcal{Z}$  is supposed to construct  $\mathcal{P}$  in the sense that

$$\mathcal{P} = \text{cl} \text{co}(\{ z \in \mathbb{R}^J \mid (x, y, z) \in \mathcal{Z} \} + \mathbb{R}_+^J),$$

---

\*Bilkent University, Department of Industrial Engineering, Ankara, Turkey.

<sup>†</sup>Ç. Ararat and Ö. Çavuş contributed equally to this work.

where  $\text{co}$  denotes the convex hull operator. Our aim is to compute  $\mathcal{P}$  approximately using a finite set of weakly efficient solutions.

Algorithms for computing upper images of vector optimization problems are extensively studied in the literature. A seminal contribution in this field is the algorithm for linear vector optimization problems by Benson (1998), which computes the set of all weakly efficient solutions of the problem and works on the outer approximation of the upper image rather than the feasible region itself. Benson's algorithm has been generalized recently in Ehrgott *et al.* (2011) and Löhne *et al.* (2014) for (ordinary) convex vector optimization problems, namely, optimization problems with a vector-valued objective function and a vector-valued constraint that are convex with respect to certain underlying cones, e.g., the positive orthants in the respective dimensions. While the algorithm in Ehrgott *et al.* (2011) relies on the differentiability of the involved functions, Löhne *et al.* (2014) makes no assumption on differentiability and obtains finer approximations of the upper image by making use of the so-called *geometric dual problem*.

In the literature, there is a limited number of studies on multi-objective two-stage stochastic optimization problems. Some examples of these studies are Abbas and Bellahcene (2000), Cardona *et al.* (2011), where the decision maker is risk-neutral, that is, one takes  $R(Cx+Qy) = \mathbb{E}[Cx + Qy] + \mathbb{R}_+^J$ . In principle, multi-objective risk-neutral two-stage stochastic optimization problems with linear constraints and continuous variables can be formulated as linear vector optimization problems and they can be solved using the algorithm in Benson (1998). If the number of scenarios is not too large, then the problem can be solved in reasonable computation time. Otherwise, one should look for an efficient method, generally, based on scenario decompositions.

To the best of our knowledge, for the risk-averse case, there is no study on multi-objective two-stage stochastic programming problems. However, single-objective mean-risk type problems can be seen as scalarizations of two-objective stochastic programming problems (*see, for instance*, Ahmed (2006), Miller and Ruszczyński (2011)). On the other hand, Dentcheva and Wolflagen (2016), Noyan *et al.* (2017) work on single-objective problems with multivariate stochastic ordering constraints. As pointed out in the recent survey Gutjahr and Pichler (2016), there is a need for a general methodology for the formulation and solution of multi-objective risk-averse stochastic problems.

The main contributions of the present study can be summarized as follows:

1. To the best of our knowledge, this is the first study focusing on multi-objective risk-averse two-stage stochastic programming problems in a general setting.
2. We propose a vector optimization formulation for our problem using multivariate convex risk measures. Such risk measures include, but are not limited to, multivariate coherent risk measures and multivariate utility-based risk measures.
3. To solve our problem, we suggest an extended version of the convex Benson algorithm in Löhne *et al.* (2014) that is developed for a convex vector optimization problem with a vector-valued constraint. Different from Löhne *et al.* (2014), we deal with set-valued risk constraints and dualize them using the dual representation of multivariate convex risk measures (*see* Hamel and Heyde (2010)) and the Lagrange duality for set-valued constraints (*see* Borwein (1981)).
4. The convex Benson algorithm in Löhne *et al.* (2014) cannot be used for some multivariate risk measures, specifically, for higher-order nonsmooth risk measures. On the other hand, our method is general and can be used for any risk measures for which subgradients can be calculated. An example of such risk measures is higher-order mean semideviation (*see* Shapiro *et al.* (2009) and the references therein).
5. Two risk-averse two-stage stochastic scalarization problems, namely, the problem of weighted sum scalarization and the problem of scalarization by a reference variable, have to be solved during the procedure of the convex Benson algorithm. As the number of scenarios gets larger, these problems cannot be solved in reasonable computation time. Therefore, based on Lagrangian duality, we propose scenario-wise decomposable dual problems for these scalarization problems and suggest a solution procedure based on the bundle algorithm (*see* Lemaréchal (1978), Ruszczyński (2006) and the references therein).
6. Our scenario-wise decomposition algorithms for the scalarization problems can be embedded into other algorithms using the same type of scalarization problems. See Jahn (2004, Chapter 12) for examples of such algorithms.

7. We propose a procedure to recover the primal solutions of the scalarization problems from the solutions of their Lagrangian dual problems.

The rest of the paper is organized as follows: In Section 2, we provide some preliminary definitions and results for multivariate convex risk measures. In Section 3, we provide the problem formulation and recall the related notions of optimality. Section 4 is devoted to the convex Benson algorithm. The two scalarization problems in this algorithm are treated separately in Section 5. In particular, we propose scenario-wise decomposition algorithms and procedures to recover primal solutions. Computational results are provided in Section 6. Some proofs related to Section 5 are collected in the appendix.

## 2 Multivariate convex risk measures

We work on a finite probability space  $\Omega = \{1, \dots, I\}$  with  $I \geq 2$ . For each  $i \in \Omega$ , let  $p_i > 0$  be the probability of the elementary event  $\{i\}$  so that  $\sum_{i \in \Omega} p_i = 1$ .

Let us introduce the notation for (random) vectors and matrices. Let  $J \geq 1$  be a given integer and  $\mathcal{J} = \{1, \dots, J\}$ .  $\mathbb{R}_+^J$  and  $\mathbb{R}_{++}^J$  denote the set of all elements of the Euclidean space  $\mathbb{R}^J$  whose components are nonnegative and positive, respectively. For  $w = (w^1, \dots, w^J)^\top, z = (z^1, \dots, z^J)^\top \in \mathbb{R}^J$ , their scalar product and Hadamard product are defined as

$$w^\top z = \sum_{j \in \mathcal{J}} w^j z^j \in \mathbb{R}, \quad w \cdot z = (w^1 z^1, \dots, w^J z^J)^\top \in \mathbb{R}^J,$$

respectively. For a set  $\mathcal{Z} \subseteq \mathbb{R}^J$ , its associated indicator function (in the sense of convex analysis) is defined by

$$I_{\mathcal{Z}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{Z}, \\ +\infty & \text{else,} \end{cases}$$

for each  $z \in \mathbb{R}^J$ . We denote by  $\mathbb{L}^J$  the set of all  $J$ -dimensional random cost vectors  $u = (u^1, \dots, u^J)^\top$ , which is clearly isomorphic to the space  $\mathbb{R}^{J \times I}$  of  $J \times I$ -dimensional real matrices. We write  $\mathbb{L} = \mathbb{L}^1$  for  $J = 1$ . For  $u \in \mathbb{L}^J$ , we denote by  $u_i = (u_i^1, \dots, u_i^J)^\top \in \mathbb{R}^J$  its realization at  $i \in \Omega$ , and define the expected value of  $u$  as

$$\mathbb{E}[u] = \sum_{i \in \Omega} p_i u_i \in \mathbb{R}^J.$$

Similarly, given another integer  $N \geq 1$ , we denote by  $\mathbb{L}^{J \times N}$  the set of all  $J \times N$ -dimensional random matrices  $Q$  with realizations  $Q_1, \dots, Q_I$ .

The elements of  $\mathbb{L}^J$  will be used to denote random cost vectors; hence, lower values are preferable. To that end, we introduce  $\mathbb{L}_+^J$ , the set of all elements in  $\mathbb{L}^J$  whose components are nonnegative random variables. Given  $u, v \in \mathbb{L}^J$ , we write  $u \leq v$  if and only if  $u_i^j \leq v_i^j$  for every  $i \in \Omega$  and  $j \in \mathcal{J}$ , that is,  $v \in u + \mathbb{L}_+^J$ . We call a set-valued function  $R: \mathbb{L}^J \rightarrow 2^{\mathbb{R}^J}$  a *multivariate convex risk measure* if it satisfies the following axioms (see Hamel and Heyde (2010)):

(A1) **Monotonicity:**  $u \leq v$  implies  $R(u) \supseteq R(v)$  for every  $u, v \in \mathbb{L}^J$ .

(A2) **Translativity:**  $R(u + z) = R(u) + z$  for every  $u \in \mathbb{L}^J$  and  $z \in \mathbb{R}^J$ .

(A3) **Finiteness:**  $R(u) \notin \{\emptyset, \mathbb{R}^J\}$  for every  $u \in \mathbb{L}^J$ .

(A4) **Convexity:**  $R(\gamma u + (1 - \gamma)v) \supseteq \gamma R(u) + (1 - \gamma)R(v)$  for every  $u, v \in \mathbb{L}^J$ ,  $\gamma \in (0, 1)$ .

(A5) **Closedness:** The *acceptance set*  $\mathcal{A} := \{u \in \mathbb{L}^J \mid 0 \in R(u)\}$  of  $R$  is a closed set.

A multivariate convex risk measure  $R$  is called *coherent* if it also satisfies the following axiom:

(A6) **Positive homogeneity:**  $R(\gamma u) = \gamma R(u)$  for every  $u \in \mathbb{L}^J$ ,  $\gamma > 0$ .

**Remark 2.1.** It is easy to check that the values of a multivariate convex risk measure  $R$  are in the collection of all *closed convex upper subsets* of  $\mathbb{R}^J$ , that is,

$$\mathcal{G} = \{E \subseteq \mathbb{R}^J \mid E = \text{cl co}(E + \mathbb{R}_+^J)\},$$

where  $\text{cl}$  and  $\text{co}$  denote the closure and convex hull operators, respectively. In other words, for every  $u \in \mathbb{L}^J$ , the set  $R(u)$  is a closed convex set with the property  $R(u) = R(u) + \mathbb{R}_+^J$ . The collection  $\mathcal{G}$ , when equipped with the superset relation  $\supseteq$ , is a complete lattice in the sense that every nonempty subset  $\mathcal{E}$  of  $\mathcal{G}$  has an infimum (and also a supremum) which is uniquely given by  $\inf \mathcal{E} = \text{cl co} \bigcup_{E \in \mathcal{E}} E$  as an element of  $\mathcal{G}$  (see Example 2.13 in Hamel *et al.* (2016)). The complete lattice property of  $\mathcal{G}$  makes it possible to study optimization problems with  $\mathcal{G}$ -valued objective functions and constraints, as will also be crucial in the approach of the present paper.

A multivariate convex risk measure  $R$  can be represented in terms of vectors  $\mu$  of probability measures and weight vectors  $w$  in the cone  $\mathbb{R}_+^J \setminus \{0\}$ , which is called its *dual representation*. To state this representation, we provide the following definitions and notation.

Let  $\mathbb{M}_1^J$  be the set of all  $J$ -dimensional vectors  $\mu = (\mu^1, \dots, \mu^J)$  of probability measures on  $\Omega$ , that is, for each  $j \in \mathcal{J}$ , the probability measure  $\mu^j$  assigns the probability  $\mu_i^j$  to the elementary event  $\{i\}$  for  $i \in \Omega$ . For  $\mu \in \mathbb{M}_1^J$  and  $i \in \Omega$ , we also write  $\mu_i := (\mu_i^1, \dots, \mu_i^J)^\top \in \mathbb{R}^J$ . Finally, for  $\mu \in \mathbb{M}_1^J$  and  $u \in \mathbb{L}^J$ , we define the expectation of  $u$  under  $\mu$  by

$$\mathbb{E}^\mu [u] = \left( \mathbb{E}^{\mu^1} [u^1], \dots, \mathbb{E}^{\mu^J} [u^J] \right)^\top = \sum_{i \in \Omega} \mu_i \cdot u_i.$$

A multivariate convex risk measure  $R$  has the following dual representation (see Theorem 6.1 in Hamel and Heyde (2010)): for every  $u \in \mathbb{L}^J$ ,

$$\begin{aligned} R(u) &= \bigcap_{\mu \in \mathbb{M}_1^J, w \in \mathbb{R}_+^J \setminus \{0\}} \left( \mathbb{E}^\mu [u] + \{z \in \mathbb{R}^J \mid w^\top z \geq -\beta(\mu, w)\} \right) \\ &= \bigcap_{w \in \mathbb{R}_+^J \setminus \{0\}} \left\{ z \in \mathbb{R}^J \mid w^\top z \geq \sup_{\mu \in \mathbb{M}_1^J} (w^\top \mathbb{E}^\mu [u] - \beta(\mu, w)) \right\}, \end{aligned}$$

where  $\beta$  is the *minimal penalty function* of  $R$  defined by

$$\beta(\mu, w) = \sup_{u \in \mathcal{A}} w^\top \mathbb{E}^\mu [u] = \sup \{w^\top \mathbb{E}^\mu [u] \mid 0 \in R(u), u \in \mathbb{L}^J\}, \quad (2.1)$$

for each  $\mu \in \mathbb{M}_1^J$ ,  $w \in \mathbb{R}_+^J \setminus \{0\}$ . Note that  $\beta(\cdot, w)$  and  $\beta(\mu, \cdot)$  are convex functions as they are suprema of linear functions.

The *scalarization* of  $R$  by a weight vector  $w \in \mathbb{R}_+^J \setminus \{0\}$  is defined as the function

$$u \mapsto \varphi_w(u) := \inf_{z \in R(u)} w^\top z \quad (2.2)$$

on  $\mathbb{L}^J$ . As an immediate consequence of the dual representation of  $R$ , we also obtain a dual representation for its scalarization:

$$\varphi_w(u) = \sup_{\mu \in \mathbb{M}_1^J} (w^\top \mathbb{E}^\mu [u] - \beta(\mu, w)). \quad (2.3)$$

Some examples of multivariate coherent and convex risk measures are the *multivariate conditional value-at-risk* (multivariate CVaR) and the *multivariate entropic risk measure*, respectively.

**Example 2.2** (Multivariate CVaR). Let  $C \subseteq \mathbb{R}^J$  be a polyhedral closed convex cone with  $\mathbb{R}_+^J \subseteq C \neq \mathbb{R}^J$ . The multivariate conditional value-at-risk is defined by

$$R(u) = (\text{CVaR}_{\nu^1}(u^1), \dots, \text{CVaR}_{\nu^J}(u^J))^\top + C, \quad (2.4)$$

where

$$CVaR_{\nu^j}(u^j) = \inf_{z^j \in \mathbb{R}} \left( z^j + \frac{1}{1 - \nu^j} \mathbb{E} [(u^j - z^j)^+] \right),$$

for each  $u \in \mathbb{L}^J$  and  $j \in \mathcal{J}$  (see Definition 2.1 and Remark 2.3 in Hamel *et al.* (2013)). Here,  $\nu^j \in (0, 1)$  is a risk-aversion parameter and  $(x)^+ := \max\{x, 0\}$  for  $x \in \mathbb{R}$ . The minimal penalty function of  $R$  is given by

$$\beta(\mu, w) = \begin{cases} 0 & \text{if } w \in C^+ \text{ and } \frac{\mu_i^j}{p_i} \leq \frac{1}{1 - \nu^j}, \forall i \in \Omega, j \in \mathcal{J}, \\ +\infty & \text{else,} \end{cases}$$

where  $C^+$  is the positive dual cone of  $C$  defined by

$$C^+ = \{w \in \mathbb{R}^J \mid w^\top z \geq 0, \forall z \in C\}.$$

Note that (2.4) is the multivariate extension of the well-known conditional value-at-risk (see Rockafellar and Uryasev (2000), Rockafellar and Uryasev (2002)).

**Example 2.3** (Multivariate entropic risk measure). Consider the vector-valued exponential utility function  $U: \mathbb{R}^J \rightarrow \mathbb{R}^J$  defined by

$$U(x) = (U^1(x^1), \dots, U^J(x^J))^\top,$$

where

$$U^j(x^j) = \frac{1 - e^{\delta^j x^j}}{\delta^j},$$

for each  $x \in \mathbb{R}^J$  and  $j \in \mathcal{J}$ . Here,  $\delta^j > 0$  is a risk-aversion parameter. Note that  $U^j(\cdot)$  is a concave decreasing function. Let  $C \subseteq \mathbb{R}^J$  be a polyhedral closed convex cone with  $\mathbb{R}_+^J \subseteq C \neq \mathbb{R}^J$ . The multivariate entropic risk measure  $R: \mathbb{L}^J \rightarrow 2^{\mathbb{R}^J}$  is defined as

$$R(u) = \{z \in \mathbb{R}^J \mid \mathbb{E}[U(u - z)] \in C\}, \quad (2.5)$$

for each  $u \in \mathbb{L}^J$  (see Section 4.1 in Ararat *et al.* (2017)). Since  $\mathbb{R}_+^J \subseteq C$ , larger values of the expected utility are preferred. Moreover, as each  $U^j(\cdot)$  is a decreasing function,  $z \in R(u)$  implies  $z' \in R(u)$  for every  $z' \geq z$ .

Finally, the minimal penalty function of  $R$  is given by (see Proposition 4.4 in Ararat *et al.* (2017))

$$\beta(\mu, w) = \sum_{j \in \mathcal{J}} \frac{w^j}{\delta^j} (H(\mu^j \| p) - 1 + \log w^j) + \inf_{s \in C^+} \sum_{j \in \mathcal{J}} \frac{1}{\delta^j} (s^j - w^j \log s^j),$$

where  $H(\mu^j \| p)$  is the *relative entropy* of  $\mu^j$  with respect to  $p$  defined by

$$H(\mu^j \| p) = \sum_{i \in \Omega} \mu_i^j \log \left( \frac{\mu_i^j}{p_i} \right).$$

Note that (2.5) is the multivariate extension of the well-known entropic risk measure (see Föllmer and Schied (2002)).

### 3 Problem formulation

We consider a multi-objective risk-averse two-stage stochastic programming problem. The decision variables and the parameters of the problem consist of deterministic and random vectors and matrices of different dimensions. To that end, let us fix some integers  $J, K, L, M, N \geq 1$  and deterministic parameters  $A \in \mathbb{R}^{K \times M}$  and  $b \in \mathbb{R}^K$ . At the first stage, the decision-maker chooses a deterministic vector  $x \in \mathbb{R}_+^M$  with associated cost  $Cx$ , where  $C \in \mathbb{R}^{J \times M}$ . At the second stage, the decision-maker chooses a random vector  $y \in \mathbb{L}_+^N$  based on the first-stage decision  $x \in \mathbb{R}^M$  as well as the random parameters  $W \in \mathbb{L}^{L \times N}, T \in \mathbb{L}^{L \times M}, h \in \mathbb{L}^L$ . The random cost associated with  $y$  is  $Qy \in \mathbb{L}^J$ , where  $Q \in \mathbb{L}^{J \times N}$ .

Given feasible choices of the decision variables  $x \in \mathbb{R}^M$  and  $y \in \mathbb{L}^N$ , the risk associated with the second-stage cost vector  $Qy \in \mathbb{L}^J$  is quantified via a multivariate convex risk measure  $R: \mathbb{L}^J \rightarrow 2^{\mathbb{R}^J}$ . The set  $R(Qy)$  consists of the deterministic cost vectors in  $\mathbb{R}^J$  that can make  $Qy$  acceptable in the following sense:

$$R(Qy) = \{z \in \mathbb{R}^J \mid Qy - z \in \mathcal{A}\},$$

where  $\mathcal{A} = \{u \in \mathbb{L}^J \mid 0 \in R(u)\}$  is the *acceptance set* of the risk measure. Hence,  $R(Qy)$  collects the deterministic cost reductions from  $Qy$  that would yield an acceptable level of risk for the resulting random cost. Together with the deterministic cost vector  $Cx$ , the overall risk associated with  $x$  and  $y$  is given by the set

$$Cx + R(Qy) = \{Cx + z \mid Qy - z \in \mathcal{A}\} = \{Cx + z \mid z \in R(Qy)\} = R(Cx + Qy),$$

where the last equality holds thanks to the translativity property (A2).

Our aim is to calculate the “minimal” vectors  $z \in R(Cx + Qy)$  over all feasible choices of  $x$  and  $y$ . Using *vector optimization*, we formulate our problem as follows:

$$\begin{aligned} \min \quad & z \text{ w.r.t. } \mathbb{R}_+^J & (P_V) \\ \text{s.t.} \quad & z \in R(Cx + Qy) \\ & Ax = b \\ & T_i x + W_i y_i = h_i, \quad \forall i \in \Omega \\ & z \in \mathbb{R}^J, \quad x \in \mathbb{R}_+^M, \quad y_i \in \mathbb{R}_+^N, \quad \forall i \in \Omega. \end{aligned}$$

Let

$$\mathcal{X} := \{(x, y) \in \mathbb{R}_+^M \times \mathbb{L}_+^N \mid Ax = b, T_i x + W_i y_i = h_i, \forall i \in \Omega\}.$$

We assume that  $\mathcal{X}$  is a compact set. Let us denote by  $\mathcal{R}$  the image of the feasible region of  $(P_V)$  under the objective function, that is,

$$\mathcal{R} = \{z \in \mathbb{R}^J \mid z \in R(Cx + Qy), (x, y) \in \mathcal{X}\} = \bigcup_{(x, y) \in \mathcal{X}} R(Cx + Qy).$$

The *upper image* of  $(P_V)$  is defined as the set

$$\mathcal{P} = \text{cl } \mathcal{R} = \text{cl } \bigcup_{(x, y) \in \mathcal{X}} R(Cx + Qy). \quad (3.1)$$

In particular, we have  $\mathcal{P} \in \mathcal{G}$ , that is,  $\mathcal{P}$  is a closed convex upper set; see Remark 2.1.

Finding the “minimal”  $z$  vectors of  $(P_V)$  is understood as computing the boundary of the set  $\mathcal{P}$ . For completeness, we recall the minimality notions for  $(P_V)$ .

**Definition 3.1.** *A point  $(x, y, z) \in \mathcal{X} \times \mathbb{R}^J$  is called a weak minimizer (weakly efficient solution) of  $(P_V)$  if  $z \in R(Cx + Qy)$  and  $z$  is a weakly minimal element of  $\mathcal{R}$ , that is, there exists no  $z' \in \mathcal{R}$  such that  $z \in z' + \mathbb{R}_{++}^J$ .*

**Definition 3.2.** *(Definition 3.2 in Löhne et al. (2014)) A set  $\mathcal{Z} \subseteq \mathcal{X} \times \mathbb{R}^J$  is called a weak solution of  $(P_V)$  if the following conditions are satisfied:*

1. **Infimality:** *it holds  $\text{cl co}(\{z \in \mathbb{R}^J \mid (x, y, z) \in \mathcal{Z}\} + \mathbb{R}_+^J) = \mathcal{P}$ ,*
2. **Minimality:** *each  $(x, y, z) \in \mathcal{Z}$  is a weak minimizer of  $(P_V)$ .*

Ideally, one would be interested in computing a weak solution  $\mathcal{Z}$  of  $(P_V)$ . However, except for some special cases (e.g. when the values of  $R$  and the upper image  $\mathcal{P}$  are polyhedral sets), such  $\mathcal{Z}$  consists of infinitely many feasible points, that is, it is impossible to recover  $\mathcal{P}$  using only finitely many values of  $R$ . Therefore, our aim is to propose algorithms to compute  $\mathcal{P}$  approximately through finitely many feasible points.

**Definition 3.3.** (Definition 3.3 in Löhne et al. (2014)) Let  $\epsilon > 0$ . A nonempty finite set  $\bar{\mathcal{Z}} \subseteq \mathcal{X} \times \mathbb{R}^J$  is called a finite weak  $\epsilon$ -solution of  $(P_V)$  if the following conditions are satisfied:

1.  **$\epsilon$ -Infimality:** it holds  $\text{co}(\{z \in \mathbb{R}^J \mid (x, y, z) \in \bar{\mathcal{Z}}\}) + \mathbb{R}_+^J - \epsilon \mathbf{1} \supseteq \mathcal{P}$ ,
2. **Minimality:** each  $(x, y, z) \in \bar{\mathcal{Z}}$  is a weak minimizer of  $(P_V)$ .

As noted in Löhne et al. (2014), a finite weak  $\epsilon$ -solution  $\bar{\mathcal{Z}}$  provides an outer and an inner approximation of  $\mathcal{P}$  in the sense that

$$\text{co}(\{z \in \mathbb{R}^J \mid (x, y, z) \in \bar{\mathcal{Z}}\}) + \mathbb{R}_+^J - \epsilon \mathbf{1} \supseteq \mathcal{P} \supseteq \text{co}(\{z \in \mathbb{R}^J \mid (x, y, z) \in \bar{\mathcal{Z}}\}) + \mathbb{R}_+^J. \quad (3.2)$$

Let us also introduce the *weighted sum scalarization* problem with weight vector  $w \in \mathbb{R}_+^J \setminus \{0\}$ :

$$\min w^\top z \quad \text{s.t. } z \in R(Cx + Qy), (x, y) \in \mathcal{X}. \quad (P_1(w))$$

Define  $\mathcal{P}_1(w)$  as the optimal value of  $(P_1(w))$ . For the remainder of this section, we provide a discussion on the existence of optimal solutions of  $(P_1(w))$  as well as the relationship between  $(P_1(w))$  and  $(P_V)$ .

**Proposition 3.4.** Let  $w \in \mathbb{R}_+^J \setminus \{0\}$ . Then, there exists an optimal solution  $(x, y, z)$  of  $(P_1(w))$ .

**Proof.** Note that  $\mathcal{P}_1(w) = \inf_{(x, y) \in \mathcal{X}} \varphi_w(Cx + Qy)$ , where  $\varphi_w(\cdot)$  is the scalarization of  $R$  by  $w$  as defined in (2.2). Since  $\varphi_w(\cdot)$  admits the dual representation in (2.3), it is a lower semicontinuous function on  $\mathbb{L}^J$ . Moreover,  $\mathcal{X}$  is a compact set by assumption. By Theorem 2.43 in Aliprantis and Border (2006), it follows that an optimal solution of  $(P_1(w))$  exists.  $\square$

**Remark 3.5.** Note that the feasible region  $\{(x, y, z) \in \mathcal{X} \times \mathbb{R}^J \mid z \in R(Cx + Qy)\}$  of  $(P_V)$  is not compact in general due to the multivariate risk measure  $R$ , which has unbounded values. However, in Löhne et al. (2014), the feasible region of a vector optimization problem is assumed to be compact. Therefore, by assuming only  $\mathcal{X}$  to be compact, Proposition 3.4 generalizes the analogous result in Löhne et al. (2014).

The following proposition is stated in Löhne et al. (2014) without a proof. It can be shown as a direct application of Theorem 5.28 in Jahn (2004).

**Proposition 3.6.** (Proposition 3.4 in Löhne et al. (2014)) Let  $w \in \mathbb{R}_+^J \setminus \{0\}$ . Every optimal solution  $(x, y, z)$  of  $(P_1(w))$  is a weak minimizer of  $(P_V)$ .

Proposition 3.6 implies that, in the weak sense, solving  $(P_V)$  is understood as solving the family  $(P_1(w))_{w \in \mathbb{R}_+^J \setminus \{0\}}$  of weighted sum scalarizations.

## 4 Convex Benson algorithms for $(P_V)$

The convex Benson algorithms have a primal and a dual variant. While the primal approximation algorithm computes a sequence of outer approximations for the upper image  $\mathcal{P}$  in the sense of (3.2), the dual approximation algorithm works on an associated vector maximization problem, called the *geometric dual problem*. To explain the details of these algorithms, we should define the concept of geometric duality as well as a new scalarization problem  $(P_2(v))$ , called the problem of *scalarization by a reference variable*  $v \in \mathbb{R}^J$ .

### 4.1 The problem of scalarization by a reference variable

The problem  $(P_2(v))$  is required to find the minimum step-length to enter the upper image  $\mathcal{P}$  from a point  $v \in \mathbb{R}^J \setminus \mathcal{P}$  along the direction  $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^J$ . It is formulated as

$$\min \alpha \quad \text{s.t. } v + \alpha \mathbf{1} \in R(Cx + Qy), (x, y) \in \mathcal{X}, \alpha \in \mathbb{R}. \quad (P_2(v))$$

Note that  $(P_2(v))$  is a scalar convex optimization problem with a set-valued constraint. We denote by  $\mathcal{P}_2(v)$  the optimal value of  $(P_2(v))$ . We relax the set-valued constraint  $v + \alpha \mathbf{1} \in R(Cx + Qy)$  in a Lagrangian fashion and obtain the following dual problem using the results of Section 3.2 in Borwein (1981):

$$\underset{\gamma \in \mathbb{R}_+^J}{\text{maximize}} \quad \inf_{(x, y) \in \mathcal{X}, \alpha \in \mathbb{R}} \left( \alpha + \inf_{z \in R(Cx + Qy) - v - \alpha \mathbf{1}} \gamma^\top z \right). \quad (LD_2(v))$$

Note that  $(LD_2(v))$  is constructed by rewriting the risk constraint of  $(P_2(v))$  as  $0 \in R(Cx + Qy) - v - \alpha \mathbf{1}$  and calculating the support function of the set  $R(Cx + Qy) - v - \alpha \mathbf{1}$  by the dual variable  $\gamma \in \mathbb{R}_+^J$ . The next proposition states the strong duality relationship between  $(P_2(v))$  and  $(LD_2(v))$ .

**Proposition 4.1.** *(Theorem 19 and Equation (3.23) in Borwein (1981)) Let  $v \in \mathbb{R}^J$ . Then, there exist optimal solutions  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  of  $(P_2(v))$  and  $\gamma_{(v)}$  of  $(LD_2(v))$ , and the optimal values of the two problems coincide.*

Finally, we recall the relationship between  $(P_2(v))$  and  $(P_V)$ . The next proposition is provided without a proof since the proof in Löhne *et al.* (2014) can be directly applied to our case.

**Proposition 4.2.** *(Proposition 4.5 in Löhne et al. (2014)) Let  $v \in \mathbb{R}^J$ . If  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  is an optimal solution of  $(P_2(v))$ , then  $(x_{(v)}, y_{(v)}, v + \alpha_{(v)} \mathbf{1})$  is a weak minimizer of  $(P_V)$ .*

## 4.2 Geometric duality

Let  $\mathcal{W}$  be the unit simplex in  $\mathbb{R}^J$ , that is,

$$\mathcal{W} = \{w \in \mathbb{R}_+^J \mid w^\top \mathbf{1} = 1\}.$$

For each  $j \in \mathcal{J}$ , let  $e_{(j)}$  be the  $j^{\text{th}}$  unit vector in  $\mathbb{R}^J$ , that is, the  $j^{\text{th}}$  entry of  $e_{(j)}$  is one and all other entries are zero.

The geometric dual of problem  $(P_V)$  is defined as the vector maximization problem

$$\begin{aligned} \max \quad & (w^1, \dots, w^{J-1}, \mathcal{P}_1(w))^\top \text{ w.r.t. } K \\ \text{s.t.} \quad & w \in \mathcal{W}, \end{aligned} \tag{D_V}$$

where  $K$  is the so-called *ordering cone* defined as  $K = \{\lambda e_{(j)} \mid \lambda \geq 0\}$ . Similar to the upper image  $\mathcal{P}$  of  $(P_V)$ , we can define the lower image  $\mathcal{D}$  of  $(D_V)$  as

$$\mathcal{D} := \{(w^1, \dots, w^{J-1}, p) \in \mathbb{R}^J \mid w = (w^1, \dots, w^{J-1}, w^J) \in \mathcal{W}, p \leq \mathcal{P}_1(w)\}.$$

**Remark 4.3.** In analogy with Remark 2.1, the lower image  $\mathcal{D}$  is a closed convex  $K$ -lower set, that is,  $\text{cl co}(\mathcal{D} - K) = \mathcal{D}$ .

Next, we state the relationship between  $\mathcal{D}$  and the optimal solutions of  $(P_1(w))$ ,  $(P_2(v))$ ,  $(LD_2(v))$ .

**Proposition 4.4.** *(Proposition 3.5 in Löhne et al. (2014)) Let  $w \in \mathcal{W}$ . If  $(P_1(w))$  has a finite optimal value  $\mathcal{P}_1(w)$ , then  $(w^1, \dots, w^{J-1}, \mathcal{P}_1(w))^\top$  is a boundary point of  $\mathcal{D}$  and it is also a  $K$ -maximal element of  $\mathcal{D}$ , that is, there is no  $d \in \mathcal{D}$  such that  $d^J > \mathcal{P}_1(w)$ .*

**Proposition 4.5.** *(Propositions 4.6, 4.7 in Löhne et al. (2014)) Let  $v \in \mathbb{R}^J$ . If  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  is an optimal solution of  $(P_2(v))$  and  $\gamma_{(v)}$  is an optimal solution of  $(LD_2(v))$ , then  $\gamma_{(v)}$  is a maximizer of  $(D_V)$ , that is,  $(\gamma_{(v)}^1, \dots, \gamma_{(v)}^{J-1}, \mathcal{P}_1(\gamma_{(v)}))^\top$  is a  $K$ -maximal element of the lower image  $\mathcal{D}$ . Moreover,  $\{z \in \mathbb{R}^J \mid \gamma_{(v)}^\top z \geq \gamma_{(v)}^\top (v + \alpha_{(v)} \mathbf{1})\}$  is a supporting halfspace of  $\mathcal{D}$  at the point  $(v + \alpha_{(v)} \mathbf{1})$ .*

**Proposition 4.6.** *Let  $w \in \mathbb{R}_+^J \setminus \{0\}$ . If  $(x_{(w)}, y_{(w)}, z_{(w)})$  is an optimal solution of  $(P_1(w))$ , then  $\{d \in \mathbb{R}^J \mid (z_{(w)}^J - z_{(w)}^1, \dots, z_{(w)}^J - z_{(w)}^{J-1}, 1)^\top d \leq z_{(w)}^J\}$  is a supporting halfspace of  $\mathcal{D}$  at the point  $(w^1, \dots, w^{J-1}, \mathcal{P}_1(w))$ .*

**Proof.** From Proposition 4.4,  $d := (w^1, \dots, w^{J-1}, \mathcal{P}_1(w))$  is a boundary point of  $\mathcal{D}$ . Moreover, it follows that

$$(z_{(w)}^J - z_{(w)}^1, \dots, z_{(w)}^J - z_{(w)}^{J-1}, 1)^\top d = -w^\top z_{(w)} + z_{(w)}^J + \mathcal{P}_1(w) = z^J$$

since  $\mathcal{P}_1(w) = w^\top z_{(w)}$ . Hence, the assertion of the proposition follows.  $\square$

**Proposition 4.7.** *(Proposition 3.10 in Löhne et al. (2014)) Let  $\epsilon > 0$ .*



(a) Let  $\bar{\mathcal{Z}}$  be a finite weak  $\epsilon$ -solution of  $(P_V)$ . Then,

$$\mathcal{P}^{in}(\bar{\mathcal{Z}}) := \text{co}(\{z \in \mathbb{R}^J \mid (x, y, z) \in \bar{\mathcal{Z}}\}) + \mathbb{R}_+^J$$

is an inner approximation of the upper image  $\mathcal{P}$ , that is,  $\mathcal{P}^{in}(\bar{\mathcal{Z}}) \subseteq \mathcal{P}$ . Moreover,

$$\mathcal{D}^{out}(\bar{\mathcal{Z}}) = \left\{ d \in \mathbb{R}^J \mid (z^J - z^1, \dots, z^J - z^{J-1}, 1)^\top d \leq z^J, \forall z \in \bar{\mathcal{Z}} \right\}$$

is an outer approximation of the lower image  $\mathcal{D}$ , that is,  $\mathcal{D} \subseteq \mathcal{D}^{out}(\bar{\mathcal{Z}})$ .

(b) Let  $\bar{\mathcal{W}}$  be a finite  $\epsilon$ -solution of  $(D_V)$ . Then,

$$\mathcal{P}^{in}(\bar{\mathcal{W}}) := \text{co}(\{(w^1, \dots, w^{J-1}, \mathcal{P}_1(w))^\top \mid w \in \bar{\mathcal{W}}\}) - K$$

is an inner approximation of  $\mathcal{P}$ , that is,  $\mathcal{P}^{in}(\bar{\mathcal{W}}) \subseteq \mathcal{P}$ . Moreover,

$$\mathcal{P}^{out}(\bar{\mathcal{W}}) = \{z \in \mathbb{R}^J \mid w^\top z \geq \mathcal{P}_1(w), \forall w \in \bar{\mathcal{W}}\}$$

is an outer approximation of  $\mathcal{P}$ , that is,  $\mathcal{P} \subseteq \mathcal{P}^{out}(\bar{\mathcal{W}})$ .

The problems  $(P_1(w))$ ,  $(P_2(v))$  and the above propositions form a basis for the primal and dual convex Benson algorithms. These algorithms are explained briefly in the following sections.

### 4.3 Primal algorithm

The primal algorithm starts with an initial outer approximation  $\mathcal{P}^0$  for the upper image  $\mathcal{P}$ . To construct  $\mathcal{P}^0$ , for each  $j \in \mathcal{J}$ , the algorithm computes the supporting halfspace of  $\mathcal{P}$  with direction vector  $e_{(j)}$  by solving the weighted-sum scalarization problem  $(P_1(e_{(j)}))$ . If  $(x_{(j)}, y_{(j)}, z_{(j)})$  is an optimal solution of  $(P_1(e_{(j)}))$ , then this halfspace supports the upper image  $\mathcal{P}$  at the point  $z_{(j)}$ . Then,  $\mathcal{P}^0$  is defined as the intersection of these  $J$  supporting halfspaces.

The algorithm iteratively obtains a sequence  $\mathcal{P}^0 \supseteq \mathcal{P}^1 \supseteq \mathcal{P}^2 \supseteq \dots \supseteq \mathcal{P}$  of finer outer approximations, it updates a set  $\bar{\mathcal{Z}}$  and  $\bar{\mathcal{W}}$  of weak minimizers and maximizers for  $(P_V)$  and  $(D_V)$ , respectively. At iteration  $k$ , the algorithm first computes  $\mathcal{V}^k$ , that is the set of all vertices of  $\mathcal{P}^k$ . For each vertex  $v \in \mathcal{V}^k$ , an optimal solution  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  to  $(P_2(v))$  is computed. The optimal  $\alpha_{(v)}$  is the minimum step-length required to find a boundary point  $(v + \alpha_{(v)}\mathbf{1})$  of  $\mathcal{P}$ . Since the triplet  $(x_{(v)}, y_{(v)}, v + \alpha_{(v)}\mathbf{1})$  is a weak minimizer of  $(P_V)$  by Proposition 4.2, it is added to the set  $\bar{\mathcal{Z}}$ . Then, an optimal solution  $\gamma_{(v)}$  of the dual problem  $(LD_2(v))$  is computed, which is a maximizer for  $(D_V)$  (see Proposition 4.5) and is added to the set  $\bar{\mathcal{W}}$ . This procedure is continued until a vertex  $v$  with a step-length greater than an error parameter  $\epsilon > 0$  is detected. For such  $v$ , using Proposition 4.5, a supporting halfspace of  $\mathcal{P}$  at point  $(v + \alpha_{(v)}\mathbf{1})$  is obtained. The outer approximation is updated as  $\mathcal{P}^{k+1}$  by intersecting  $\mathcal{P}^k$  with this supporting halfspace. The algorithm terminates when all the vertices are in  $\epsilon$ -distance to the upper image  $\mathcal{P}$ .

At the termination, the algorithm computes inner and outer approximations  $\mathcal{P}^{in}(\bar{\mathcal{Z}})$ ,  $\mathcal{P}^{out}(\bar{\mathcal{W}})$  for the upper image  $\mathcal{P}$  and  $\mathcal{P}^{in}(\bar{\mathcal{W}})$ ,  $\mathcal{D}^{out}(\bar{\mathcal{Z}})$  for the lower image  $\mathcal{D}$  using Proposition 4.7. Note that both  $\mathcal{P}^{out}(\bar{\mathcal{W}})$  and  $\mathcal{P}^k$  are outer approximations for  $\mathcal{P}$ . However,  $\mathcal{P}^{out}(\bar{\mathcal{W}})$  is a finer outer approximation than  $\mathcal{P}^k$ . The reason is that when  $\mathcal{P}^k$  is updated, only the vertices in more than  $\epsilon$ -distance to  $\mathcal{P}$  are used. On the other hand, all the vertices are considered when calculating  $\mathcal{P}^{out}(\bar{\mathcal{W}})$ . Furthermore, the algorithm returns a finite weak  $\epsilon$ -solution  $\bar{\mathcal{Z}}$  to  $(P_V)$  and a finite  $\epsilon$ -solution  $\bar{\mathcal{W}}$  to  $(D_V)$  (see Theorem 4.9 in Löhne *et al.* (2014)).

The steps of the primal algorithm are provided as Algorithm 1.

### 4.4 Dual algorithm

The steps of the dual algorithm follow in a way that is similar to the primal algorithm; however, as a major difference, at each iteration, an outer approximation for the dual image  $\mathcal{D}$  is obtained. Moreover, the dual algorithm does not require solving  $(P_2(v))$ ; only  $(P_1(w))$  is solved for different weights  $w$  in the initialization step as well as in the iterations. An optimal solution of  $(P_1(w))$  is used to update the outer approximation of  $\mathcal{D}$  as in Proposition 4.6.

---

**Algorithm 1** Primal Approximation Algorithm

---

```
1: Compute an optimal solution  $(x_{(j)}, y_{(j)}, z_{(j)})$  to  $(P_1(e_{(j)}))$  for each  $j \in \mathcal{J}$ ;  
2: Let  $\mathcal{P}^0 = \{z \in \mathbb{R}^J : e_{(j)}^\top z \geq \mathcal{P}_1(e_{(j)}), \forall j \in \mathcal{J}\}$ ;  
3:  $k \leftarrow 0$ ;  
    $\bar{\mathcal{Z}} \leftarrow \{(x_{(j)}, y_{(j)}, z_{(j)}) \mid j \in \mathcal{J}\}$ ;  
    $\bar{\mathcal{W}} \leftarrow \{e_{(j)} \mid j \in \mathcal{J}\}$ ;  
4: repeat  
5:    $\mathcal{M} \leftarrow \mathbb{R}^J$ ;  
6:   Compute the set  $\mathcal{V}^k$  of the vertices of  $\mathcal{P}^k$ ;  
7:   for each  $v \in \mathcal{V}^k$  do  
8:     Compute an optimal solution  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  of  $(P_2(v))$  and an optimal solution  $\gamma_{(v)}$  of  $(LD_2(v))$ ;  
9:      $\bar{\mathcal{Z}} \leftarrow \bar{\mathcal{Z}} \cup \{(x_{(v)}, y_{(v)}, v + \alpha_{(v)}\mathbf{1})\}$ ;  
      $\bar{\mathcal{W}} \leftarrow \bar{\mathcal{W}} \cup \{\gamma_{(v)}\}$ ;  
10:    if  $\alpha_{(v)} > \epsilon$  then  
11:       $\mathcal{M} \leftarrow \mathcal{M} \cap \{z \in \mathbb{R}^J : \gamma_{(v)}^\top z \geq \gamma_{(v)}^\top (v + \alpha_{(v)}\mathbf{1})\}$ ;  
12:      break;  
13:    end if  
14:  end for  
15:  if  $\mathcal{M} \neq \mathbb{R}^J$  then  
16:     $\mathcal{P}^{k+1} \leftarrow \mathcal{P}^k \cap \mathcal{M}$ ,  $k \leftarrow k + 1$ ;  
17:  end if  
18: until  $\mathcal{M} = \mathbb{R}^J$ ;  
19: Compute  $\mathcal{P}^{in}(\bar{\mathcal{Z}})$ ,  $\mathcal{P}^{out}(\bar{\mathcal{W}})$ ,  $\mathcal{P}^{in}(\bar{\mathcal{W}})$ ,  $\mathcal{P}^{out}(\bar{\mathcal{Z}})$  as in Proposition 4.7;  
20: return  $\begin{cases} \bar{\mathcal{Z}}: \text{A finite weak } \epsilon\text{-solution to } (P_V); \\ \bar{\mathcal{W}}: \text{A finite } \epsilon\text{-solution to } (D_V); \\ \mathcal{P}^{in}(\bar{\mathcal{Z}}), \mathcal{P}^{out}(\bar{\mathcal{W}}), \mathcal{P}^{in}(\bar{\mathcal{W}}), \mathcal{P}^{out}(\bar{\mathcal{Z}}); \end{cases}$ 
```

---

At the termination, the algorithm computes inner and outer approximations for the upper image  $\mathcal{P}$  and lower image  $\mathcal{D}$  using Proposition 4.7. Furthermore, the algorithm returns a finite weak  $\epsilon$ -solution  $\bar{\mathcal{Z}}$  to  $(P_V)$  and a finite  $\epsilon$ -solution  $\bar{\mathcal{W}}$  to  $(D_V)$  (see Theorem 4.14 in Löhne *et al.* (2014)).

The steps of the dual algorithm are provided as Algorithm 2.

## 5 Scenario decomposition for scalar problems

In this section, we are interested in solving the scalarization problems  $(P_1(w))$  and  $(P_2(v))$ . Note that these problems are single-objective multivariate risk-averse two-stage stochastic programming problems. For such problems, the problem size increases as the number of scenarios,  $I$ , gets larger. An efficient solution procedure is possible by scenario-wise decompositions. In the univariate case, for risk-neutral two-stage stochastic programming problems, see Birge and Louveaux (1997), Birge and Louveaux (1988), Kall and Mayer (2005), Ruszczyński (2003), Van Slyke and Wets (1969) for solution methodologies by scenario-wise decomposition. For scenario decompositions in two-stage risk-averse stochastic programming problems, the reader is referred to Ahmed (2006), Miller and Ruszczyński (2011), Fábíán (2008), Kristoffersen (2005) for problems with a single coherent risk-averse objective function and to Liu *et al.* (2016) for chance-constrained problems. Scenario-wise decomposition solution methodology is also possible for multi-stage stochastic programming problems with dynamic coherent risk measure as suggested in Collado *et al.* (2012). Different from these studies, the scalarization problems we solve are two-stage risk-averse stochastic programming problems with *multivariate convex* risk measures; therefore, these problems require different solutions techniques than the existing ones.

---

**Algorithm 2** Dual Approximation Algorithm
 

---

```

1: Compute an optimal solution  $(x_{(\eta)}, y_{(\eta)}, z_{(\eta)})$  to  $(P_1(\eta))$  for  $\eta = (\frac{1}{J}, \dots, \frac{1}{J})^\top$ ;
2: Let  $\mathcal{D}^0 = \{d \in \mathbb{R}^J \mid \mathcal{P}_1(\eta) \geq d^J\}$ ;
3:  $k \leftarrow 0$ ;
    $\bar{\mathcal{Z}} \leftarrow \{(x_{(\eta)}, y_{(\eta)}, z_{(\eta)})\}$ ;
    $\bar{\mathcal{W}} \leftarrow \{\eta\}$ ;
4: repeat
5:    $\mathcal{M} \leftarrow \mathbb{R}^J$ ;
6:   Compute the set  $\mathcal{V}^k$  of vertices of  $\mathcal{D}^k$ ;
7:   for each  $t = (t^1, \dots, t^{J-1}, t^J)^\top \in \mathcal{V}^k$  do
8:     Let  $w = (t^1, \dots, t^{J-1}, 1 - \sum_{j=1}^{J-1} t^j)^\top$ ;
9:     Compute an optimal solution  $(x_{(w)}, y_{(w)}, z_{(w)})$  to  $(P_1(w))$ ;
10:     $\bar{\mathcal{Z}} \leftarrow \bar{\mathcal{Z}} \cup \{(x_{(w)}, y_{(w)}, z_{(w)})\}$ ;
11:    if  $w \in \mathbb{R}_{++}^J$  or  $t^J - \mathcal{P}_1(w) \leq \epsilon$  then
12:       $\bar{\mathcal{W}} \leftarrow \bar{\mathcal{W}} \cup \{w\}$ ;
13:    end if
14:    if  $t^J - \mathcal{P}_1(w) > \epsilon$  then
15:       $\mathcal{M} \leftarrow \mathcal{M} \cap \{d \in \mathbb{R}^J \mid (z_{(w)}^J - z_{(w)}^1, \dots, z_{(w)}^J - z_{(w)}^{J-1}, 1)^\top d \leq z_{(w)}^J\}$ ;
16:      break;
17:    end if
18:  end for
19:  if  $\mathcal{M} \neq \mathbb{R}^J$  then
20:     $\mathcal{D}^{k+1} \leftarrow \mathcal{D}^k \cap \mathcal{M}$ ,  $k \leftarrow k + 1$ ;
21:  end if
22: until  $\mathcal{M} = \mathbb{R}^J$ ;
23: Compute  $\mathcal{P}^{in}(\bar{\mathcal{Z}})$ ,  $\mathcal{P}^{out}(\bar{\mathcal{W}})$ ;  $\mathcal{P}^{in}(\bar{\mathcal{W}})$ ,  $\mathcal{P}^{out}(\bar{\mathcal{Z}})$  as in Proposition 4.7;
24: return  $\begin{cases} \bar{\mathcal{Z}}: \text{A finite weak } \epsilon\text{-solution to } (P_V); \\ \bar{\mathcal{W}}: \text{A finite } \epsilon\text{-solution to } (D_V); \\ \mathcal{P}^{in}(\bar{\mathcal{Z}}), \mathcal{P}^{out}(\bar{\mathcal{W}}), \mathcal{P}^{in}(\bar{\mathcal{W}}), \mathcal{P}^{out}(\bar{\mathcal{Z}}); \end{cases}$ 

```

---

## 5.1 The problem of weighted sum scalarization

Let  $w \in \mathbb{R}_+^J \setminus \{0\}$ . The weighted sum scalarization problem  $(P_1(w))$  defined in Section 3 can be rewritten more explicitly as:

$$\begin{aligned}
 \min \quad & w^\top z && (P_1(w)) \\
 \text{s.t.} \quad & z \in R(Cx + Qy) \\
 & Ax = b \\
 & T_i x + W_i y_i = h_i, \quad \forall i \in \Omega \\
 & z \in \mathbb{R}^J, x \in \mathbb{R}_+^M, y_i \in \mathbb{R}_+^N, \quad \forall i \in \Omega.
 \end{aligned}$$

We propose a Lagrangian dual reformulation of  $(P_1(w))$  whose objective function is scenario-wise decomposable. The details are provided in Section 5.1.1. Based on this dual reformulation, in Section 5.1.2, we propose a dual cutting-plane algorithm for  $(P_1(w))$ , called the *dual bundle method*, which provides an optimal dual solution. As the Benson algorithms in Section 4 require an optimal primal solution in addition to an optimal dual solution, in Section 5.1.3, we show that such a primal solution can be obtained from the dual of the so-called *master problem* in the dual bundle method.

### 5.1.1 Scenario decomposition

To derive a decomposition algorithm for  $(P_1(w))$ , we *randomize* the first stage variable  $x \in \mathbb{R}^M$  and treat it as an element  $x \in \mathbb{L}^M$  with realizations  $x_1, \dots, x_I \in \mathbb{R}^M$ . To ensure the equivalence of the new formulation

with the previous one, we add the so-called *nonanticipativity constraints*

$$p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega,$$

which are equivalent to  $x_1 = \dots = x_I$ .

Let us introduce

$$\mathcal{F} := \{(x, y) \in \mathbb{L}^M \times \mathbb{L}^N \mid (x_i, y_i) \in \mathcal{F}_i, \forall i \in \Omega\}, \quad (5.1)$$

where, for each  $i \in \Omega$ ,

$$\mathcal{F}_i := \{(x_i, y_i) \in \mathbb{R}_+^M \times \mathbb{R}_+^N \mid Ax_i = b, T_i x_i + W_i y_i = h_i\}.$$

With this notation and using the nonanticipativity constraints, we may rewrite  $(P_1(w))$  as follows:

$$\begin{aligned} \min \quad & w^\top z & (P'_1(w)) \\ \text{s.t.} \quad & z \in R(Cx + Qy) \\ & p_i(x_i - \mathbb{E}[x]) = 0, \quad \forall i \in \Omega \\ & (x, y) \in \mathcal{F}, z \in \mathbb{R}^J. \end{aligned}$$

Note that the optimal value of  $(P'_1(w))$  is  $\mathcal{P}_1(w)$ .

The following theorem provides a dual formulation of  $(P'_1(w))$  by relaxing the nonanticipativity constraints in a Lagrangian fashion. We call this dual formulation as  $(D_1(w))$ .

**Theorem 5.1.** *It holds*

$$\mathcal{P}_1(w) = \sup_{\mu \in \mathbb{M}_1^J, \lambda \in \mathbb{L}^M} \left\{ \sum_{i \in \Omega} f_i(\mu_i, \lambda_i, w) - \beta(\mu, w) \mid \mathbb{E}[\lambda] = 0 \right\}, \quad (D_1(w))$$

where, for each  $i \in \Omega$ ,  $\mu_i \in \mathbb{R}_+^J$ ,  $\lambda_i \in \mathbb{R}^M$ ,

$$f_i(\mu_i, \lambda_i, w) := \inf_{(x_i, y_i) \in \mathcal{F}_i} (w^\top [\mu_i \cdot (Cx_i + Q_i y_i)] + p_i \lambda_i^\top x_i), \quad (5.2)$$

and  $\beta$  is defined by (2.1).

**Proof.** We may write

$$\mathcal{P}_1(w) = \inf_{(x, y) \in \mathcal{F}, z \in \mathbb{R}^J} \{w^\top z \mid z \in R(Cx + Qy), p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega\} \quad (5.3)$$

$$= \inf_{(x, y) \in \mathcal{F}} \left\{ \inf_{z \in R(Cx + Qy)} w^\top z \mid p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega \right\} \quad (5.4)$$

$$= \inf_{(x, y) \in \mathcal{F}} \left\{ \sup_{\mu \in \mathbb{M}_1^J} (w^\top \mathbb{E}^\mu [Cx + Qy] - \beta(\mu, w)) \mid p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega \right\}, \quad (5.5)$$

where the passage to the last line is by (2.3). Using the minimax theorem of Sion (1958), we may interchange the infimum and the supremum in the last line. This yields

$$\mathcal{P}_1(w) = \sup_{\mu \in \mathbb{M}_1^J} (F(\mu, w) - \beta(\mu, w)), \quad (5.6)$$

where, for each  $\mu \in \mathbb{M}_1^J$ ,

$$F(\mu, w) := \inf_{(x, y) \in \mathcal{F}} \{w^\top \mathbb{E}^\mu [Cx + Qy] \mid p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega\}. \quad (5.7)$$

Let us fix  $\mu \in \mathbb{M}_1^J$ . Note that  $F(\mu, w)$  is the optimal value of a large-scale linear program where the only coupling constraints between the decision variables for different scenarios are the nonanticipativity constraints. To obtain a formulation of this problem that can be decomposed into a subproblem for each

scenario, we dualize the nonanticipativity constraints. The reader is referred to Section 2.4.2 of Shapiro *et al.* (2009) for the details on the dualization of nonanticipativity constraints. To that end, let us assign Lagrange multipliers  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_I \in \mathbb{R}^M$  for the non-anticipativity constraints. Note that we may consider them as the realizations of a random Lagrange multiplier  $\tilde{\lambda} \in \mathbb{L}^M$ . By strong duality for linear programming,

$$F(\mu, w) = \sup_{\tilde{\lambda} \in \mathbb{L}^M} \inf_{(x, y) \in \mathcal{F}} \ell(x, y, \tilde{\lambda}),$$

where the Lagrangian  $\ell$  is defined by

$$\begin{aligned} \ell(x, y, \tilde{\lambda}) &:= w^\top \mathbb{E}^\mu [Cx + Qy] + \sum_{i \in \Omega} p_i \tilde{\lambda}_i^\top (x_i - \mathbb{E}[x]) \\ &= w^\top \mathbb{E}^\mu [Cx + Qy] + \sum_{i \in \Omega} p_i (\tilde{\lambda}_i - \mathbb{E}[\tilde{\lambda}])^\top x_i, \end{aligned}$$

for each  $x \in \mathbb{L}^M, y \in \mathbb{L}^N, \tilde{\lambda} \in \mathbb{L}^M$ . Given such  $x, y, \tilde{\lambda}$ , note that

$$\begin{aligned} \ell(x, y, \tilde{\lambda}) &= \ell(x, y, \lambda) = w^\top \mathbb{E}^\mu [Cx + Qy] + \sum_{i \in \Omega} p_i \lambda_i^\top x_i \\ &= w^\top \mathbb{E}^\mu [Cx + Qy] + \mathbb{E}[\lambda^\top x], \end{aligned}$$

if we set  $\lambda := \tilde{\lambda} - \mathbb{E}[\tilde{\lambda}]$ . In this case,  $\mathbb{E}[\lambda] = 0$ . Therefore, we obtain

$$\begin{aligned} F(\mu, w) &= \sup_{\lambda \in \mathbb{L}^M} \left\{ \inf_{(x, y) \in \mathcal{F}} \ell(x, y, \lambda) \mid \mathbb{E}[\lambda] = 0 \right\} \\ &= \sup_{\lambda \in \mathbb{L}^M} \left\{ \inf_{(x, y) \in \mathcal{F}} (w^\top \mathbb{E}^\mu [Cx + Qy] + \mathbb{E}[\lambda^\top x]) \mid \mathbb{E}[\lambda] = 0 \right\} \\ &= \sup_{\lambda \in \mathbb{L}^M} \left\{ \sum_{i \in \Omega} f_i(\mu_i, \lambda_i, w) \mid \mathbb{E}[\lambda] = 0 \right\}, \end{aligned} \tag{5.8}$$

where  $f_i(\mu_i, \lambda_i, w)$ , defined by (5.2), is the optimal value of the subproblem for scenario  $i \in \Omega$ . The assertion of the theorem follows from (5.6) and (5.8).  $\square$

### 5.1.2 The dual bundle method

To solve  $(D_1(w))$  given in Theorem 5.1, we propose a dual bundle method which constructs *affine* upper approximations for  $f_i(\cdot, \cdot, w)$ ,  $i \in \Omega$ , and  $-\beta(\cdot, w)$ . The upper approximations are based on the subgradients of these functions at points  $(\mu^{(\ell)}, \lambda^{(\ell)})$  that are generated iteratively by solving the so-called *master problem*. The reader is referred to Ruszczyński (2006) for the details of the bundle method.

For  $i \in \Omega, \mu'_i \in \mathbb{R}_+^J, \lambda'_i \in \mathbb{R}^M$ , we denote by  $\partial_{\mu_i, \lambda_i} f_i(\mu'_i, \lambda'_i, w)$  the subdifferential of the concave function  $f_i(\cdot, \cdot, w)$  at the point  $(\mu'_i, \lambda'_i)$ , that is,  $\partial_{\mu_i, \lambda_i} f_i(\mu'_i, \lambda'_i, w)$  is the set of all vectors  $(g_{\mu'_i}, g_{\lambda'_i}) \in \mathbb{R}^{J+M}$  such that

$$f_i(\mu_i, \lambda_i, w) \leq f_i(\mu'_i, \lambda'_i, w) + g_{\mu'_i}^\top (\mu_i - \mu'_i) + g_{\lambda'_i}^\top (\lambda_i - \lambda'_i), \tag{5.9}$$

for all  $\mu_i \in \mathbb{R}_+^J, \lambda_i \in \mathbb{R}^M$ . Note that the right hand side of (5.9) provides an *affine* upper approximation for  $f_i(\cdot, \cdot, w)$ . For this reason, (5.9) is called as a *cut*.

Similarly, we denote by  $\partial_\mu(-\beta)(\mu', w)$  the subdifferential of the concave function  $-\beta(\cdot, w)$  at a point  $\mu' \in \mathbb{M}_1^J$ , which is the set of all vectors  $\rho_{\mu'} = (\rho_{\mu'_1}, \dots, \rho_{\mu'_I}) \in \mathbb{R}^{J \times I}$  such that

$$-\beta(\mu, w) \leq -\beta(\mu', w) + \sum_{i \in \Omega} \rho_{\mu'_i}^\top (\mu_i - \mu'_i) \tag{5.10}$$

for all  $\mu \in \mathbb{M}_1^J$ . We call (5.10) a *cut* for  $-\beta(\cdot, w)$ .

In the next proposition, we show how to compute the subdifferential of the function  $f_i(\cdot, \cdot, w)$  at a point  $(\mu'_i, \lambda'_i)$ .

**Proposition 5.2.** For  $i \in \Omega$ ,  $\mu'_i \in \mathbb{R}_+^J$ ,  $\lambda'_i \in \mathbb{R}^M$ , let

$$A_i(\mu'_i, \lambda'_i, w) := \arg \min_{(x_i, y_i) \in \mathcal{F}_i} (w^\top [\mu'_i \cdot (Cx_i + Q_i y_i)] + p_i (\lambda'_i)^\top x_i).$$

Then,

$$\partial_{\mu_i, \lambda_i} f_i(\mu'_i, \lambda'_i, w) = \text{co} \{(w \cdot (Cx_i + Q_i y_i), p_i x_i) \mid (x_i, y_i) \in A_i(\mu'_i, \lambda'_i, w)\}.$$

**Proof.** Let  $\varphi_i(x_i, y_i, \mu_i, \lambda_i, w) := w^\top [\mu_i \cdot (Cx_i + Q_i y_i)] + p_i \lambda_i^\top x_i$ . The function  $\varphi_i(x_i, y_i, \cdot, \cdot, w)$  is affine for all  $(x_i, y_i) \in \mathcal{F}_i$ . The function  $\varphi_i(\cdot, \cdot, \mu_i, \lambda_i, w)$  is also affine and continuous for all  $\mu_i \in \mathbb{R}_+^J$ ,  $\lambda_i \in \mathbb{R}^M$ . Finally, the set  $\mathcal{F}_i$  is compact by assumption. By Theorem 2.87 in Ruszczyński (2006), the assertion of the proposition follows.  $\square$

Next, we show how to compute a subgradient of the function  $-\beta(\cdot, w)$  at a point  $\mu'$ .

**Proposition 5.3.** Recall that the set  $\mathcal{A} = \{u \in \mathbb{L}^J \mid 0 \in R(u)\}$  is the acceptance set of  $R$ . For  $\mu' \in \mathbb{L}_+^J$ , let

$$B(\mu', w) := \arg \max_{u \in \mathcal{A}} w^\top \mathbb{E}^{\mu'} [u]$$

and assume that  $B(\mu', w) \neq \emptyset$ . Then,

$$\partial_\mu(-\beta)(\mu', w) \supseteq \text{co} \{(w \cdot u_1, \dots, w \cdot u_I) \mid u = (u_1, \dots, u_I) \in B(\mu', w)\}.$$

**Proof.** Let  $\varphi(u, \mu, w) := w^\top \mathbb{E}^\mu [u]$ . The function  $\varphi(u, \cdot, w)$  is affine for all  $u \in \mathcal{A}$ . The function  $\varphi(\cdot, \mu, w)$  is also affine and continuous for all  $\mu \in \mathbb{M}_1$ . By Theorem 2.87 in Ruszczyński (2006), the assertion of the proposition follows.  $\square$

**Remark 5.4.** For practical risk measures, such as the multivariate entropic risk measure (see Example 2.3), the function  $-\beta(\cdot, w)$  is differentiable and the subdifferential is a singleton. For coherent multivariate risk measures, such as the multivariate CVaR (see Example 2.2), there exists a convex cone  $\mathcal{Q} \subseteq \mathbb{M}_1^J$  such that  $-\beta(\mu, w) = 0$  if  $\mu \in \mathcal{Q}$  and  $-\beta(\mu, w) = -\infty$  otherwise. For multivariate CVaR with risk-aversion parameter  $\nu \in (0, 1)^J$ ,

$$\mathcal{Q} = \left\{ \mu \in \mathbb{M}_1^J \mid \frac{\mu_i^j}{p_i} \leq \frac{1}{1 - \nu^j}, \forall i \in \Omega, j \in \mathcal{J} \right\}.$$

For a coherent multivariate risk measure,  $\partial_\mu(-\beta)(\mu')$  is the set of all normal directions of  $\mathcal{Q}$  at  $\mu'$ . It follows that the cut (5.10) is always satisfied; therefore, it can be ignored.

At each iteration  $k$  of the bundle method, we solve the *master problem*

$$\max \sum_{i \in \Omega} \vartheta_i + \eta - \sum_{i \in \Omega} \varrho \left\| \mu_i - \bar{\mu}_i^{(k)} \right\|^2 - \sum_{i \in \Omega} \varrho \left\| \lambda_i - \bar{\lambda}_i^{(k)} \right\|^2 \quad (MP_1(w))$$

$$\text{s.t.} \quad \vartheta_i \leq f_i(\mu_i^{(\ell)}, \lambda_i^{(\ell)}, w) + g_{\mu_i^{(\ell)}}^\top (\mu_i - \mu_i^{(\ell)}) + g_{\lambda_i^{(\ell)}}^\top (\lambda_i - \lambda_i^{(\ell)}), \quad \forall i \in \Omega, \ell \in \mathcal{L} \quad (5.11)$$

$$\eta \leq -\beta(\mu^{(\ell)}, w) + \sum_{i \in \Omega} \rho_{\mu_i^{(\ell)}}^\top (\mu_i - \mu_i^{(\ell)}), \quad \forall \ell \in \mathcal{L} \quad (5.12)$$

$$\sum_{i \in \Omega} p_i \lambda_i = 0 \quad (5.13)$$

$$\sum_{i \in \Omega} \mu_i = \mathbf{1} \quad (5.14)$$

$$\mu_i \in \mathbb{R}_+^J, \lambda_i \in \mathbb{R}^M, \vartheta_i \in \mathbb{R}, \quad \forall i \in \Omega \quad (5.15)$$

$$\eta \in \mathbb{R}, \quad (5.16)$$

with  $\mathcal{L} = \{1, \dots, k\}$ ,  $\varrho > 0$ . Here,  $\|\cdot\|$  denotes the Euclidean norm on an appropriate dimension. Note that constraints (5.14) and (5.15) for  $\mu$  are equivalent to having  $\mu \in \mathbb{M}_1^J$ , and constraint (5.13) for  $\lambda$  is equivalent to having  $\mathbb{E}[\lambda] = 0$ .  $\bar{\mu}^{(k)} \in \mathbb{M}_1^J$ ,  $\bar{\lambda}^{(k)} \in \mathbb{L}^M$  with  $\mathbb{E}[\bar{\lambda}^{(k)}] = 0$  are parameters of the problem, called the *centers*, that are initialized and updated within the bundle method. The quadratic terms in the objective function are Moreau-Yosida regularization terms and they make the overall objective function strictly convex. These regularization terms enforce an optimal solution of  $(MP_1(w))$  to be close to the centers.

Let  $(\mu^{(k+1)}, \lambda^{(k+1)}, \vartheta^{(k+1)}, \eta^{(k+1)})$  be an optimal solution for  $(MP_1(w))$ . Computing the subgradients

$$\begin{aligned} (g_{\mu_i^{(k+1)}}, g_{\lambda_i^{(k+1)}}) &\in \partial_{\mu_i, \lambda_i} f_i(\mu_i^{(k+1)}, \lambda_i^{(k+1)}), \forall i \in \Omega, \\ (\rho_{\mu_1^{(k+1)}}, \dots, \rho_{\mu_I^{(k+1)}}) &\in \partial_{\mu} \beta(\mu^{(k+1)}, w) \end{aligned}$$

at this optimal solution and using (5.9) and (5.10), a cut for each of the functions  $f_i(\cdot, \cdot, w)$ ,  $i \in \Omega$ , and  $-\beta(\cdot, w)$  are added to  $(MP_1(w))$  at the next iteration in order to improve the upper approximations for these functions.

The centers are updated in the following fashion. At iteration  $k$ , one checks if the difference between the objective value of  $(D_1(w))$  evaluated at the point  $(\mu^{(k)}, \lambda^{(k)})$ , that is,  $\sum_{i \in \Omega} f_i(\mu_i^{(k)}, \lambda_i^{(k)}, w) - \beta(\mu^{(k)}, w)$ , and the objective value evaluated at the centers  $\bar{\mu}^{(k-1)}, \bar{\lambda}^{(k-1)}$ , that is,  $\sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k-1)}, \bar{\lambda}_i^{(k-1)}, w) - \beta(\bar{\mu}^{(k-1)}, w)$ , is larger than a threshold. If so, this means that an optimal solution of  $(D_1(w))$  is close to  $(\mu^{(k)}, \lambda^{(k)})$ . Therefore, the new centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  are set to  $\mu^{(k)}, \lambda^{(k)}$ , respectively. This is called a *descent step*. Otherwise, the centers remain unchanged, that is,  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  are set to  $\bar{\mu}^{(k-1)}, \bar{\lambda}^{(k-1)}$ , respectively.

The steps of our dual bundle method are provided as Algorithm 3. By (Ruszczynski, 2006, Theorem 7.16), the bundle method generates a sequence  $(\bar{\mu}^{(k)}, \bar{\lambda}^{(k)})_{k \in \mathbb{N}}$  that converges to an optimal solution of  $(D_1(w))$  as  $k \rightarrow \infty$ . In practice, the stopping condition in line 22 of Algorithm 3 is not satisfied. Therefore, it is a general practice to stop the algorithm when

$$\sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} - \bar{F}^{(k+1)} \leq \varepsilon \quad (5.17)$$

for some small constant  $\varepsilon > 0$ .

**Remark 5.5.** Note that the objective function of  $(MP_1(w))$  can be replaced with

$$\vartheta + \eta - \sum_{i \in \Omega} \varrho \left\| \mu_i - \bar{\mu}_i^{(k)} \right\|^2 - \sum_{i \in \Omega} \varrho \left\| \lambda_i - \bar{\lambda}_i^{(k)} \right\|^2$$

and constraint (5.11) can be replaced with

$$\vartheta \leq \sum_{i \in \Omega} \left( f_i(\mu_i^{(\ell)}, \lambda_i^{(\ell)}, w) + g_{\mu_i^{(\ell)}}^T(\mu_i - \mu_i^{(\ell)}) + g_{\lambda_i^{(\ell)}}^T(\lambda_i - \lambda_i^{(\ell)}) \right), \forall \ell \in \mathcal{L}.$$

This way one would obtain an upper approximation for the sum  $\sum_{i \in \Omega} f_i(\cdot, \cdot, w)$ . Compared to the multiple cuts in (5.11), this provides a looser upper approximation for  $\sum_{i \in \Omega} f_i(\cdot, \cdot, w)$ . However, while one adds  $I = |\Omega|$  cuts at each iteration in the multiple cuts version, this approach adds a single cut.

### 5.1.3 Recovery of primal solution

Both the primal and the dual Benson algorithms require an optimal solution  $(x_{(w)}, y_{(w)}, z_{(w)})$  of the problem  $(P_1'(w))$ . Therefore, in Theorem 5.6, we suggest a procedure to recover an optimal primal solution from the solution of the master problem  $(MP_1(w))$ .

**Theorem 5.6.** *Let  $\mathcal{L} = \{1, \dots, k\}$  be the index set at the last iteration of the dual bundle method with the approximate stopping condition (5.17) for some  $\varepsilon > 0$ . Let  $n + 1$  be the first descent iteration after the approximate stopping condition is satisfied and let  $\mathcal{L}' = \{1, \dots, n\}$ . For  $(MP_1(w))$  with centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ , let  $\tau = (\tau_i^{(\ell)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta = (\theta^{(\ell)})_{\ell \in \mathcal{L}'}, \sigma \in \mathbb{R}^M, \Psi \in \mathbb{R}^J, \nu = (\nu_i)_{i \in \Omega}$  be the Lagrangian dual*

---

**Algorithm 3** A Dual Bundle Method for  $(P_1(w))$ 


---

1:  $k \leftarrow 0, \mathcal{L} \leftarrow \emptyset, \gamma \in (0, 1), \vartheta_i^{(1)} \leftarrow \infty$  for each  $i \in \Omega, \eta^{(1)} \leftarrow \infty, \bar{F}^{(1)} \leftarrow 0$ ;  
2: Let  $\mu^{(1)} \in \mathbb{M}_1^J, \lambda^{(1)} \in \mathbb{L}^M$  be such that  $\mathbb{E}[\lambda^{(1)}] = 0$ ;  
3: **repeat**  
4:    $k \leftarrow k + 1$ ;  
5:   **for** each  $i \in \Omega$  **do**  
6:     Compute an optimal solution  $(x_i^{(k)}, y_i^{(k)})$  and the optimal value  $f_i(\mu_i^{(k)}, \lambda_i^{(k)}, w)$  of the subproblem
$$\min_{(x_i, y_i) \in \mathcal{F}_i} \left( w^\top \left[ \mu_i^{(k)} \cdot (Cx_i + Q_i y_i) \right] + p_i(\lambda_i^{(k)})^\top x_i \right);$$
7:     Compute subgradients  $g_{\mu_i^{(k)}} = w \cdot (Cx_i^{(k)} + Q_i y_i^{(k)}), g_{\lambda_i^{(k)}} = p_i x_i^{(k)}$ ;  
8:   **end for**  
9:   Compute  $\beta(\mu^{(k)}, w)$  and subgradient  $(\rho_{\mu_1^{(k)}}, \dots, \rho_{\mu_I^{(k)}}) \in \partial_{\mu}(-\beta)(\mu^{(k)}, w)$ ;  
10:    $F^{(k)} \leftarrow \sum_{i \in \Omega} f_i(\mu_i^{(k)}, \lambda_i^{(k)}, w) - \beta(\mu^{(k)}, w)$ ;  
11:   **if**  $F^{(k)} < \sum_{i \in \Omega} \vartheta_i^{(k)} + \eta^{(k)}$  **then**  
12:      $\mathcal{L} \leftarrow \mathcal{L} \cup \{k\}$ ;  
13:   **end if**  
14:   **if**  $(k = 1)$  or  $(k \geq 2$  and  $F^{(k)} \geq (1 - \gamma)\bar{F}^{(k)} + \gamma(\sum_{i \in \Omega} \vartheta_i^{(k)} + \eta^{(k)}))$  **then**  
15:      $\bar{\mu}^{(k)} \leftarrow \mu^{(k)}, \bar{\lambda}^{(k)} \leftarrow \lambda^{(k)}$ ;  
16:   **else**  
17:      $\bar{\mu}^{(k)} \leftarrow \bar{\mu}^{(k-1)}, \bar{\lambda}^{(k)} \leftarrow \bar{\lambda}^{(k-1)}$ ;  
18:   **end if**  
19:   Solve the master problem. Let  $(\mu^{(k+1)}, \lambda^{(k+1)}, \vartheta^{(k+1)}, \eta^{(k+1)})$  be an optimal solution;  
20:   (Optional) Remove all cuts whose dual variables at the solution of master problem are zero;  
21:    $\bar{F}^{(k+1)} \leftarrow \sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k)}, \bar{\lambda}_i^{(k)}, w) - \beta(\bar{\mu}^{(k)}, w)$ ;  
22: **until**  $\sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} = \bar{F}^{(k+1)}$ ;  
23: **return**  $\begin{cases} \bar{F}^{(k+1)} & : \text{Optimal value } \mathcal{P}_1(w); \\ (\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}) & : \text{An optimal solution of } (D_1(w)); \end{cases}$

---

variables assigned to the constraints (5.11), (5.12), (5.13), (5.14), (5.15), respectively, with  $\tau_i^{(\ell)} \geq 0, \theta^{(\ell)} \geq 0, \nu_i \in \mathbb{R}_+^J$  for each  $i \in \Omega, \ell \in \mathcal{L}'$ . Let  $(x_i^{(\ell)}, y_i^{(\ell)})$  be an optimal solution of the subproblem in line 6 of Algorithm 3 for each  $i \in \Omega$  and  $\ell \in \mathcal{L}'$ . Let

$$\left( \tau^{(n+1)} = (\tau_i^{(\ell, n+1)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta^{(n+1)} = (\theta^{(\ell, n+1)})_{\ell \in \mathcal{L}'}, \sigma^{(n+1)}, \Psi^{(n+1)}, \nu^{(n+1)} = (\nu_i^{(n+1)})_{i \in \Omega} \right)$$

be a dual optimal solution for  $(MP_1(w))$ . Let  $x_{(w)} = ((x_{(w)})_i)_{i \in \Omega}, y_{(w)} = ((y_{(w)})_i)_{i \in \Omega}$  be defined by

$$(x_{(w)})_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} x_i^{(\ell)}, \quad (y_{(w)})_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} y_i^{(\ell)}.$$

Moreover, let  $z_{(w)}$  be a minimizer of the problem

$$\inf_{z \in R(Cx_{(w)} + Qy_{(w)})} w^\top z.$$

Then,  $(x_{(w)}, y_{(w)}, z_{(w)})$  is an approximately optimal solution of  $(P'_1(w))$  in the following sense:

- (a)  $((x_{(w)})_i, (y_{(w)})_i) \in \mathcal{F}_i$  for each  $i \in \Omega$ .
- (b)  $z_{(w)} \in R(Cx_{(w)} + Qy_{(w)})$ .
- (c) As  $\varepsilon \rightarrow 0$ , it holds  $(x_{(w)})_i - \sigma^{(n+1)} \rightarrow 0$  for each  $i \in \Omega$ .



(d) As  $\varepsilon \rightarrow 0$ , it holds  $w^\top z_{(w)} \rightarrow \mathcal{P}_1(w)$ .

The proof of Theorem 5.6 is given in Appendix A.

## 5.2 The problem of scalarization by a reference variable

Let  $v \in \mathbb{R}^J \setminus \mathcal{P}$ . The problem  $(P_2(v))$  defined in Section 4.1 is formulated to find the minimum step-length to enter  $\mathcal{P}$  from  $v$  along the direction  $\mathbf{1} \in \mathbb{R}^J$  and it can be rewritten more explicitly as

$$\begin{aligned} \min \quad & \alpha & (P_2(v)) \\ \text{s.t.} \quad & v + \alpha \mathbf{1} \in R(Cx + Qy) \\ & Ax = b \\ & T_i x + W_i y_i = h_i \quad \forall i \in \Omega \\ & \alpha \in \mathbb{R}, x \in \mathbb{R}^M, y_i \in \mathbb{R}_+^N \quad \forall i \in \Omega. \end{aligned}$$

We propose a scenario-wise decomposition solution methodology for  $(P_2(v))$ . Even the steps we follow are similar to the ones for  $(P_1(w))$ , the decomposition is more complicated because the weights are not parameters but instead they are decision variables in the dual problem of  $(P_2(v))$  (see Theorem 5.11 below). Therefore, following the same steps as in  $(P_1(w))$  results in a nonconvex optimization problem. In order to resolve this convexity issue, we propose a new formulation for  $(P_2(v))$  by introducing finite measures to the dual representation of  $R$ .

The flow of this section is as follows: in Sections 5.2.1 and 5.2.2, we propose a scenario-wise decomposition solution methodology for  $(P_2(v))$ . Section 5.2.3 is devoted to the recovery of a primal solution.

### 5.2.1 Scenario decomposition

To derive a decomposition algorithm for  $(P_2(v))$ , we randomize the first stage variable  $x \in \mathbb{R}^M$  as in  $(P_1(w))$  and add the nonanticipativity constraints

$$p_i(x_i - \mathbb{E}[x]) = 0, \quad \forall i \in \Omega.$$

Using the feasible region  $\mathcal{F}$  defined by (5.1), we may rewrite  $(P_2(v))$  as follows:

$$\begin{aligned} \min \quad & \alpha & (P'_2(v)) \\ \text{s.t.} \quad & v + \alpha \mathbf{1} \in R(Cx + Qy) \\ & p_i(x_i - \mathbb{E}[x]) = 0 \quad \forall i \in \Omega \\ & (x, y) \in \mathcal{F} \\ & \alpha \in \mathbb{R} \end{aligned}$$

Note that the optimal value of  $(P'_2(v))$  is  $\mathcal{P}_2(v)$ .

Different from the approach for  $(P'_1(w))$ , in order to obtain a convex dual problem for  $(P'_2(v))$ , we use finite measures  $m$  instead of probability measures  $\mu$  in the dual representation of  $R$ . To that end, let  $\mathbb{M}_f^J$  be the set of all  $J$ -dimensional vectors  $m = (m^1, \dots, m^J)^\top$  of finite measures on  $\Omega$ , that is, for each  $j \in \mathcal{J}$ , the finite measure  $m^j$  assigns  $m_i^j$  to the elementary event  $\{i\}$  for  $i \in \Omega$ . For  $m \in \mathbb{M}_f^J$  and  $i \in \Omega$ , we also write  $m_i := (m_i^1, \dots, m_i^J)^\top \in \mathbb{R}^J$ .

The following lemma provides the relationship between  $\mu$  and  $m$ .

**Lemma 5.7.** *For every  $\mu \in \mathbb{M}_1^J$  and  $\gamma \in \mathbb{R}_+^J \setminus \{0\}$ , there exists  $m \in \mathbb{M}_f^J$  such that*

$$\gamma^\top \mathbb{E}^\mu [u] = \sum_{i \in \Omega} m_i^\top u_i, \quad \gamma^\top \mathbf{1} = \sum_{i \in \Omega} m_i^\top \mathbf{1}, \quad (5.18)$$

for every  $u \in \mathbb{L}^J$ . Conversely, for every  $m \in \mathbb{M}_f^J$ , there exist  $\mu \in \mathbb{M}_1^J$  and  $\gamma \in \mathbb{R}_+^J \setminus \{0\}$  such that (5.18) holds for every  $u \in \mathbb{L}^J$ .

**Proof.** Let  $\mu \in \mathbb{M}_1^J$  and  $\gamma \in \mathbb{R}_+^J \setminus \{0\}$ . Define  $m \in \mathbb{M}_f^J$  by

$$m_i^j = \gamma^j \mu_i^j, \quad \forall i \in \Omega, j \in \mathcal{J}.$$

Then, trivially, (5.18) holds for every  $u \in \mathbb{L}^J$ . Conversely, let  $m \in \mathbb{M}_f^J$ . Define  $\mu \in \mathbb{M}_1^J$  and  $\gamma \in \mathbb{R}_+^J \setminus \{0\}$  by

$$\gamma^j = \sum_{i \in \Omega} m_i^j, \quad \mu_i^j = \begin{cases} \frac{m_i^j}{\gamma^j} & \text{if } \gamma^j > 0, \\ \frac{1}{I} & \text{if } \gamma^j = 0, \end{cases} \quad \forall i \in \Omega, j \in \mathcal{J}.$$

Then, trivially, (5.18) holds for every  $u \in \mathbb{L}^J$ . □

Recall that  $\beta$  is the minimal penalty function of  $R$  as defined in (2.1). For  $m \in \mathbb{M}_f^J$ , let us define

$$\tilde{\beta}(m) = \sup_{u \in \mathcal{A}} \sum_{i \in \Omega} m_i^\top u_i. \quad (5.19)$$

Similarly, recall the function  $f_i$  defined in (5.2). For  $m \in \mathbb{M}_f^J$  and  $\lambda \in \mathbb{L}^M$ , let us define

$$\tilde{f}_i(m_i, \lambda_i) = \inf_{(x_i, y_i) \in \mathcal{F}_i} (m_i^\top (C x_i + Q_i y_i) + p_i \lambda_i^\top x_i). \quad (5.20)$$

Therefore, if  $\mu \in \mathbb{M}_1^J, \gamma \in \mathbb{R}_+^J \setminus \{0\}$  and  $m \in \mathbb{M}_f^J$  are related as in Lemma 5.7 and  $\lambda \in \mathbb{L}^M$ , then it is clear that

$$\beta(\mu, \gamma) = \tilde{\beta}(m), \quad f_i(\mu_i, \lambda_i, \gamma) = \tilde{f}_i(m_i, \lambda_i).$$

**Example 5.8.** Recall Example 2.3 on the multivariate entropic risk measure. The function  $\tilde{\beta}(\cdot)$  takes the form

$$\tilde{\beta}(m) = \sum_{j \in \mathcal{J}} \frac{1}{\delta^j} \left( H(m^j \| p) - m^{j \top} \mathbf{1} \right) + \inf_{s \in C^+} \sum_{j \in \mathcal{J}} \frac{1}{\delta^j} \left( s^j - (m^{j \top} \mathbf{1}) \log s^j \right),$$

where  $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^I$  and  $H(m^j \| p)$  is the relative entropy of  $m^j$  with respect to  $p$  defined by

$$H(m^j \| p) = \sum_{i \in \Omega} m_i^j \log \left( \frac{m_i^j}{p_i} \right).$$

**Example 5.9.** Recall Example 2.2 on the multivariate CVaR. The function  $\tilde{\beta}(\cdot)$  takes the form

$$\tilde{\beta}(m) = \begin{cases} 0 & \text{if } \frac{m_i^j}{p_i} \leq \frac{1}{1-\nu^j} \sum_{i \in \Omega} m_i^j, \quad \forall i \in \Omega, j \in \mathcal{J}, \\ +\infty & \text{else,} \end{cases}$$

for every  $m \in \mathbb{M}_f^J$ .

**Remark 5.10.** Note that, in general,  $\beta(\cdot, \cdot)$  is not a convex function since  $(\mu, \gamma) \mapsto \gamma^\top \mathbb{E}^\mu [u]$  is not a convex function. On the other hand,  $\tilde{\beta}(\cdot)$  is a convex function. Indeed, for each  $i \in \Omega$  and  $u \in \mathcal{A}$ ,  $m_i \mapsto m_i^\top u_i$  is a linear function so that  $m \mapsto \tilde{\beta}(m)$  is a convex function since it is the supremum of linear functions indexed by  $u \in \mathcal{A}$ . Similarly,  $f_i(\cdot, \cdot, \cdot)$  is not a concave function in general. However,  $(m_i, \lambda_i) \mapsto \tilde{f}_i(m_i, \lambda_i)$  is the infimum of linear functions indexed by  $(x_i, y_i) \in \mathcal{F}_i$ ; therefore, it is a concave function.

**Theorem 5.11.** *It holds*

$$\begin{aligned} \mathcal{P}_2(v) &= \sup_{\mu \in \mathbb{M}_1^J, \lambda \in \mathbb{L}^M, \gamma \in \mathbb{R}_+^J} \left\{ \sum_{i \in \Omega} f_i(\mu_i, \lambda_i, \gamma) - \gamma^\top v - \beta(\mu, \gamma) \mid \gamma^\top \mathbf{1} = 1, \mathbb{E}[\lambda] = 0 \right\} \\ &= \sup_{m \in \mathbb{M}_f^J, \lambda \in \mathbb{L}^M} \left\{ \sum_{i \in \Omega} \tilde{f}_i(m_i, \lambda_i) - \sum_{i \in \Omega} m_i^\top v - \tilde{\beta}(m) \mid \sum_{i \in \Omega} m_i^\top \mathbf{1} = 1, \mathbb{E}[\lambda] = 0 \right\}. \end{aligned} \quad (D_2(v))$$

In view of Remark 5.10, while the first reformulation of  $(P'_2(v))$  provided in Theorem 5.11 is not a convex optimization problem, the second reformulation, that is  $(D_2(v))$ , is a convex optimization problem.

The proof of Theorem 5.11 uses Lemma 5.7 and the following lemma of independent interest.

**Lemma 5.12.** *For every  $u \in \mathbb{L}^J$ ,*

$$\inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(u) \} = \sup \{ \gamma^\top (\mathbb{E}^\mu [u] - v) - \beta(\mu, \gamma) \mid \mu \in \mathbb{M}_1^J, \gamma^\top \mathbf{1} = 1, \gamma \in \mathbb{R}_+^J \}.$$

**Proof.** Let  $u \in \mathbb{L}^J$ . Note that

$$\inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(u) \} = \inf \{ \alpha \in \mathbb{R} \mid 0 \in R(u) - v - \alpha \mathbf{1} \}$$

is the optimal value of a single-objective optimization problem with a set-valued constraint function  $\alpha \mapsto H(\alpha) = R(u) - v - \alpha \mathbf{1}$ . Using the Lagrange duality in Borwein (1981) for such problems, in particular, Theorem 19, we have

$$\inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(u) \} = \sup_{\gamma \in \mathbb{R}_+^J} \inf_{\alpha \in \mathbb{R}} \left( \alpha + \inf_{z \in R(u) - v - \alpha \mathbf{1}} \gamma^\top z \right). \quad (5.21)$$

To be able to use this result, we check the following constraint qualification:  $H$  is *open* at  $0 \in \mathbb{R}^J$  in the sense that for every  $\alpha \in \mathbb{R}$  with  $0 \in H(\alpha)$  and for every  $\varepsilon > 0$ , there exists an open ball  $V$  around  $0 \in \mathbb{R}^J$  such that

$$V \subseteq \bigcup_{\tilde{\alpha} \in (\alpha - \varepsilon, \alpha + \varepsilon)} H(\tilde{\alpha}). \quad (5.22)$$

To that end, let  $\alpha \in \mathbb{R}$  with  $0 \in H(\alpha)$ , that is,  $v + \alpha \mathbf{1} \in R(u)$ . Let  $\varepsilon > 0$ . Since  $\mathbf{1}$  is an interior point of  $\mathbb{R}_+^J$  and  $R(u) + \mathbb{R}_+^J = R(u)$  due to the monotonicity and translativity of  $R$ , it follows that  $v + (\alpha + \varepsilon) \mathbf{1}$  is an interior point of  $R(u)$ . On the other hand, note that

$$\begin{aligned} \bigcup_{\tilde{\alpha} \in (\alpha - \varepsilon, \alpha + \varepsilon)} H(\tilde{\alpha}) &= \bigcup_{\tilde{\alpha} \in (\alpha - \varepsilon, \alpha + \varepsilon)} R(u - v - \tilde{\alpha} \mathbf{1}) \\ &= R(u - v - (\alpha + \varepsilon) \mathbf{1}) \\ &= R(u) - v - (\alpha + \varepsilon) \mathbf{1} \end{aligned}$$

thanks to the monotonicity and translativity of  $R$ . Hence,  $0 \in \mathbb{R}^J$  is an interior point of the above union. Therefore, (5.22) holds for some open ball  $V$  around  $0 \in \mathbb{R}^J$  and (5.21) follows.

Since  $R(u) + \mathbb{R}_+^J = R(u)$  and  $R(u)$  is a convex set as a consequence of the convexity of  $R$ , one can check that  $\inf_{z \in R(u)} \gamma^\top z = -\infty$  for every  $\gamma \notin \mathbb{R}_+^J$ . Hence, the supremum in (5.21) can be evaluated over all  $\gamma \in \mathbb{R}_+^J$ . Finally, using (2.3), we obtain

$$\begin{aligned} &\inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(u) \} \\ &= \sup_{\gamma \in \mathbb{R}_+^J} \inf_{\alpha \in \mathbb{R}} \left( \alpha + \inf_{z \in R(u) - v - \alpha \mathbf{1}} \gamma^\top z \right) \\ &= \sup_{\gamma \in \mathbb{R}_+^J} \inf_{\alpha \in \mathbb{R}} \left( \alpha - \gamma^\top (v + \alpha \mathbf{1}) + \sup_{\mu \in \mathbb{M}_1^J} (\gamma^\top \mathbb{E}^\mu [u] - \beta(\mu, \gamma)) \right) \\ &= \sup_{\gamma \in \mathbb{R}_+^J} \left[ \inf_{\alpha \in \mathbb{R}} (1 - \gamma^\top \mathbf{1}) \alpha + \sup_{\mu \in \mathbb{M}_1^J} (\gamma^\top (\mathbb{E}^\mu [u] - v) - \beta(\mu, \gamma)) \right] \\ &= \sup \{ \gamma^\top (\mathbb{E}^\mu [u] - v) - \beta(\mu, \gamma) \mid \mu \in \mathbb{M}_1^J, \gamma^\top \mathbf{1} = 1, \gamma \in \mathbb{R}_+^J \}, \end{aligned}$$

where, in the last equality, we use the observation that  $\inf_{\alpha \in \mathbb{R}} (1 - \gamma^\top \mathbf{1}) \alpha = 0$  if  $\gamma^\top \mathbf{1} = 1$  and  $\inf_{\alpha \in \mathbb{R}} (1 - \gamma^\top \mathbf{1}) \alpha = -\infty$  if  $\gamma^\top \mathbf{1} \neq 1$ .  $\square$

**Proof of Theorem 5.11.** Using Lemma 5.12, we may write

$$\begin{aligned}\mathcal{P}_2(v) &= \inf_{(x,y) \in \mathcal{F}, \alpha \in \mathbb{R}} \{ \alpha \mid v + \alpha \mathbf{1} \in R(Cx + Qy), p_i(x_i - \mathbb{E}[x]) = 0 \forall i \in \Omega \} \\ &= \inf_{(x,y) \in \mathcal{F}} \{ \inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(Cx + Qy) \} \mid p_i(x_i - \mathbb{E}[x]) = 0 \forall i \in \Omega \} \\ &= \inf_{(x,y) \in \mathcal{F}} \left\{ \sup_{\mu \in \mathbb{M}_1^J, \gamma \in \mathbb{R}_+^J, \gamma^\top \mathbf{1} = 1} (\gamma^\top (\mathbb{E}^\mu [Cx + Qy] - v) - \beta(\mu, \gamma)) \mid p_i(x_i - \mathbb{E}[x]) = 0 \forall i \in \Omega \right\}.\end{aligned}$$

Using the minimax theorem of Sion (1958), we may interchange the infimum and the supremum, and obtain

$$\mathcal{P}_2(v) = \sup_{\mu \in \mathbb{M}_1^J, \gamma \in \mathbb{R}_+^J, \gamma^\top \mathbf{1} = 1} (F(\mu, \gamma) - \gamma^\top v - \beta(\mu, \gamma)),$$

where, for each  $\mu \in \mathbb{M}_1^J$ ,  $\gamma \in \mathbb{R}_+^J$ ,  $F(\mu, \gamma)$  is defined by (5.7). Hence, using (5.8), we obtain

$$\mathcal{P}_2(v) = \sup_{\mu \in \mathbb{M}_1^J, \lambda \in \mathbb{L}^M, \gamma \in \mathbb{R}_+^J} \left\{ \sum_{i \in \Omega} f_i(\mu_i, \lambda_i, \gamma) - \gamma^\top v - \beta(\mu, \gamma) \mid \gamma^\top \mathbf{1} = 1, \mathbb{E}[\lambda] = 0 \right\},$$

where  $f_i$  is defined by (5.2). Hence, the first reformulation follows. The second reformulation follows from the first reformulation and Lemma 5.7.  $\square$

## 5.2.2 The dual bundle method

To solve  $(D_2(v))$  provided in Theorem 5.11, we propose a dual bundle method similar to the one in Section 5.1.2.

At each iteration  $k$  of the dual bundle method, we solve the master problem  $(MP_2(v))$  given below. Here,  $(g_{m_i}, g_{\lambda_i}) \in \mathbb{R}^{J+M}$  denotes a subgradient of the concave function  $\tilde{f}_i(\cdot, \cdot)$  at the point  $(m_i, \lambda_i) \in \mathbb{R}_+^J \times \mathbb{R}^M$ . Similarly,  $(\rho_{m_1}, \dots, \rho_{m_I}) \in \mathbb{R}^{J \times I}$  denotes a subgradient of the concave function  $-\tilde{\beta}(\cdot)$  at the point  $m \in \mathbb{M}_1^J$ . We call (5.23) and (5.24) a cut for  $\tilde{f}_i(\cdot, \cdot)$  and  $-\tilde{\beta}(\cdot)$ , respectively.

$$\max \quad \sum_{i \in \Omega} \vartheta_i + \eta - \sum_{i \in \Omega} m_i^\top v - \sum_{i \in \Omega} \rho \left\| m_i - \bar{m}_i^{(k)} \right\|^2 - \sum_{i \in \Omega} \rho \left\| \lambda_i - \bar{\lambda}_i^{(k)} \right\|^2 \quad (MP_2(v))$$

$$\text{s.t.} \quad \vartheta_i \leq \tilde{f}_i(m_i^{(\ell)}, \lambda_i^{(\ell)}) + g_{m_i^{(\ell)}}^\top (m_i - m_i^{(\ell)}) + g_{\lambda_i^{(\ell)}}^\top (\lambda_i - \lambda_i^{(\ell)}), \quad \forall i \in \Omega, \ell \in \mathcal{L} \quad (5.23)$$

$$\eta \leq -\tilde{\beta}(m^{(\ell)}) + \sum_{i \in \Omega} \rho_{m_i^{(\ell)}}^\top (m_i - m_i^{(\ell)}), \quad \forall \ell \in \mathcal{L} \quad (5.24)$$

$$\sum_{i \in \Omega} p_i \lambda_i = 0 \quad (5.25)$$

$$\sum_{i \in \Omega} m_i^\top \mathbf{1} = 1 \quad (5.26)$$

$$m_i \in \mathbb{R}_+^J, \lambda_i \in \mathbb{R}^M, \vartheta_i \in \mathbb{R}, \quad \forall i \in \Omega \quad (5.27)$$

$$\eta \in \mathbb{R}. \quad (5.28)$$

Here,  $\mathcal{L} = \{1, \dots, k\}$ ,  $\rho > 0$ .

The steps of the dual bundle method are provided in Algorithm 4. Similar to (5.17), the algorithm stops in practice when

$$\sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} - \sum_{i \in \Omega} (m_i^{(k+1)})^\top v - \bar{F}^{(k+1)} \leq \varepsilon \quad (5.29)$$

for some  $\varepsilon > 0$ . Since the construction of this algorithm is similar to the construction of Algorithm 3, the details are omitted for brevity.

---

**Algorithm 4** A Dual Bundle Method for  $(P_2(v))$ 


---

- 1:  $k \leftarrow 0, \mathcal{L} \leftarrow \emptyset, \gamma \in (0, 1), \vartheta_i^{(1)} \leftarrow \infty$  for each  $i \in \Omega, \eta^{(1)} \leftarrow \infty, \bar{F}^{(1)} \leftarrow 0$ ;
  - 2: Let  $m^{(1)} \in \mathbb{M}_f^J, \lambda^{(1)} \in \mathbb{L}^M$  be such that  $\mathbb{E}[\lambda^{(1)}] = 0$ ;
  - 3: **repeat**
  - 4:    $k \leftarrow k + 1$ ;
  - 5:   **for** each  $i \in \Omega$  **do**
  - 6:     Compute an optimal solution  $(x_i^{(k)}, y_i^{(k)})$  and the optimal value  $\tilde{f}_i(m_i^{(k)}, \lambda_i^{(k)})$  of the subproblem
 
$$\min_{(x_i, y_i) \in \mathcal{F}_i} \left( (m_i^{(k)})^\top (Cx_i + Q_i y_i) + p_i (\lambda_i^{(k)})^\top x_i \right);$$
  - 7:     Compute subgradients  $g_{m_i^{(k)}} = Cx_i^{(k)} + Q_i y_i^{(k)}, g_{\lambda_i^{(k)}} = p_i x_i^{(k+1)}$ ;
  - 8:   **end for**
  - 9:   Compute  $\tilde{\beta}(m^{(k)})$  and subgradient  $(\rho_{m_1^{(k)}}, \dots, \rho_{m_I^{(k)}}) \in \partial_m(-\tilde{\beta})(m^{(k)})$ ;
  - 10:    $F^{(k)} \leftarrow \sum_{i \in \Omega} \tilde{f}_i(m_i^{(k)}, \lambda_i^{(k)}) - \tilde{\beta}(m^{(k)}) - \sum_{i \in \Omega} (m_i^{(k)})^\top v$ ;
  - 11:   **if**  $F^{(k)} < \sum_{i \in \Omega} \vartheta_i^{(k)} + \eta^{(k)} - \sum_{i \in \Omega} (m_i^{(k)})^\top v$  **then**
  - 12:      $\mathcal{L} \leftarrow \mathcal{L} \cup \{k\}$ ;
  - 13:   **end if**
  - 14:   **if**  $(k = 1)$  or  $(k \geq 2$  and  $F^{(k)} \geq (1 - \gamma)\bar{F}^{(k)} + \gamma(\sum_{i \in \Omega} \vartheta_i^{(k)} + \eta^{(k)} - \sum_{i \in \Omega} (m_i^{(k)})^\top v)$ ) **then**
  - 15:      $\bar{m}^{(k)} \leftarrow m^{(k)}, \bar{\lambda}^{(k)} \leftarrow \lambda^{(k)}$ ;
  - 16:   **else**
  - 17:      $\bar{m}^{(k)} \leftarrow \bar{m}^{(k-1)}, \bar{\lambda}^{(k)} \leftarrow \bar{\lambda}^{(k-1)}$ ;
  - 18:   **end if**
  - 19:   Solve the master problem. Let  $(m^{(k+1)}, \lambda^{(k+1)}, \vartheta^{(k+1)}, \eta^{(k+1)})$  be an optimal solution;
  - 20:   (Optional) Remove all cuts whose dual variables at the solution of master problem are zero;
  - 21:    $\bar{F}^{(k+1)} \leftarrow \sum_{i \in \Omega} \tilde{f}_i(\bar{m}_i^{(k)}, \bar{\lambda}_i^{(k)}) - \tilde{\beta}(\bar{m}^{(k)}) - \sum_{i \in \Omega} (\bar{m}_i^{(k)})^\top v$ ;
  - 22: **until**  $\sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} - \sum_{i \in \Omega} (m_i^{(k+1)})^\top v = \bar{F}^{(k+1)}$ ;
  - 23: **return**  $\begin{cases} \bar{F}^{(k+1)} & : \text{Optimal value } \mathcal{P}_2(v); \\ (\bar{m}^{(k)}, \bar{\lambda}^{(k)}) & : \text{An optimal solution of } (D_2(v)); \end{cases}$
- 

Next, we provide a recipe for computing the subgradients  $g_{m_i}, g_{\lambda_i}, \rho_m$ . Let us denote by  $\partial_{m_i, \lambda_i} \tilde{f}_i(m'_i, \lambda'_i)$  the subdifferential of the function  $\tilde{f}_i(\cdot, \cdot)$  at a point  $(m'_i, \lambda'_i) \in \mathbb{R}_+^J \times \mathbb{R}^M$ , and by  $\partial_m(-\tilde{\beta})(m')$  the subdifferential of the function  $-\tilde{\beta}(\cdot)$  at a point  $m' \in \mathbb{L}_+^J$ . In the next proposition, we show how to compute  $\partial_{m_i, \lambda_i} \tilde{f}_i(m'_i, \lambda'_i)$  and a subgradient  $\rho_m$  of the function  $-\tilde{\beta}(\cdot)$ .

**Proposition 5.13.** (a) For  $i \in \Omega, m'_i \in \mathbb{R}_+^J, \lambda'_i \in \mathbb{R}^M$ , let

$$\tilde{A}_i(m'_i, \lambda'_i) := \arg \min_{(x_i, y_i) \in \mathcal{F}_i} \left( (m'_i)^\top (Cx_i + Q_i y_i) + p_i (\lambda'_i)^\top x_i \right).$$

Then,

$$\partial_{m_i, \lambda_i} \tilde{f}_i(m'_i, \lambda'_i) = \text{co} \left\{ (Cx_i + Q_i y_i, p_i x_i) \mid (x_i, y_i) \in \tilde{A}_i(m'_i, \lambda'_i) \right\}.$$

(b) Recall that the set  $\mathcal{A} = \{u \in \mathbb{L}^J \mid 0 \in R(u)\}$  is the acceptance set of  $R$ . For  $m' \in \mathbb{L}_+^J$ , let

$$\tilde{B}(m') := \arg \max_{u \in \mathcal{A}} \sum_{i \in \Omega} (m'_i)^\top u_i$$

and assume that  $\tilde{B}(m') \neq \emptyset$ . Then,

$$\partial_m(-\tilde{\beta})(m') \supseteq \text{co} \left\{ u = (u_1, \dots, u_I) \mid u \in \tilde{B}(m') \right\}.$$

**Proof.** The proof of this proposition is similar to the proofs of Propositions 5.2 and 5.3. Therefore, it is omitted.  $\square$

### 5.2.3 Recovery of primal solution

The primal Benson algorithm requires an optimal solution  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  of the problem  $(P'_2(v))$ . Therefore, in Theorem 5.14, we suggest a procedure to recover an optimal primal solution from the solution of the master problem  $(MP_2(v))$ .

**Theorem 5.14.** *Let  $\mathcal{L} = \{1, \dots, k\}$  be the index set at the last iteration of the dual bundle method with the approximate stopping condition (5.29) for some  $\varepsilon > 0$ . Let  $n + 1$  be the first descent iteration after the approximate stopping condition is satisfied and let  $\mathcal{L}' = \{1, \dots, n\}$ . For  $(MP_2(v))$  with centers  $\bar{m}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ , let  $\tau = (\tau_i^{(\ell)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta = (\theta^{(\ell)})_{\ell \in \mathcal{L}'}, \sigma \in \mathbb{R}^M, \psi \in \mathbb{R}, \nu = (\nu_i)_{i \in \Omega}$  be the Lagrangian dual variables assigned to the constraints (5.23), (5.24), (5.25), (5.26), (5.27), respectively, with  $\tau_i^{(\ell)} \geq 0, \theta^{(\ell)} \geq 0, \nu_i \in \mathbb{R}_+^J$  for each  $i \in \Omega, \ell \in \mathcal{L}'$ . Let  $(x_i^{(\ell)}, y_i^{(\ell)})$  be an optimal solution of the subproblem in line 6 of Algorithm 4 for each  $i \in \Omega$  and  $\ell \in \mathcal{L}'$ . Let*

$$\left( \tau^{(n+1)} = (\tau_i^{(\ell, n+1)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta^{(n+1)} = (\theta^{(\ell, n+1)})_{\ell \in \mathcal{L}'}, \sigma^{(n+1)}, \psi^{(n+1)}, \nu^{(n+1)} = (\nu_i^{(n+1)})_{i \in \Omega} \right)$$

be a dual optimal solution for  $(MP_2(v))$ . Let  $x_{(v)} = ((x_{(v)})_i)_{i \in \Omega}, y_{(v)} = ((y_{(v)})_i)_{i \in \Omega}$  be defined by

$$(x_{(v)})_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} x_i^{(\ell)}, \quad (y_{(v)})_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} y_i^{(\ell)}.$$

Let

$$\alpha_{(v)} := \inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(C\bar{x} + Q\bar{y}) \}.$$

Then,  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$  is an approximately optimal solution of  $(P'_2(v))$  in the following sense:

- (a)  $((x_{(v)})_i, (y_{(v)})_i) \in \mathcal{F}_i$  for each  $i \in \Omega$ .
- (b)  $v + \alpha_{(v)} \mathbf{1} \in R(Cx_{(v)} + Qy_{(v)})$ .
- (c) As  $\varepsilon \rightarrow 0$ , it holds  $(x_{(v)})_i - \sigma^{(n+1)} \rightarrow 0$  for each  $i \in \Omega$ .
- (d) As  $\varepsilon \rightarrow 0$ , it holds  $\alpha_{(v)} \rightarrow \mathcal{P}_2(v)$ .

The proof of Theorem 5.14 is given in Appendix B.

### 5.2.4 Recovery of a solution to $(LD_2(v))$

In addition to a primal optimal solution  $(x_{(v)}, y_{(v)}, \alpha_{(v)})$ , the primal Benson algorithm also requires an optimal solution  $\gamma_{(v)}$  of the dual problem  $(LD_2(v))$  (see Section 4.1). Therefore, in Theorem 5.15, we suggest a procedure to recover this solution from the solution of the master problem  $(MP_2(v))$ .

**Theorem 5.15.** *In the setting of Theorem 5.14, let*

$$\gamma_{(v)} = \sum_{i \in \Omega} m_i^{(n+1)}. \tag{5.30}$$

Then,  $\gamma_{(v)}$  is an approximately optimal solution of  $(LD_2(v))$  in the following sense: as  $\varepsilon \rightarrow 0$ , it holds

$$\left| \inf_{(x, y) \in \mathcal{X}, \alpha \in \mathbb{R}} \left( \alpha + \inf_{z \in R(Cx + Qy) - v - \alpha \mathbf{1}} \gamma_{(v)}^\top z \right) - \mathcal{P}_2(v) \right| \rightarrow 0. \tag{5.31}$$

The proof of Theorem 5.15 is given in Appendix C.

## 6 Computational Study

In order to test our methods, we solve a multi-objective risk-averse portfolio optimization problem under transaction costs. We consider a one-period market with  $J$  risky assets. Each asset  $j \in \mathcal{J} = \{1, \dots, J\}$  has a random return  $r^j \in \mathbb{L}$ . At the beginning of the period, it costs  $\theta^{jk} \in \mathbb{R}$  units of asset  $j$  for an agent to buy one unit of asset  $k \in \mathcal{J}$ . At the end of the period, the random transaction cost of buying one unit of asset  $k$  is  $\pi^{jk} \in \mathbb{L}$  units of asset  $j$ .

The risk-averse agent has a capital  $c \in \mathbb{R}_{++}$  units of asset 1 to be invested in the  $J$  assets. Let  $x^j \in \mathbb{R}_+$  denote the number of physical units of asset  $j$  purchased by the agent; hence, she spends  $x^j \theta^{1j}$  units of asset 1 for this purchase. At the end of the period, the agent observes the random return of each asset as well as the random transaction costs between the assets. The value of each asset  $j$  is  $(1 + r^j)x^j$  and it is transacted to purchase the  $J$  assets with a transaction cost of  $\pi^{jk}$  for asset  $k$ . Let  $q^{jk} \in \mathbb{L}_+$  denote the number of physical units of asset  $k$  purchased by selling some units of asset  $j$ . Let  $y^k \in \mathbb{L}_+$  denote the total number of physical units of asset  $k$  purchased by the agent so that  $y^k = \sum_{j \in \mathcal{J}} q^{jk}$ . The objective is to minimize the risk of the random cost vector  $-y \in \mathbb{L}^J$  using a multivariate convex risk measure  $R$ . This problem can be formulated as follows:

$$\begin{aligned}
 \min \quad & z \text{ w.r.t. } \mathbb{R}_+^J \\
 \text{s.t.} \quad & z \in R(-y) \\
 & \sum_{j \in \mathcal{J}} \theta^{1j} x^j = c \\
 & (1 + r_i^j) x^j = \sum_{k \in \mathcal{J}} \pi_i^{jk} q_i^{jk}, \quad \forall j \in \mathcal{J}, i \in \Omega \\
 & y_i^j = \sum_{k \in \mathcal{J}} q_i^{kj}, \quad \forall j \in \mathcal{J}, i \in \Omega \\
 & z \in \mathbb{R}^J, x \in \mathbb{R}_+^J, y_i \in \mathbb{R}_+^J, q_i \in \mathbb{R}_+^{J \times J}, \quad \forall i \in \Omega.
 \end{aligned}$$

Note that  $x \in \mathbb{R}_+^J$  is the first-stage,  $y_i \in \mathbb{R}_+^J$ ,  $q_i \in \mathbb{R}_+^{J \times J}$ ,  $i \in \Omega$  are the second-stage decision variables.

All computational experiments are conducted on a PC with 8.00 GB of RAM and an Intel(R) Core(TM) i7-4790 CPU@3.60 GHz processor. We use Matlab implementations of Algorithms 3 and 4 where CVX 1.22 is used to solve master problems and CPLEX 12.6 is used to solve subproblems.

We generate two classes of instances where the number of assets  $J$  is either 2 or 3. In both cases, we assume  $c = 1$ . We set  $\theta^{12} = 1.0815$ ,  $\theta^{13} = 0.9094$ . The return of asset 1 is uniformly distributed between  $-0.1$  and  $0.2$ , denoted by  $r^1 \sim U[-0.1, 0.2]$ . Similarly, we assume  $r^2 \sim U[-0.05, 0.1]$  and  $r^3 \sim U[-0.15, 0.3]$ . The random transaction costs among the assets are assumed to have the following distributions:  $\pi^{12} \sim U[1, 1.1]$ ,  $\pi^{21} \sim U[0.9, 1]$ ,  $\pi^{13} \sim U[0.9, 1]$ ,  $\pi^{31} \sim U[1, 1.1]$ ,  $\pi^{23} \sim U[0.8, 1]$ ,  $\pi^{32} \sim U[1, 1.2]$ , and  $\pi^{11} = \pi^{22} = \pi^{33} = 1$ .

First of all, in Example 6.1, we compare our dual bundle method with CVX on the problem  $(P_1(w))$  of weighted sum scalarization. Our dual bundle method takes the advantage of scenario-wise decompositions while CVX solves the problem as a standard convex optimization problem without decompositions.

**Example 6.1.** We compare the CPU times (in seconds) of the dual bundle method and the CVX solver on  $(P_1(w))$  instances with two- and three-dimensional multivariate entropic risk measure and different numbers of scenarios ( $I$ ). In each instance, we use a fixed weight vector  $w$ .

As observed in Table 1 and Table 2, the CVX solver overperforms the dual bundle method for smaller numbers of scenarios. However, as the number of scenarios increases, the dual bundle method overperforms the CVX solver. For instance, for  $I = 10000$  in Table 1, the CVX solver cannot solve the problem due to a memory error. The same situation is observed for  $I = 500$  in Table 2.

For the remainder of this section, we use the multivariate CVaR (*see* Example 2.2) and the multivariate entropic risk measure (*see* Example 2.3) for the choice of  $R$ . For each risk measure, we consider the biobjective ( $J = 2$  assets) and the three-objective ( $J = 3$  assets) cases. We run the primal and dual Benson algorithms with different error parameters ( $\epsilon$ ) and report the total number of scalar optimization problems solved

$I$	Dual Bundle Method	CVX
1000	869.22	75.98
2500	2130.85	588.85
5000	4170.55	3091.44
10000	8452.47	**

Table 1: Computational performances of the dual bundle method and CVX for a two-dimensional multivariate entropic risk measure instance with weight vector  $w = (1/2, 1/2)$

$I$	Dual Bundle Method	CVX
50	56.39	35.22
100	252.73	98.73
250	1838.38	346.87
500	6309.39	**

Table 2: Computational performances of the dual bundle method and CVX for a three-dimensional multivariate entropic risk measure instance with weight vector  $w = (1/3, 1/3, 1/3)$

(#opt.), the number of vertices in the final outer approximation (#vert.) and the CPU time in seconds (time).

**Example 6.2.** (Two-dimensional multivariate CVaR) We consider  $J = 2$  assets under  $I = 500$  scenarios. The parameters of the multivariate CVaR are chosen as  $\nu^1 = 0.8, \nu^2 = 0.9$ . We use error parameter values  $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ . The computational results are reported in Table 3. It can be seen that the performances of the primal and dual algorithms are close to each other.

The inner (red lines) and outer (blue lines) approximations of the upper image  $\mathcal{P}$  and the lower image  $\mathcal{D}$  are given in Figures 1 and 2. These figures are obtained by the primal algorithm. Since the corresponding figures for the dual algorithm are similar, they are omitted. Clearly, the algorithm provides finer approximations for the upper and lower images when  $\epsilon$  is reduced from  $10^{-3}$  to  $10^{-4}$ .

	$\epsilon$	#opt.	#vert.	time
Primal Algorithm	$10^{-2}$	5	3	2675.69
	$10^{-3}$	11	6	10513.06
	$10^{-4}$	23	13	11391.12
Dual Algorithm	$10^{-2}$	5	4	2819.92
	$10^{-3}$	13	8	7021.55
	$10^{-4}$	25	15	10007.75

Table 3: Computational results for the two-dimensional multivariate CVaR



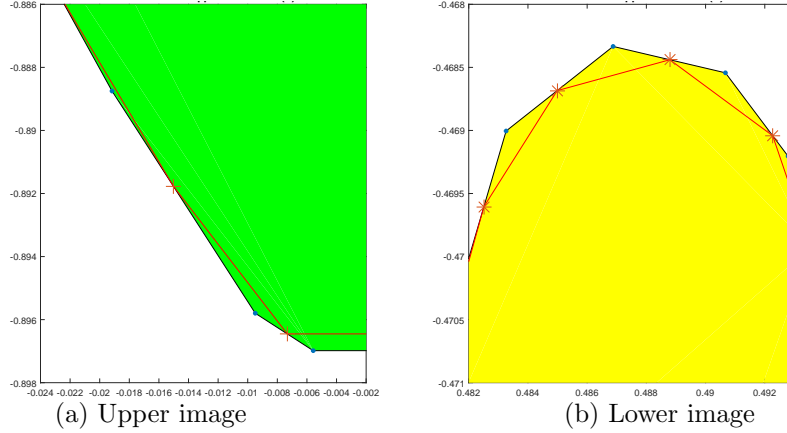


Figure 1: Inner and outer approximations obtained by the primal algorithm for  $\epsilon = 10^{-3}$

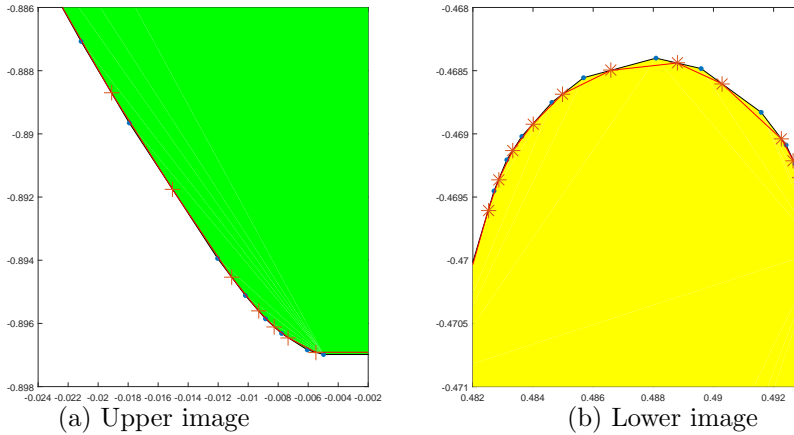


Figure 2: Inner and outer approximations obtained by the primal algorithm for  $\epsilon = 10^{-4}$

**Example 6.3.** (Two-dimensional multivariate entropic risk measure) We consider  $J = 2$  assets under  $I = 500$  scenarios. The parameters of the multivariate entropic risk measure are chosen as  $\delta^1 = \delta^2 = 0.1$  and the cone  $C$  is generated by the vectors  $(2, 1)$  and  $(1, 2)$ . We use error parameter values  $\epsilon \in \{0.1, 0.05, 0.01\}$ .

The computational results are reported in Table 4. In this example, the dual algorithm solves more optimization problems and enumerates more vertices than the primal algorithm in significantly shorter time. The inner and outer approximations of the upper image  $\mathcal{P}$  and the lower image  $\mathcal{D}$  obtained by the primal algorithm are given in Figures 3 and 4. Since the corresponding figures for the dual algorithm are similar, they are omitted.

	$\epsilon$	#opt.	#vert.	time
Primal Algorithm	0.1	25	13	37706.90
	0.05	37	19	84730.81
	0.01	83	42	144848.62
Dual Algorithm	0.1	31	17	13955.43
	0.05	47	25	14088.08
	0.01	85	44	17121.26

Table 4: Computational results for the two-dimensional multivariate entropic risk measure

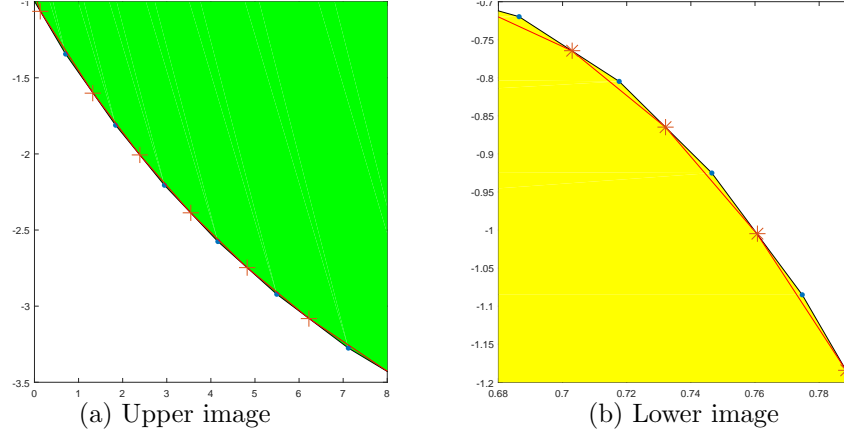


Figure 3: Inner and outer approximations obtained by the primal algorithm for  $\epsilon = 0.05$

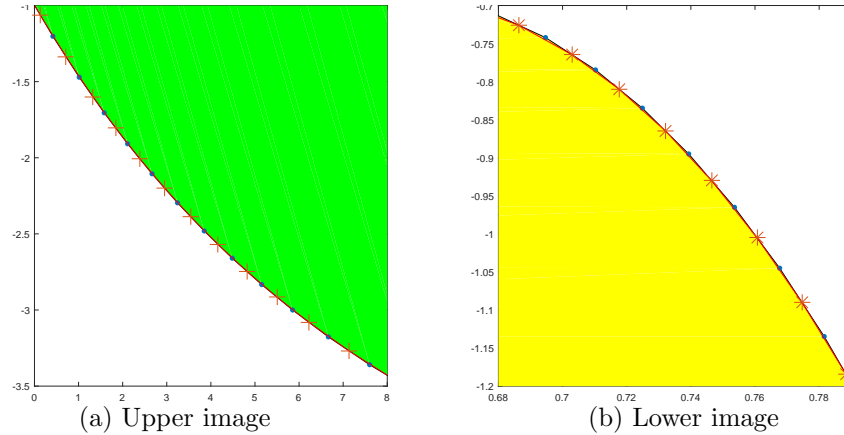


Figure 4: Inner and outer approximations obtained by the primal algorithm for  $\epsilon = 0.01$

**Example 6.4.** (Three-dimensional multivariate CVaR) We consider  $J = 3$  assets under  $I = 250$  scenarios. The parameters of the multivariate CVaR are chosen as  $\nu^1 = 0.8, \nu^2 = 0.9$ . We use error parameter values  $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ .

The computational results are reported in Table 5. For  $\epsilon = 10^{-2}$  and  $\epsilon = 10^{-3}$ , the primal algorithm terminates in shorter time while, for  $\epsilon = 10^{-4}$ , the dual algorithm is faster.

The outer approximations of the upper image  $\mathcal{P}$  and the lower image  $\mathcal{D}$  obtained by the primal and dual algorithms are given in Figures 5-7. Note that the dots represent the vertices of some polyhedra even if they are not connected by line segments.

As observed in these figures, the primal algorithm provides a better approximation of the lower image compared to the dual algorithm. However, the approximation of the upper image provided by the dual algorithm is better than the one by the primal algorithm.

The multivariate CVaR is defined in terms of the positive part function  $(\cdot)^+$ , which is piecewise linear. As a result, the upper and lower images of the problem are polyhedral sets and the vertices of their outer approximations in Figures 5-7 are generally dense around certain line segments.

	$\epsilon$	#opt.	#vert.	time
Primal Algorithm	$10^{-2}$	21	9	16856.84
	$10^{-3}$	82	32	67555.09
	$10^{-4}$	468	162	319862.68
Dual Algorithm	$10^{-2}$	24	11	20303.43
	$10^{-3}$	98	36	79397.49
	$10^{-4}$	448	152	249081.86

Table 5: Computational results for the three-dimensional multivariate CVaR

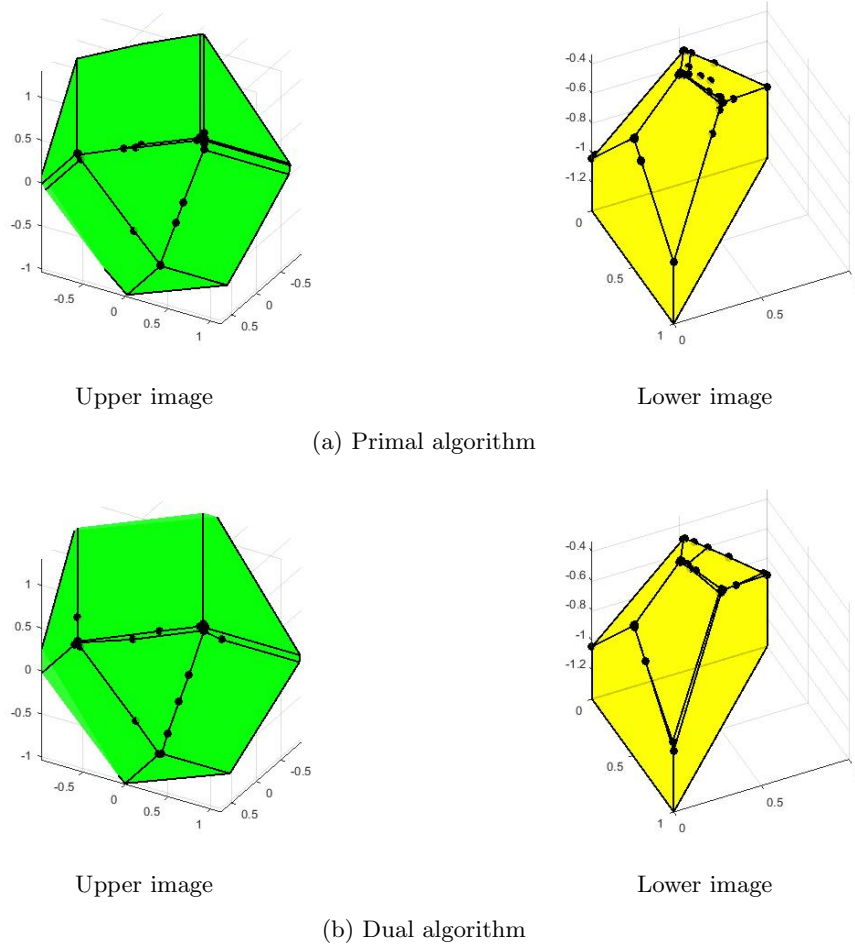
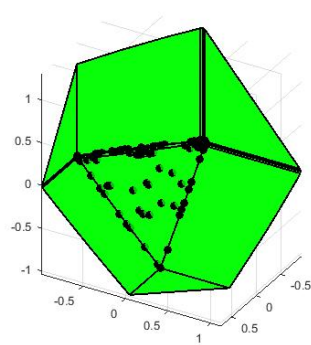
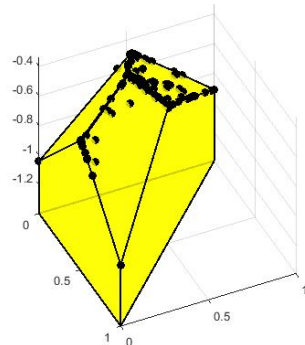


Figure 5: Outer approximations obtained by the primal and dual algorithms for  $I = 250$  and  $\epsilon = 10^{-2}$

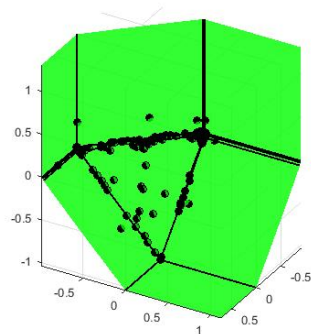


Upper image

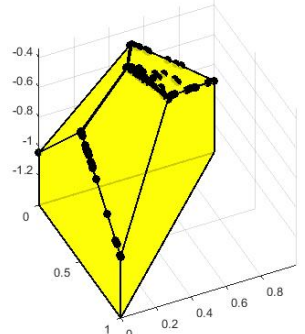


Lower image

(a) Primal algorithm



Upper image



Lower image

(b) Dual algorithm

Figure 6: Outer approximations obtained by the primal and dual algorithms for  $I = 250$  and  $\epsilon = 10^{-3}$

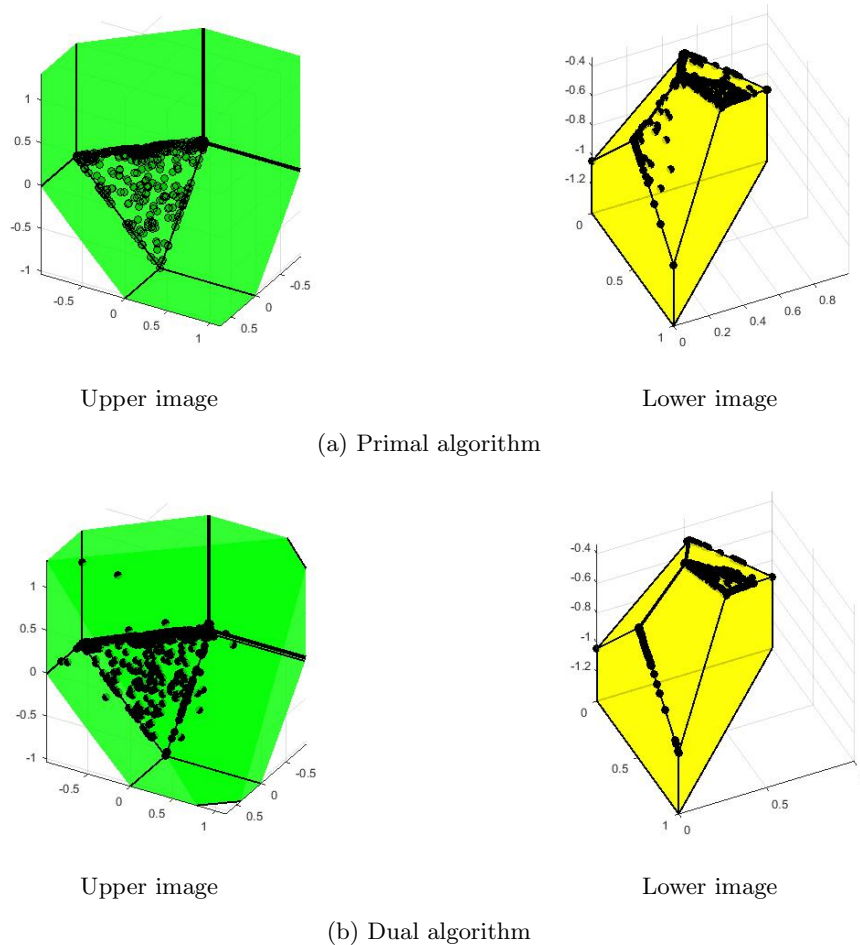


Figure 7: Outer approximations obtained by the primal and dual algorithms for  $I = 250$  and  $\epsilon = 10^{-4}$

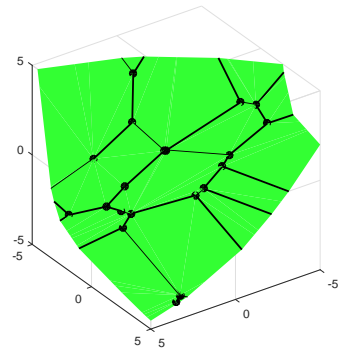
**Example 6.5.** (Three-dimensional multivariate entropic risk measure) We consider  $J = 3$  assets under  $I = 100$  scenarios. The parameters of the multivariate entropic risk measure are chosen as  $\delta^1 = \delta^2 = \delta^3 = 0.1$  and the cone  $C$  is generated by the vectors  $(1, 2, 3), (3, 2, 1)$ . We use error parameter values  $\epsilon \in \{0.1, 0.05, 0.01\}$ .

The computational results are reported in Table 4. We are not able to solve this problem using the primal algorithm as the dual bundle method for  $(P_2(v))$  does not converge for some vertices  $v$ . This is in line with what is reported in Löhne *et al.* (2014, Example 5.4) for a four-objective problem with multivariate entropic risk measure. The results of the dual algorithm are provided in Table 6 and Figure 8.

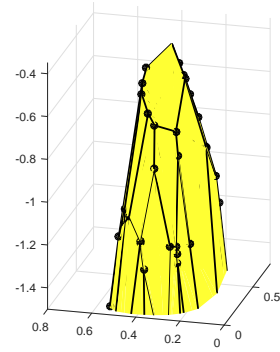
As the multivariate entropic risk measure is defined in terms of the exponential utility function, which is strictly convex, the upper and lower images are non-polyhedral sets. For this reason, the polyhedral outer approximations of these sets have a more uniform density of vertices over their surfaces compared to the outer approximations for the multivariate CVaR.

	$\epsilon$	#opt.	#vert.	time
Dual Algorithm	0.1	196	61	48742.57
	0.05	319	99	82237.89
	0.01	670	211	168460.45

Table 6: Computational results for the three-dimensional multivariate entropic risk measure

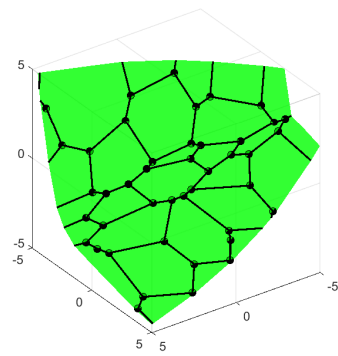


Upper image

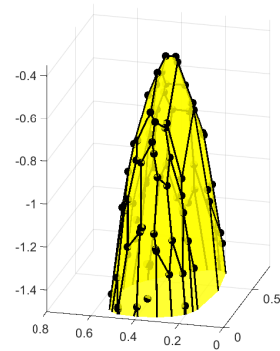


Lower image

(a)  $\epsilon = 0.1$

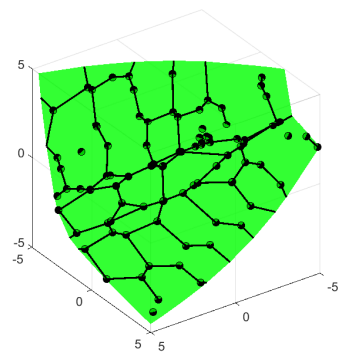


Upper image

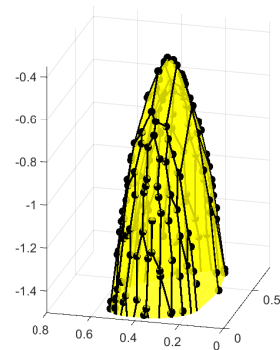


Lower image

(b)  $\epsilon = 0.05$



Upper image



Lower image

(c)  $\epsilon = 0.01$

Figure 8: Outer approximation obtained by the dual algorithm for  $I = 100$

## References

M. Abbas, F. Bellahcene, *Cutting plane method for multiple objective stochastic integer linear programming*, European Journal of Operational Research, **168**(3), 967–984 (2006)

- S. Ahmed, *Convexity and decomposition of mean-risk stochastic programs*, Mathematical Programming, **106**(3), 433–446 (2006)
- C. D. Aliprantis, K. C. Border, *Infinite dimensional analysis: a hitchhiker’s guide*. Springer, Heidelberg, Germany (2006)
- Ç. Ararat, A. H. Hamel, B. Rudloff, *Set-valued shortfall and divergence risk measures*, International Journal of Theoretical and Applied Finance, doi: 10.1142/S0219024917500261 (2017)
- H. P. Benson, *An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem*, Journal of Global Optimization, **13**(1), 1–24 (1998)
- J. R. Birge, F. V. Louveaux. *A multicut algorithm for two-stage stochastic linear programs*, European Journal of Operational Research, **34**(3), 384–392 (1988)
- J. R. Birge, F. V. Louveaux, *Introduction to Stochastic Programming*. Springer, New York, USA (1997)
- J. M. Borwein, *Convex relations in analysis and optimization*, in: S. Schaible, W. T. Ziemba (eds.) Generalized Concavity in Optimization and Economics, Academic Press Inc., New York, 335–377 (1981)
- Y. Cardona-Valdés, A. Álvarez, D. Özdemir, *A bi-objective supply chain design problem with uncertainty*, Transportation Research Part C: Emerging Technologies, **19**(5), 821–832 (2011)
- R. A. Collado, D. Papp, A. Ruszczyński, *Scenario decomposition of risk-averse multistage stochastic programming problems*, Annals of Operations Research, **200**(1), 147–170 (2012)
- D. Dentcheva, A. Ruszczyński, *Optimization with multivariate stochastic dominance constraints*, Mathematical Programming, **117**(1), 111–127 (2009)
- D. Dentcheva, E. Wolfhagen, *Two-stage optimization problems with multivariate stochastic order constraints*, Mathematics of Operations Research, **41**(1), 1–22 (2016)
- M. Ehrgott, L. Shao, A. Schöbel, *An approximation algorithm for convex multi-objective programming problems*, Journal of Global Optimization, **50**(3), 397–416 (2011)
- C. I. Fábián, *Handling CVaR objectives and constraints in two-stage stochastic models*, European Journal of Operational Research, **191**(3), 888–911 (2008)
- H. Föllmer, A. Schied, *Convex measures of risk and trading constraints*, Finance and Stochastics, **6**(4), 429–447 (2002)
- W. J. Gutjahr, A. Pichler, *Stochastic multi-objective optimization: a survey on non-scalarizing methods*, Annals of Operations Research, **236**(2) 475–499 (2016)
- A. H. Hamel, F. Heyde, *Duality for set-valued measures of risk*, SIAM Journal on Financial Mathematics, **1**(1), 66–95 (2010)
- A. H. Hamel, F. Heyde, A. Löhne, B. Rudloff and C. Schrage, *Set optimization - a rather short introduction*, In: (ed.) A. Hamel, F. Heyde, A. Löhne, B. Rudloff and C. Schrage, *Set optimization and applications in finance - the state of the art*, Proceedings in Mathematics & Statistics, Springer, 65–141 (2016)
- A. H. Hamel, B. Rudloff, M. Yankova, *Set-valued average value at risk and its computation*, Mathematics and Financial Economics, **7**(2), 229–246 (2013)
- J. Jahn, *Vector Optimization - Theory, Applications, and Extensions*. Springer, Berlin Heidelberg, Germany (2004)
- P. Kall, J. Mayer, *Stochastic Linear Programming*. Springer, New York, USA (2005)
- T. K. Kristoffersen, *Deviation measures in linear two-stage stochastic programming*, Mathematical Methods of Operations Research, **62**(2), 255–274 (2005)

- C. Lemaréchal, *Nonsmooth optimization and descent methods*, Research Report 78-2, International Institute for Applied Systems Analysis, Laxenburg (1978)
- X. Liu, S. Küçükyavuz, J. Luedtke, *Decomposition algorithms for two-stage chance-constrained programs*, *Mathematical Programming*, **157**(1), 219–243 (2016)
- A. Löhne, B. Rudloff, F. Ulus, *Primal and dual approximation algorithms for convex vector optimization problems*, *Journal of Global Optimization*, **60**(4), 713–736 (2014)
- A. Löhne, C. Schrage, *An algorithm to solve polyhedral convex set optimization problems*, *Optimization*, **62**(1), 131–141 (2013)
- N. Miller, A. Ruszczyński, *Risk-averse two-stage stochastic linear programming: modeling and decomposition*, *Operations Research*, **59**(1), 125–132 (2011)
- N. Noyan, M. Meraklı, S. Küçükyavuz, *Two-stage stochastic programming under multivariate risk constraints with an application to humanitarian relief network design*, preprint, (2017)
- R. T. Rockafellar, S. Uryasev, *Optimization of conditional value-at-risk*, *Journal of Risk*, **2**(3), 21–41 (2000)
- R. T. Rockafellar, S. Uryasev, *Conditional value-at-risk for general loss distributions*, *Journal of Banking & Finance*, **26**(7), 1443–1471 (2002)
- A. Ruszczyński, *Decomposition methods* in A. Ruszczyński, A. Shapiro, eds. *Stochastic Programming*. Elsevier, Amsterdam, the Netherlands (2003)
- A. Ruszczyński, *Nonlinear optimization*. Princeton University Press, Princeton, New Jersey, USA (2006)
- A. Shapiro, D. Dentcheva, A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM Series on Optimization, No. 9, MPS-SIAM, Philadelphia, Pennsylvania, USA (2009)
- M. Sion, *On general minimax theorems*, *Pacific Journal of Mathematics*, **8**(1), 171–176 (1958)
- R. Van Slyke, R. J.-B. Wets, *L-shaped linear programs with applications to optimal control and stochastic programming*, *SIAM Journal on Applied Mathematics*, **17**(4), 638–663 (1969)

## A Proof of Theorem 5.6

In the setting of Theorem 5.6, let  $(\mu^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)})$  be an optimal solution for  $(MP_1(w))$  with index set  $\mathcal{L}' = \{1, \dots, n\}$ . Recall that

$$F^{(k)} = \sum_{i \in \Omega} f_i(\mu_i^{(k)}, \lambda_i^{(k)}, w) - \beta(\mu^{(k)}, w),$$

$$\bar{F}^{(k)} = \sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k-1)}, \bar{\lambda}_i^{(k-1)}, w) - \beta(\bar{\mu}^{(k-1)}, w).$$

Let us define

$$\phi^{(n+1,k)} := \sum_{i \in \Omega} \varrho \left\| \mu_i^{(n+1)} - \bar{\mu}_i^{(k)} \right\|^2 + \sum_{i \in \Omega} \varrho \left\| \lambda_i^{(n+1)} - \bar{\lambda}_i^{(k)} \right\|^2,$$

$$\bar{\varepsilon}^{(n+1)} := \varepsilon + \phi^{(n+1,k)}.$$

Furthermore, let  $\bar{x} = (\bar{x}_i)_{i \in \Omega}$ ,  $\bar{y} = (\bar{y}_i)_{i \in \Omega}$  be defined as

$$\bar{x}_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} x_i^{(\ell)}, \quad \bar{y}_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} y_i^{(\ell)}, \quad (\text{A.1})$$

for every  $i \in \Omega$ .



**Lemma A.1.** *The following relationships hold for  $(\mu^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)}, \tau^{(n+1)}, \theta^{(n+1)}, \sigma^{(n+1)}, \Psi^{(n+1)}, \nu^{(n+1)})$ :*

$$\sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} = 1, \quad \forall i \in \Omega, \quad (\text{A.2})$$

$$\sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} = 1, \quad (\text{A.3})$$

$$\varrho(\lambda_i^{(n+1)} - \bar{\lambda}_i^{(k)}) = \frac{1}{2} \left( \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} g_{\lambda_i^{(\ell)}} - \sigma^{(n+1)} p_i \right), \quad \forall i \in \Omega \quad (\text{A.4})$$

$$-2\varrho(\mu_i^{(n+1)} - \bar{\mu}_i^{(k)}) + \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} g_{\mu_i^{(\ell)}} + \sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} \rho_{\mu_i^{(\ell)}} - \Psi^{(n+1)} + \nu_i^{(n+1)} = 0, \quad \forall i \in \Omega. \quad (\text{A.5})$$

**Proof.** The Lagrangian for  $(MP_1(w))$  with centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$  is

$$\begin{aligned} L(\mu, \lambda, \vartheta, \eta, \tau, \theta, \sigma, \Psi, \nu) & \quad (\text{A.6}) \\ &= \sum_{i \in \Omega} \vartheta_i + \eta - \sum_{i \in \Omega} \varrho \left\| \mu_i - \bar{\mu}_i^{(k)} \right\|^2 - \sum_{i \in \Omega} \varrho \left\| \lambda_i - \bar{\lambda}_i^{(k)} \right\|^2 \\ &+ \sum_{i \in \Omega} \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} \left( f_i(\mu_i^{(\ell)}, \lambda_i^{(\ell)}, w) + g_{\mu_i^{(\ell)}}^\top (\mu_i - \mu_i^{(\ell)}) + g_{\lambda_i^{(\ell)}}^\top (\lambda_i - \lambda_i^{(\ell)}) - \vartheta_i \right) \\ &+ \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \left( -\beta(\mu^{(\ell)}, w) + \sum_{i \in \Omega} \rho_{\mu_i^{(\ell)}}^\top (\mu_i - \mu_i^{(\ell)}) - \eta \right) \\ &- \sigma^\top \sum_{i \in \Omega} p_i \lambda_i + \Psi^\top \left( \mathbf{1} - \sum_{i \in \Omega} \mu_i \right) + \sum_{i \in \Omega} \nu_i^\top \mu_i. \end{aligned}$$

The dual objective function is defined by

$$h(\tau, \theta, \sigma, \Psi, \nu) = \sup_{\mu \in \mathbb{L}^J, \lambda \in \mathbb{L}^M, \vartheta \in \mathbb{L}, \eta \in \mathbb{R}} L(\mu, \lambda, \vartheta, \eta, \tau, \theta, \sigma, \Psi, \nu),$$

and the dual problem is

$$\begin{aligned} \min \quad & h(\tau, \theta, \sigma, \Psi, \nu) & (D - MP_1(w)) \\ \text{s.t.} \quad & \tau_i^{(\ell)} \geq 0, \theta^{(\ell)} \geq 0, \sigma \in \mathbb{R}^M, \Psi \in \mathbb{R}^J, \nu_i \in \mathbb{R}_+^J. \end{aligned}$$

Note that  $(\mu^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)})$  is an optimal solution for  $(MP_1(w))$  with centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ , and

$$\left( \tau^{(n+1)} = (\tau_i^{(\ell, n+1)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta^{(n+1)} = (\theta^{(\ell, n+1)})_{\ell \in \mathcal{L}'}, \sigma^{(n+1)}, \Psi^{(n+1)}, \nu^{(n+1)} = (\nu_i^{(n+1)})_{i \in \Omega} \right)$$

is the corresponding optimal solution for  $(D - MP_1(w))$ . The maximization of the Lagrangian over  $\vartheta \in \mathbb{L}$  and  $\eta \in \mathbb{R}$  gives (A.2) and (A.3), respectively, as constraints for the dual problem. The maximization of the Lagrangian over  $\lambda \in \mathbb{L}^M$  gives the first-order condition (A.4). Finally, the maximization of the Lagrangian over  $\mu \in \mathbb{L}^J$  gives the first-order condition (A.5).  $\square$

**Lemma A.2.** *The following statements hold for every  $i \in \Omega$ :*

(a) *As  $\varepsilon \rightarrow 0$ ,  $\bar{x}_i - \sigma^{(n+1)} \rightarrow 0$ .*

(b)  *$(\bar{x}_i, \bar{y}_i) \in \mathcal{F}_i$ .*

(c)  *$(\bar{x}, \bar{y})$  will be an element of the set  $\{(x, y) \in \mathcal{F} \mid p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega\}$  as  $\varepsilon \rightarrow 0$  in the sense that  $(\bar{x}, \bar{y}) \in \mathcal{F}$  and  $p_i(\bar{x}_i - \mathbb{E}[\bar{x}]) \rightarrow 0$  for every  $i \in \Omega$  as  $\varepsilon \rightarrow 0$ .*

**Proof.** (a) We have, for every  $i \in \Omega$ ,

$$\varrho \sum_{i \in \Omega} \left\| \lambda_i^{(n+1)} - \bar{\lambda}_i^{(k)} \right\|^2 \leq \phi^{(n+1,k)}. \quad (\text{A.7})$$

As  $\varepsilon \rightarrow 0$ , we have  $n \rightarrow \infty$  and, by Lemma 7.17 in Ruszczyński (2006),

$$\phi^{(n+1,k)} \rightarrow 0. \quad (\text{A.8})$$

Hence, by (A.4), (A.1), (A.7), (A.8), we get

$$\frac{1}{2} \left( \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} p_i x_i^{(\ell)} - \sigma^{(n+1)} p_i \right) = \frac{1}{2} p_i (\bar{x}_i - \sigma^{(n+1)}) \rightarrow 0$$

as  $\varepsilon \rightarrow 0$ . Then, the claim follows.

- (b) By line 6 of Algorithm 3, we have  $(x_i^{(\ell)}, y_i^{(\ell)}) \in \mathcal{F}_i$  for every  $i \in \Omega$  and  $\ell \in \mathcal{L}'$ . Since  $\mathcal{F}_i$  is convex,  $\tau_i^{(\ell, n+1)} \geq 0$  and  $\sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} = 1$ , we have  $(\bar{x}_i, \bar{y}_i) \in \mathcal{F}_i$  for every  $i \in \Omega$ .
- (c) Note that  $(\bar{x}, \bar{y}) \in \mathcal{F}$  for every  $\varepsilon > 0$ . Since  $\mathcal{F}$  is a closed set, the limit of this point as  $\varepsilon \rightarrow 0$  is also an element of  $\mathcal{F}$ . On the other hand, for every  $i \in \Omega$ ,  $\bar{x}_i - \sigma^{(n+1)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$  by part (a). Hence,  $\mathbb{E} [\sigma^{(n+1)} - \bar{x}] = \sum_{i \in \Omega} p_i (\sigma^{(n+1)} - \bar{x}_i) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . Therefore, for every  $i \in \Omega$ ,

$$p_i (\bar{x}_i - \mathbb{E} [\bar{x}]) = p_i (\bar{x}_i - \sigma^{(n+1)}) + p_i \mathbb{E} [\sigma^{(n+1)} - \bar{x}] \rightarrow 0$$

as  $\varepsilon \rightarrow 0$ . □

Recall that  $\rho_{\mu^{(\ell)}} = (\rho_{\mu_1^{(\ell)}}, \dots, \rho_{\mu_I^{(\ell)}})$  is a subgradient of  $-\beta(\cdot, w)$  at  $\mu^{(\ell)}$ , for each  $\ell \in \mathcal{L}'$ . From (2.1), there exists some  $u^{(\ell)} \in \mathcal{A}$ ,  $\ell \in \mathcal{L}'$ , such that

$$\rho_{\mu_i^{(\ell)}} = -w \cdot u_i^{(\ell)} \quad (\text{A.9})$$

for all  $i \in \Omega$ . Since  $\mathcal{A}$  is convex,  $\theta^{(\ell)} \geq 0$  and  $\sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} = 1$ , it follows that

$$\bar{u} := \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u^{(\ell)} = \left( \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u_1^{(\ell)}, \dots, \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u_I^{(\ell)} \right) \in \mathcal{A}. \quad (\text{A.10})$$

**Lemma A.3.** *The followings hold:*

(a) For each  $i \in \Omega$ ,

$$\vartheta_i^{(n+1)} = w^\top \left[ \mu_i^{(n+1)} \cdot (C\bar{x}_i + Q_i\bar{y}_i) \right] + p_i (\lambda_i^{(n+1)})^\top \bar{x}_i. \quad (\text{A.11})$$

(b) For each  $i \in \Omega$ ,

$$\vartheta_i^{(n+1)} - f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) \leq \bar{\varepsilon}^{(n+1)}. \quad (\text{A.12})$$

Moreover,

$$\sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) + \beta(\mu^{(n+1)}, w) \leq \bar{\varepsilon}^{(n+1)}. \quad (\text{A.13})$$

(c)

$$\eta^{(n+1)} + \beta(\mu^{(n+1)}, w) \leq \bar{\varepsilon}^{(n+1)}. \quad (\text{A.14})$$

(d)  $-w \cdot \bar{u}$  is an  $\bar{\varepsilon}^{(n+1)}$ -subgradient of  $-\beta(\cdot, w)$  at  $\mu^{(n+1)}$  in the sense that, for every  $\mu \in \mathbb{M}_1^J$ ,

$$-\beta(\mu, w) \leq -\beta(\mu^{(n+1)}, w) + \bar{\varepsilon}^{(n+1)} - \sum_{i \in \Omega} w^\top (\mu_i - \mu_i^{(n+1)}) \cdot \bar{u}_i. \quad (\text{A.15})$$

**Proof.** Consider  $(MP_1(w))$  with centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ .

(a) Note that constraint (5.11) can be rewritten as

$$\vartheta_i \leq w^\top \left[ \mu_i \cdot (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) \right] + p_i \lambda_i^\top x_i^{(\ell)}, \quad \forall i \in \Omega, \ell \in \mathcal{L}'$$

since

$$f_i(\mu_i^{(\ell)}, \lambda_i^{(\ell)}, w) = w^\top \left[ \mu_i^{(\ell)} \cdot (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) \right] + p_i (\lambda_i^{(\ell)})^\top x_i^{(\ell)}$$

and  $g_{\mu_i^{(\ell)}} = w \cdot (Cx_i^{(\ell)} + Q_i y_i^{(\ell)})$ ,  $g_{\lambda_i^{(\ell)}} = p_i x_i^{(\ell)}$ . From the complementary slackness conditions for constraint (5.11) and using (A.1), (A.2), we get

$$\begin{aligned} \vartheta_i^{(n+1)} &= \sum_{\ell \in \mathcal{L}'} \vartheta_i^{(n+1)} \tau_i^{(\ell, n+1)} \\ &= \sum_{\ell \in \mathcal{L}'} \left( w^\top \left[ \mu_i^{(n+1)} \cdot (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) \right] + p_i (\lambda_i^{(n+1)})^\top x_i^{(\ell)} \right) \tau_i^{(\ell, n+1)} \\ &= w^\top \left[ \mu_i^{(n+1)} \cdot (C\bar{x}_i + Q_i \bar{y}_i) \right] + p_i (\lambda_i^{(n+1)})^\top \bar{x}_i. \end{aligned}$$

Hence, (A.11) follows.

(b) Note that

$$\begin{aligned} &\vartheta_i^{(n+1)} - f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) + \beta(\mu^{(n+1)}, w) - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k)}, \bar{\lambda}_i^{(k)}, w) + \beta(\bar{\mu}^{(k)}, w) - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} - \sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k)}, \bar{\lambda}_i^{(k)}, w) + \beta(\bar{\mu}^{(k)}, w) - \phi^{(k+1, k)} \\ &\leq \varepsilon. \end{aligned}$$

Here, the first inequality follows since  $\vartheta_i^{(n+1)} \geq f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w)$  for each  $i \in \Omega$  and  $\eta^{(n+1)} \geq -\beta(\mu^{(n+1)}, w)$ , the second inequality follows since

$$\begin{aligned} F^{(n+1)} &= \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) - \beta(\mu^{(n+1)}, w) \\ &\geq \bar{F}^{(k+1)} = \sum_{i \in \Omega} f_i(\bar{\mu}_i^{(k)}, \bar{\lambda}_i^{(k)}, w) - \beta(\bar{\mu}^{(k)}, w) \end{aligned}$$

due to the center update rule in line 14 of Algorithm 3, the third inequality follows since the master problem with index set  $\mathcal{L} = \{1, \dots, k\}$  and center  $(\bar{\mu}^{(k)}, \bar{\lambda}^{(k)})$  has a smaller optimal value than the master problem with index set  $\mathcal{L}' = \{1, \dots, n\}$  and center  $(\bar{\mu}^{(k)}, \bar{\lambda}^{(k)})$ . Finally, the last inequality is by the approximate stopping condition (5.17). Hence, (A.12) and (A.13) follow.

(c) Similar to part (b), we can show that (A.14) holds.

(d) Note that constraint (5.12) can be rewritten as

$$\eta \leq \sum_{i \in \Omega} \rho_{\mu_i^{(\ell)}}^\top \mu_i, \quad \forall \ell \in \mathcal{L}'$$

since  $-\beta(\mu^{(\ell)}, w) = \sum_{i \in \Omega} \rho_{\mu_i^{(\ell)}}^\top \mu_i^{(\ell)}$  by the definition of subgradient. From the complementary slackness conditions,

$$\eta^{(n+1)} = \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \sum_{i \in \Omega} \rho_{\mu_i^{(\ell)}}^\top \mu_i^{(n+1)} = \sum_{i \in \Omega} \mu_i^{(n+1)} \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \rho_{\mu_i^{(\ell)}}^\top. \quad (\text{A.16})$$

From (A.9), (A.10), (A.16), it follows that

$$\eta^{(n+1)} = -w^\top \sum_{i \in \Omega} \mu_i^{(n+1)} \cdot \bar{u}_i = -w^\top \mathbb{E}^{\mu^{(n+1)}} [\bar{u}]. \quad (\text{A.17})$$

For every  $\mu \in \mathbb{M}_1^J$ ,

$$\begin{aligned} -\beta(\mu, w) &\leq -w^\top \mathbb{E}^\mu [\bar{u}] \\ &= -w^\top \mathbb{E}^{\mu^{(n+1)}} [\bar{u}] + w^\top \left( \mathbb{E}^{\mu^{(n+1)}} [\bar{u}] - \mathbb{E}^\mu [\bar{u}] \right) \\ &= -w^\top \mathbb{E}^{\mu^{(n+1)}} [\bar{u}] - \sum_{i \in \Omega} w^\top (\mu_i - \mu_i^{(n+1)}) \cdot \bar{u}_i \\ &\leq -\beta(\mu^{(n+1)}, w) + \bar{\varepsilon}^{(n+1)} - \sum_{i \in \Omega} w^\top (\mu_i - \mu_i^{(n+1)}) \cdot \bar{u}_i, \end{aligned}$$

where the first inequality follows from (2.1), and the last inequality follows from (A.14) and (A.17). Hence, the claim follows.  $\square$

**Lemma A.4.** *The followings hold:*

(a) As  $\varepsilon \rightarrow 0$ ,

$$\begin{aligned} w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] - \beta(\mu^{(n+1)}, w) &\rightarrow \mathcal{P}_1(w), \\ \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) - \beta(\mu^{(n+1)}, w) &\rightarrow \mathcal{P}_1(w). \end{aligned}$$

(b) Let  $z_{(w)}$  be a minimizer of the problem

$$\inf_{z \in R(C\bar{x} + Q\bar{y})} w^\top z.$$

Then,  $w^\top z_{(w)} \rightarrow \mathcal{P}_1(w)$  as  $\varepsilon \rightarrow 0$ .

**Proof.** (a) Note that

$$\begin{aligned} \sum_{i \in \Omega} \vartheta_i^{(n+1)} &= \sum_{i \in \Omega} \left( w^\top \left[ \mu_i^{(n+1)} \cdot (C\bar{x}_i + Q_i\bar{y}_i) \right] + p_i(\lambda_i^{(n+1)})^\top \bar{x}_i \right) \\ &= w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] + \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i, \end{aligned} \quad (\text{A.18})$$

where the first equality follows from (A.11). We may rewrite  $\sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i$  as

$$\begin{aligned} \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i &= \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top (\bar{x}_i - \sigma^{(n+1)}) + \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \sigma^{(n+1)} \\ &= \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top (\bar{x}_i - \sigma^{(n+1)}) \end{aligned} \quad (\text{A.19})$$

since  $\sigma^{(n+1)}$  is deterministic and  $\mathbb{E}[\lambda^{(n+1)}] = 0$ . Note that

$$\left| \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top (\bar{x}_i - \sigma^{(n+1)}) \right| \leq \sum_{i \in \Omega} p_i |\lambda_i^{(n+1)}|^\top |\bar{x}_i - \sigma^{(n+1)}|, \quad (\text{A.20})$$

where  $|z| := (|z^1|, \dots, |z^J|)$  for  $z \in \mathbb{R}^J$ . By the proof of Theorem 7.16 in Ruszczyński (2006), we have  $|(\lambda_i^{(n+1)})^j| \leq \tilde{C}$  for every  $i \in \Omega, j \in \mathcal{J}$ , where  $\tilde{C} > 0$  is some constant. Using (A.19) and (A.20), we get

$$\left| \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i \right| \leq \tilde{C} \sum_{i \in \Omega} p_i \mathbf{1}^\top |\bar{x}_i - \sigma^{(n+1)}|. \quad (\text{A.21})$$

Therefore, by Lemma A.2(a), (A.18) and (A.21),

$$\left| \sum_{i \in \Omega} \vartheta_i^{(n+1)} - w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] \right| = \left| \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i \right| \rightarrow 0 \quad (\text{A.22})$$

as  $\varepsilon \rightarrow 0$ . On the other hand, by (A.13) and using the fact that  $\vartheta_i^{(n+1)}$ ,  $\eta^{(n+1)}$  are upper approximations for  $f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w)$ ,  $-\beta(\mu^{(n+1)}, w)$ , respectively, we get

$$\begin{aligned} 0 &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} - \beta(\mu^{(n+1)}, w) - \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) + \beta(\mu^{(n+1)}, w) \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) + \beta(\mu^{(n+1)}, w) \leq \bar{\varepsilon}^{(n+1)}. \end{aligned} \quad (\text{A.23})$$

By (A.8),  $\bar{\varepsilon}^{(n+1)} = \varepsilon + \phi^{(n+1,k)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . By Lemma 7.17 in Ruszczyński (2006), as  $\varepsilon \rightarrow 0$ ,  $\sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)}$  converges to the optimal value of  $(D_1(w))$ , which is  $\mathcal{P}_1(w)$ . Hence,  $\sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) - \beta(\mu^{(n+1)}, w)$  also converges to  $\mathcal{P}_1(w)$  as  $\varepsilon \rightarrow 0$ . Finally, by triangle inequality, (A.22) and (A.23) yield

$$\left| w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] - \beta(\mu^{(n+1)}, w) - \sum_{i \in \Omega} f_i(\mu_i^{(n+1)}, \lambda_i^{(n+1)}, w) + \beta(\mu^{(n+1)}, w) \right| \leq \left| \sum_{i \in \Omega} p_i(\lambda_i^{(n+1)})^\top \bar{x}_i \right| + \bar{\varepsilon}^{(n+1)}.$$

From (A.22), the right hand side converges to zero as  $\varepsilon \rightarrow 0$ . We conclude that  $w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] - \beta(\mu^{(n+1)}, w)$  also converges to  $\mathcal{P}_1(w)$  as  $\varepsilon \rightarrow 0$ .

(b) By (2.3), we have

$$\begin{aligned} w^\top z(w) &= \sup_{\mu \in \mathbb{M}_1^J} (w^\top \mathbb{E}^\mu [C\bar{x} + Q\bar{y}] - \beta(\mu, w)) \\ &= \sup_{\mu \in \mathbb{L}^J} \left\{ w^\top \mathbb{E}^\mu [C\bar{x} + Q\bar{y}] - \beta(\mu, w) \mid \sum_{i \in \Omega} \mu_i = \mathbf{1}, \mu_i \in \mathbb{R}_+^J, \forall i \in \Omega \right\}. \end{aligned}$$

The corresponding Lagrangian dual problem is given by

$$\inf_{\Psi \in \mathbb{R}^J, \nu \in \mathbb{L}_+^J} \sup_{\mu \in \mathbb{L}^J} \left( w^\top \mathbb{E}^\mu [C\bar{x} + Q\bar{y}] - \beta(\mu, w) + \Psi^\top \left( \mathbf{1} - \sum_{i \in \Omega} \mu_i \right) + \sum_{i \in \Omega} \nu_i^\top \mu_i \right).$$

The optimization over  $\mu_1, \dots, \mu_I$  gives the first-order condition

$$w \cdot (C\bar{x}_i + Q_i\bar{y}_i) + \rho_{\mu_i} - \Psi + \nu_i = 0, \quad \forall i \in \Omega, \quad (\text{A.24})$$

for some  $\rho_\mu = (\rho_{\mu_1}, \dots, \rho_{\mu_I}) \in \partial_\mu(-\beta)(\mu, w)$ . We argue that  $(\mu^{(n+1)}, \Psi^{(n+1)}, \nu^{(n+1)})$  satisfies (A.24) in an approximate sense. Note that we may rewrite (A.5) as

$$\begin{aligned} &-2\varrho(\mu_i^{(n+1)} - \bar{\mu}_i^{(k)}) + \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} w \cdot (C\bar{x}_i^{(\ell)} + Q_i\bar{y}_i^{(\ell)}) - \sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} w \cdot u_i^{(\ell)} - \Psi^{(n+1)} + \nu_i^{(n+1)} \\ &= -2\varrho(\mu_i^{(n+1)} - \bar{\mu}_i^{(k)}) + w \cdot (C\bar{x}_i + Q_i\bar{y}_i) - w \cdot \bar{u}_i - \Psi^{(n+1)} + \nu_i^{(n+1)} \\ &= 0. \end{aligned} \quad (\text{A.25})$$

As shown in (A.15),  $-w \cdot \bar{u} = -w \cdot (\bar{u}_1, \dots, \bar{u}_I)$  is an  $\bar{\varepsilon}^{(n)}$ -subgradient of  $-\beta(\cdot, w)$  at  $\mu^{(n+1)}$ . Hence, there exists a subgradient  $\rho_{\mu^{(n+1)}}$  of  $-\beta(\cdot, w)$  at  $\mu^{(n+1)}$  for each  $n$  such that  $-w \cdot \bar{u} - \rho_{\mu^{(n+1)}} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . On the other hand, by (A.8),  $\mu_i^{(n+1)} - \bar{\mu}_i^{(k)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . Therefore, from (A.25), (A.24) is satisfied by  $(\mu^{(n+1)}, \Psi^{(n+1)}, \nu^{(n+1)})$  and  $-w \cdot \bar{u}$  approximately in the sense that

$$w \cdot (C\bar{x}_i + Q_i\bar{y}_i) - w \cdot \bar{u}_i - \Psi^{(n+1)} + \nu_i^{(n+1)} \rightarrow 0$$

and  $\left| w^\top \mathbb{E}^{\mu^{(n+1)}} [C\bar{x} + Q\bar{y}] - \beta(\mu^{(n+1)}, w) - w^\top z(w) \right| \rightarrow 0$  in the limit as  $\varepsilon \rightarrow 0$ . □

**Proof of Theorem 5.6.** Note that  $x_{(w)} = \bar{x}$  and  $y_{(w)} = \bar{y}$ , where  $\bar{x}, \bar{y}$  are defined by (A.1). Parts (a) and (c) follow directly from Lemma A.2. Part (b) follows directly from the definition of  $z_{(w)}$ . Part (d) follows from Lemma A.4.  $\square$

## B Proof of Theorem 5.14

In the setting of Theorem 5.14, let  $(m^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)})$  be an optimal solution for  $(MP_2(v))$  with index set  $\mathcal{L}' = \{1, \dots, n\}$ . Recall that

$$\begin{aligned} F^{(k)} &= \sum_{i \in \Omega} \tilde{f}_i(m_i^{(k)}, \lambda_i^{(k)}) - \tilde{\beta}(m^{(k)}) - \sum_{i \in \Omega} (m_i^{(k)})^\top v, \\ \bar{F}^{(k)} &= \sum_{i \in \Omega} \tilde{f}_i(\bar{m}_i^{(k-1)}, \bar{\lambda}_i^{(k-1)}) - \tilde{\beta}(\bar{m}^{(k-1)}) - \sum_{i \in \Omega} (\bar{m}_i^{(k-1)})^\top v. \end{aligned}$$

Let us re-define

$$\begin{aligned} \phi^{(n+1, k)} &:= \sum_{i \in \Omega} \varrho \left\| m_i^{(n+1)} - \bar{m}_i^{(k)} \right\|^2 + \sum_{i \in \Omega} \varrho \left\| \lambda_i^{(n+1)} - \bar{\lambda}_i^{(k)} \right\|^2, \\ \bar{\varepsilon}^{(n+1)} &:= \varepsilon + \phi^{(n+1, k)}. \end{aligned}$$

Furthermore, let  $\bar{x} = (\bar{x}_i)_{i \in \Omega}, \bar{y} = (\bar{y}_i)_{i \in \Omega}$  be defined as

$$\bar{x}_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} x_i^{(\ell)}, \quad \bar{y}_i := \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} y_i^{(\ell)}, \quad (\text{B.1})$$

for every  $i \in \Omega$ .

**Lemma B.1.** *The following relationships hold for  $(m^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)}, \tau^{(n+1)}, \theta^{(n+1)}, \sigma^{(n+1)}, \psi^{(n+1)}, \nu^{(n+1)})$ :*

$$\sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} = 1, \quad \forall i \in \Omega, \quad (\text{B.2})$$

$$\sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} = 1, \quad (\text{B.3})$$

$$\varrho(\lambda_i^{(n+1)} - \bar{\lambda}_i^{(k)}) = \frac{1}{2} \left( \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} g_{\lambda_i^{(\ell)}} - \sigma^{(n+1)} p_i \right), \quad \forall i \in \Omega \quad (\text{B.4})$$

$$-2\varrho(m_i^{(n+1)} - \bar{m}_i^{(k)}) + \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} g_{m_i^{(\ell)}} + \sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} \rho_{m_i^{(\ell)}} - \psi^{(n+1)} \mathbf{1} + \nu_i^{(n+1)} = 0, \quad \forall i \in \Omega. \quad (\text{B.5})$$

**Proof.** The Lagrangian for  $(MP_2(v))$  with centers  $\bar{\mu}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$  is

$$\begin{aligned} &L(m, \lambda, \vartheta, \eta, \tau, \theta, \sigma, \psi, \nu) \quad (\text{B.6}) \\ &= \sum_{i \in \Omega} \vartheta_i + \eta - \sum_{i \in \Omega} \varrho \left\| m_i - \bar{m}_i^{(k)} \right\|^2 - \sum_{i \in \Omega} \varrho \left\| \lambda_i - \bar{\lambda}_i^{(k)} \right\|^2 \\ &\quad + \sum_{i \in \Omega} \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell)} \left( \tilde{f}_i(m_i^{(\ell)}, \lambda_i^{(\ell)}) + g_{m_i^{(\ell)}}^\top (m_i - m_i^{(\ell)}) + g_{\lambda_i^{(\ell)}}^\top (\lambda_i - \lambda_i^{(\ell)}) - \vartheta_i \right) \\ &\quad + \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \left( -\tilde{\beta}(m^{(\ell)}) + \sum_{i \in \Omega} \rho_{m_i^{(\ell)}}^\top (m_i - m_i^{(\ell)}) - \eta \right) \\ &\quad - \sigma^\top \sum_{i \in \Omega} p_i \lambda_i + \psi \left( 1 - \sum_{i \in \Omega} m_i^\top \mathbf{1} \right) + \sum_{i \in \Omega} \nu_i^\top m_i. \end{aligned}$$

The dual objective function is defined by

$$h(\tau, \theta, \sigma, \psi, \nu) = \sup_{m \in \mathbb{L}^J, \lambda \in \mathbb{L}^M, \vartheta \in \mathbb{L}, \eta \in \mathbb{R}} L(m, \lambda, \vartheta, \eta, \tau, \theta, \sigma, \psi, \nu),$$

and the dual problem is

$$\begin{aligned} \min \quad & h(\tau, \theta, \sigma, \psi, \nu) && (D - MP_2(v)) \\ \text{s.t.} \quad & \tau_i^{(\ell)} \geq 0, \theta^{(\ell)} \geq 0, \sigma \in \mathbb{R}^M, \psi \in \mathbb{R}, \nu_i \in \mathbb{R}_+^J. \end{aligned}$$

Note that  $(m^{(n+1)}, \lambda^{(n+1)}, \vartheta^{(n+1)}, \eta^{(n+1)})$  is an optimal solution for  $(MP_2(v))$  with centers  $\bar{m}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ , and

$$\left( \tau^{(n+1)} = (\tau_i^{(\ell, n+1)})_{i \in \Omega, \ell \in \mathcal{L}'}, \theta^{(n+1)} = (\theta^{(\ell, n+1)})_{\ell \in \mathcal{L}'}, \sigma^{(n+1)}, \psi^{(n+1)}, \nu^{(n+1)} = (\nu_i^{(n+1)})_{i \in \Omega} \right)$$

is the corresponding optimal solution for  $(D - MP_2(v))$ . The maximization of the Lagrangian over  $\vartheta \in \mathbb{L}$  and  $\eta \in \mathbb{R}$  gives (B.2) and (B.3), respectively, as constraints for the dual problem. The maximization of the Lagrangian over  $\lambda \in \mathbb{L}^M$  gives the first-order condition (B.4). Finally, the maximization of the Lagrangian over  $m \in \mathbb{L}^J$  gives the first-order condition (B.5).  $\square$

**Lemma B.2.** *The following statements hold for every  $i \in \Omega$ :*

- (a) *As  $\varepsilon \rightarrow 0$ ,  $\bar{x}_i - \sigma^{(n+1)} \rightarrow 0$ .*
- (b)  *$(\bar{x}_i, \bar{y}_i) \in \mathcal{F}_i$ .*
- (c)  *$(\bar{x}, \bar{y})$  will eventually be an element of the set  $\{(x, y) \in \mathcal{F} \mid p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega\}$  as  $\varepsilon \rightarrow 0$  in the sense that  $(\bar{x}, \bar{y}) \in \mathcal{F}$  and  $p_i(\bar{x}_i - \mathbb{E}[\bar{x}]) \rightarrow 0$  for every  $i \in \Omega$  as  $\varepsilon \rightarrow 0$ .*

**Proof.** The proof of this lemma is similar to the proof of Lemma A.2. Therefore, it is omitted.  $\square$

Recall that  $\rho_{m^{(\ell)}} = (\rho_{m_1^{(\ell)}}, \dots, \rho_{m_I^{(\ell)}})$  is a subgradient of  $-\tilde{\beta}(\cdot, w)$  at  $m^{(\ell)}$ , for each  $\ell \in \mathcal{L}'$ . From (5.19), there exists some  $u^{(\ell)} \in \mathcal{A}$ ,  $\ell \in \mathcal{L}'$ , such that

$$\rho_{m^{(\ell)}} = -u_i^{(\ell)} \tag{B.7}$$

for all  $i \in \Omega$ . Since  $\mathcal{A}$  is convex,  $\theta^{(\ell)} \geq 0$  and  $\sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} = 1$ , it follows that

$$\bar{u} := \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u^{(\ell)} = \left( \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u_1^{(\ell)}, \dots, \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} u_I^{(\ell)} \right) \in \mathcal{A}. \tag{B.8}$$

**Lemma B.3.** *The followings hold:*

- (a) *For each  $i \in \Omega$ ,*

$$\vartheta_i^{(n+1)} = (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) + p_i(\lambda_i^{(n+1)})^\top \bar{x}_i. \tag{B.9}$$

- (b) *For each  $i \in \Omega$ ,*

$$\vartheta_i^{(n+1)} - \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) \leq \bar{\varepsilon}^{(n+1)}. \tag{B.10}$$

Moreover,

$$\sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) + \tilde{\beta}(m^{(n+1)}) \leq \bar{\varepsilon}^{(n+1)}. \tag{B.11}$$

- (c)

$$\eta^{(n+1)} + \tilde{\beta}(m^{(n+1)}) \leq \bar{\varepsilon}^{(n+1)}. \tag{B.12}$$

(d)  $-\bar{u}$  is an  $\bar{\varepsilon}^{(n+1)}$ -subgradient of  $-\tilde{\beta}(\cdot)$  at  $m^{(n+1)}$  in the sense that, for every  $m \in \mathbb{M}_f^J$ ,

$$-\tilde{\beta}(m) \leq -\tilde{\beta}(m^{(n+1)}) + \bar{\varepsilon}^{(n+1)} - \sum_{i \in \Omega} (m_i - m_i^{(n+1)})^\top \bar{u}_i. \quad (\text{B.13})$$

**Proof.** Consider  $(MP_2(v))$  with centers  $\bar{m}^{(k)}, \bar{\lambda}^{(k)}$  and index set  $\mathcal{L}'$ .

(a) Note that constraint (5.23) can be rewritten as

$$\vartheta_i \leq m_i^\top (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) + p_i \lambda_i^\top x_i^{(\ell)}, \quad \forall i \in \Omega, \ell \in \mathcal{L}'$$

since

$$\tilde{f}_i(m_i^{(\ell)}, \lambda_i^{(\ell)}) = (m_i^{(\ell)})^\top (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) + p_i (\lambda_i^{(\ell)})^\top x_i^{(\ell)}$$

and  $g_{m_i^{(\ell)}} = Cx_i^{(\ell)} + Q_i y_i^{(\ell)}$ ,  $g_{\lambda_i^{(\ell)}} = p_i x_i^{(\ell)}$ . From the complementary slackness conditions for constraint (5.23) and using (B.1), (B.2), we get

$$\begin{aligned} \vartheta_i^{(n)} &= \sum_{\ell \in \mathcal{L}'} \vartheta_i^{(n+1)} \tau_i^{(\ell, n+1)} \\ &= \sum_{\ell \in \mathcal{L}'} \left( (m_i^{(n+1)})^\top (Cx_i^{(\ell)} + Q_i y_i^{(\ell)}) + p_i (\lambda_i^{(n+1)})^\top x_i^{(\ell)} \right) \tau_i^{(\ell, n+1)} \\ &= (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i \bar{y}_i) + p_i (\lambda_i^{(n+1)})^\top \bar{x}_i. \end{aligned}$$

Hence, (B.9) follows.

(b) Note that

$$\begin{aligned} &\vartheta_i^{(n+1)} - \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v - \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) + \tilde{\beta}(m^{(n+1)}) + \sum_{i \in \Omega} (m_i^{(n+1)})^\top v - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v - \sum_{i \in \Omega} \tilde{f}_i(\bar{m}_i^{(k)}, \bar{\lambda}_i^{(k)}) + \tilde{\beta}(\bar{m}^{(k)}) + \sum_{i \in \Omega} (\bar{m}_i^{(k)})^\top v - \phi^{(n+1, k)} \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(k+1)} + \eta^{(k+1)} - \sum_{i \in \Omega} (m_i^{(k+1)})^\top v - \sum_{i \in \Omega} \tilde{f}_i(\bar{m}_i^{(k)}, \bar{\lambda}_i^{(k)}) + \tilde{\beta}(\bar{m}^{(k)}) + \sum_{i \in \Omega} (\bar{m}_i^{(k)})^\top v - \phi^{(k+1, k)} \\ &\leq \varepsilon. \end{aligned}$$

Here, the first inequality follows since  $\vartheta_i^{(n+1)} \geq \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)})$  for each  $i \in \Omega$  and  $\eta^{(n+1)} \geq -\tilde{\beta}(m^{(n+1)})$ , the second inequality follows since

$$\begin{aligned} F^{(n+1)} &= \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v \\ &\geq \bar{F}^{(k+1)} = \sum_{i \in \Omega} \tilde{f}_i(\bar{m}_i^{(k)}, \bar{\lambda}_i^{(k)}) - \tilde{\beta}(\bar{m}^{(k)}) - \sum_{i \in \Omega} (\bar{m}_i^{(k)})^\top v \end{aligned}$$

due to the center update rule in line 14 of Algorithm 4, the third inequality follows since the master problem with index set  $\mathcal{L} = \{1, \dots, k\}$  and center  $(\bar{m}^{(k)}, \bar{\lambda}^{(k)})$  has a smaller optimal value than the master problem with index set  $\mathcal{L}' = \{1, \dots, n\}$  and center  $(\bar{m}^{(k)}, \bar{\lambda}^{(k)})$ . Finally, the last inequality is by the approximate stopping condition (5.29). Hence, (B.10) and (B.11) follow.

(c) Similar to part (b), we can show that (B.12) holds.

(d) Note that constraint (5.24) can be rewritten as

$$\eta \leq \sum_{i \in \Omega} \rho_{m_i^{(\ell)}}^\top m_i, \quad \forall \ell \in \mathcal{L}'$$



since  $-\tilde{\beta}(m^{(\ell)}) = \sum_{i \in \Omega} \rho_{m_i^{(\ell)}}^\top m_i^{(\ell)}$  by the definition of subgradient. From the complementary slackness conditions,

$$\eta^{(n+1)} = \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \sum_{i \in \Omega} \rho_{m_i^{(\ell)}}^\top m_i^{(n+1)} = \sum_{i \in \Omega} m_i^{(n+1)} \sum_{\ell \in \mathcal{L}'} \theta^{(\ell)} \rho_{m_i^{(\ell)}}^\top. \quad (\text{B.14})$$

From (B.7), (B.8), (B.14), it follows that

$$\eta^{(n+1)} = - \sum_{i \in \Omega} (m_i^{(n+1)})^\top \bar{u}_i. \quad (\text{B.15})$$

For every  $m \in \mathbb{M}_f^J$ ,

$$\begin{aligned} -\tilde{\beta}(m) &\leq - \sum_{i \in \Omega} m_i^\top \bar{u}_i \\ &= - \sum_{i \in \Omega} (m_i^{(n+1)})^\top \bar{u}_i - \sum_{i \in \Omega} (m_i - m_i^{(n+1)})^\top \bar{u}_i \\ &\leq -\tilde{\beta}(m^{(n+1)}) + \bar{\varepsilon}^{(n+1)} - \sum_{i \in \Omega} (m_i - m_i^{(n+1)})^\top \bar{u}_i, \end{aligned}$$

where the first inequality follows from (5.19), and the last inequality follows from (B.12) and (B.15). Hence, the claim follows.  $\square$

**Lemma B.4.** *The followings hold:*

(a) As  $\varepsilon \rightarrow 0$ ,

$$\begin{aligned} \sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v &\rightarrow \mathcal{P}_2(v), \\ \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v &\rightarrow \mathcal{P}_2(v). \end{aligned}$$

(b) Let

$$\bar{\alpha} := \inf \{ \alpha \in \mathbb{R} \mid v + \alpha \mathbf{1} \in R(C\bar{x} + Q\bar{y}) \}. \quad (\text{B.16})$$

Then,  $\bar{\alpha} \rightarrow \mathcal{P}_2(v)$  as  $\varepsilon \rightarrow 0$ .

**Proof.** (a) By (B.9),

$$\sum_{i \in \Omega} \vartheta_i^{(n+1)} = \sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) + \sum_{i \in \Omega} p_i (\lambda_i^{(n+1)})^\top \bar{x}_i. \quad (\text{B.17})$$

Using Lemma B.2(a), (B.17) and (A.21), we obtain

$$\left| \sum_{i \in \Omega} \vartheta_i^{(n+1)} - \sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) \right| = \left| \sum_{i \in \Omega} p_i (\lambda_i^{(n+1)})^\top \bar{x}_i \right| \rightarrow 0 \quad (\text{B.18})$$

as  $\varepsilon \rightarrow 0$ . On the other hand, by (B.11) and using the fact that  $\vartheta_i^{(n+1)}, \eta^{(n+1)}$  are upper approximations for  $\tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}), -\tilde{\beta}(m^{(n+1)})$ , respectively, we get

$$\begin{aligned} 0 &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v - \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) + \tilde{\beta}(m^{(n+1)}) + \sum_{i \in \Omega} (m_i^{(n+1)})^\top v \\ &\leq \sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) + \tilde{\beta}(m^{(n+1)}) \leq \bar{\varepsilon}^{(n+1)}. \end{aligned} \quad (\text{B.19})$$

By an analog of (A.8),  $\bar{\varepsilon}^{(n+1)} = \varepsilon + \phi^{(n+1,k)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . By Lemma 7.17 in Ruszczyński (2006), as  $\varepsilon \rightarrow 0$ ,  $\sum_{i \in \Omega} \vartheta_i^{(n+1)} + \eta^{(n+1)} - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v$  converges to the optimal value of  $(D_2(v))$ , which is  $\mathcal{P}_2(v)$ . Hence,  $\sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v$  also converges to  $\mathcal{P}_2(v)$  as  $\varepsilon \rightarrow 0$ . Finally, by triangle inequality, (B.18) and (B.19) yield

$$\left| \sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v \right. \\ \left. - \left( \sum_{i \in \Omega} \tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v \right) \right| \leq \left| \sum_{i \in \Omega} p_i (\lambda_i^{(n+1)})^\top \bar{x}_i \right| + \bar{\varepsilon}^{(n+1)}.$$

From (B.18), the right hand side of the above inequality converges to zero as  $\varepsilon \rightarrow 0$ . We conclude that  $\sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) - \tilde{\beta}(m^{(n+1)}) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v$  also converges to  $\mathcal{P}_2(v)$  as  $\varepsilon \rightarrow 0$ .

(b) By Lemma 5.7 and Lemma 5.12, we have

$$\begin{aligned} \bar{\alpha} &= \sup \left\{ \gamma^\top (\mathbb{E}^\mu [C\bar{x} + Q\bar{y}] - v) - \beta(\mu, \gamma) \mid \mu \in \mathbb{M}_1^J, \gamma^\top \mathbf{1} = 1, \gamma \in \mathbb{R}_+^J \right\} \\ &= \sup \left\{ \sum_{i \in \Omega} m_i^\top (C\bar{x}_i + Q_i\bar{y}_i) - \sum_{i \in \Omega} m_i^\top v - \tilde{\beta}(m) \mid m \in \mathbb{M}_f^J, \sum_{i \in \Omega} m_i^\top \mathbf{1} = 1 \right\} \\ &= \sup_{m \in \mathbb{L}^J} \left\{ \sum_{i \in \Omega} m_i^\top (C\bar{x}_i + Q_i\bar{y}_i) - \sum_{i \in \Omega} m_i^\top v - \tilde{\beta}(m) \mid \sum_{i \in \Omega} m_i^\top \mathbf{1} = 1, m_i \in \mathbb{R}_+^J, \forall i \in \Omega \right\}. \end{aligned}$$

The corresponding Lagrangian dual problem is given by

$$\inf_{\psi \in \mathbb{R}^J, \nu \in \mathbb{L}_+^J} \sup_{m \in \mathbb{L}^J} \left( \sum_{i \in \Omega} m_i^\top (C\bar{x}_i + Q_i\bar{y}_i) - \sum_{i \in \Omega} m_i^\top v - \tilde{\beta}(m) + \psi^\top \left( \mathbf{1} - \sum_{i \in \Omega} m_i \right) + \sum_{i \in \Omega} \nu_i^\top m_i \right).$$

The optimization over  $m_1, \dots, m_I$  gives the first-order condition

$$C\bar{x}_i + Q_i\bar{y}_i - v + \rho_{m_i} - \psi + \nu_i = 0, \quad \forall i \in \Omega, \quad (\text{B.20})$$

for some  $\rho_m = (\rho_{m_1}, \dots, \rho_{m_I}) \in \partial_m(-\tilde{\beta})(m)$ . We argue that  $(m^{(n+1)}, \psi^{(n+1)}, \nu^{(n+1)})$  satisfies (B.20) in an approximate sense. Note that we may rewrite (B.5) as

$$\begin{aligned} &-2\varrho(m_i^{(n+1)} - \bar{m}_i^{(k)}) + \sum_{\ell \in \mathcal{L}'} \tau_i^{(\ell, n+1)} \left( Cx_i^{(\ell)} + Q_i y_i^{(\ell)} \right) + \sum_{\ell \in \mathcal{L}'} \theta^{(\ell, n+1)} u_i^{(\ell)} - \psi^{(n+1)} \mathbf{1} + \nu_i^{(n+1)} \\ &= -2\varrho(m_i^{(n+1)} - \bar{m}_i^{(k)}) + (C\bar{x}_i + Q_i\bar{y}_i) + \bar{u}_i - \psi^{(n+1)} \mathbf{1} + \nu_i^{(n+1)} \\ &= 0. \end{aligned} \quad (\text{B.21})$$

As shown in (B.13),  $-\bar{u} = -(\bar{u}_1, \dots, \bar{u}_I)$  is an  $\bar{\varepsilon}^{(n+1)}$ -subgradient of  $-\tilde{\beta}(\cdot)$  at  $m^{(n+1)}$ . Hence, there exists a subgradient  $\rho_{m^{(n+1)}}$  of  $-\tilde{\beta}(\cdot)$  at  $m^{(n+1)}$  for each  $n$  such that  $-\bar{u} - \rho_{m^{(n+1)}} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . On the other hand, by an analog of (A.8),  $m_i^{(n+1)} - \bar{m}_i^{(k)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . Therefore, from (B.21), (B.20) is satisfied by  $(m^{(n+1)}, \psi^{(n+1)}, \nu^{(n+1)})$  and  $-\bar{u}$  approximately in the sense that

$$C\bar{x}_i + Q_i\bar{y}_i - \bar{u}_i - \psi^{(n+1)} + \nu_i^{(n+1)} \rightarrow 0$$

and

$$\left| \sum_{i \in \Omega} (m_i^{(n+1)})^\top (C\bar{x}_i + Q_i\bar{y}_i) - \sum_{i \in \Omega} (m_i^{(n+1)})^\top v - \tilde{\beta}(m^{(n+1)}) - \bar{\alpha} \right| \rightarrow 0 \quad (\text{B.22})$$

in the limit as  $\varepsilon \rightarrow 0$ . □

**Proof of Theorem 5.14.** Note that  $x_{(v)} = \bar{x}$ ,  $y_{(v)} = \bar{y}$ ,  $\alpha_{(v)} = \bar{\alpha}$ , where  $\bar{x}, \bar{y}, \bar{\alpha}$  are defined by (B.1) and (B.16). Parts (a) and (c) follow directly from Lemma B.2. Parts (b) and (d) follow from Lemma B.4. □

## C Proof of Theorem 5.15

**Proof of Theorem 5.15.** For the objective function of  $(LD_2(v))$  evaluated at  $\gamma = \gamma_{(v)}$ , we have

$$\begin{aligned}
& \inf_{(x,y) \in \mathcal{X}, \alpha \in \mathbb{R}} \left( \alpha + \inf_{z \in R(Cx+Qy)-v-\alpha \mathbf{1}} \gamma_{(v)}^\top z \right) \\
&= \inf_{\substack{(x,y) \in \mathcal{F}, \alpha \in \mathbb{R}, \\ p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega}} \left( \alpha + \inf_{z \in R(Cx+Qy)-v-\alpha \mathbf{1}} \gamma_{(v)}^\top z \right) \\
&= \inf_{\substack{(x,y) \in \mathcal{F}, \\ p_i(x_i - \mathbb{E}[x]) = 0, \forall i \in \Omega}} \sup_{\substack{m \in \mathbb{M}_f^J, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} m_i^\top (Cx_i + Q_i y_i) - \tilde{\beta}(m) - \gamma_{(v)}^\top v \right) \\
&= \sup_{\lambda \in \mathbb{L}^M} \inf_{(x,y) \in \mathcal{F}} \left( \sup_{\substack{m \in \mathbb{M}_f^J, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} m_i^\top (Cx_i + Q_i y_i) - \tilde{\beta}(m) - \sum_{i \in \Omega} m_i^\top v \right) + \sum_{i \in \Omega} p_i \lambda_i^\top (x_i - \mathbb{E}[x]) \right) \\
&= \sup_{\substack{\lambda \in \mathbb{L}^M, \\ \mathbb{E}[\lambda] = 0}} \inf_{(x,y) \in \mathcal{F}} \left( \sup_{\substack{m \in \mathbb{M}_f^J, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} m_i^\top (Cx_i + Q_i y_i) - \tilde{\beta}(m) - \sum_{i \in \Omega} m_i^\top v \right) + \sum_{i \in \Omega} p_i \lambda_i^\top x_i \right) \\
&= \sup_{\substack{\lambda \in \mathbb{L}^M, m \in \mathbb{M}_f^J, \\ \mathbb{E}[\lambda] = 0, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} \left( \inf_{(x_i, y_i) \in \mathcal{F}_i} (m_i^\top (Cx_i + Q_i y_i) + p_i \lambda_i^\top x_i) - m_i^\top v \right) - \tilde{\beta}(m) \right) \\
&= \sup_{\substack{\lambda \in \mathbb{L}^M, m \in \mathbb{M}_f^J, \\ \mathbb{E}[\lambda] = 0, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} (\tilde{f}_i(m_i, \lambda_i) - m_i^\top v) - \tilde{\beta}(m) \right) \tag{C.1}
\end{aligned}$$

$$\begin{aligned}
&\leq \sup_{\substack{\lambda \in \mathbb{L}^M, m \in \mathbb{M}_f^J, \\ \mathbb{E}[\lambda] = 0, \\ \sum_{i \in \Omega} m_i^\top \mathbf{1} = 1}} \left( \sum_{i \in \Omega} (\tilde{f}_i(m_i, \lambda_i) - m_i^\top v) - \tilde{\beta}(m) \right) \tag{C.2} \\
&= \mathcal{P}_2(v),
\end{aligned}$$

where the first equality is obvious, the second equality follows by (2.3), (5.26) and Lemma 5.7, the third and fourth equalities follow by the dualization of the nonanticipativity constraints as in the proof of Theorem 5.1, the fifth equality is by the interchange of infimum and supremum using Sion (1958), the sixth equality follows by the definition of  $\tilde{f}_i(\cdot, \cdot)$  in (5.20). The inequality in (C.2) follows since the feasible region of the problem in (C.1) is a subset of the feasible region of the problem in (C.2). The last equality is by Theorem 5.11.

Since  $(\lambda^{(n+1)}, m^{(n+1)})$  is a feasible solution for the maximization problem in (C.1), we have

$$\sum_{i \in \Omega} (\tilde{f}_i(m_i^{(n+1)}, \lambda_i^{(n+1)}) - (m_i^{(n+1)})^\top v) - \tilde{\beta}(m^{(n+1)}) \leq \sup_{\substack{\lambda \in \mathbb{L}^M, m \in \mathbb{M}_f^J, \\ \mathbb{E}[\lambda] = 0, \\ \sum_{i \in \Omega} m_i = \gamma_{(v)}}} \left( \sum_{i \in \Omega} (\tilde{f}_i(m_i, \lambda_i) - m_i^\top v) - \tilde{\beta}(m) \right) \leq \mathcal{P}_2(v).$$

As  $\varepsilon \rightarrow \infty$ , by Lemma B.4(a), the first expression converges to  $\mathcal{P}_2(v)$ , and so does the second by sandwich theorem. This finishes the proof of (5.31).  $\square$