RUTCOR
RESEARCH
REPORT

Scenario decomposition of risk-averse two stage stochastic programming problems

Ricardo A. Collado[a] Dávid Papp[b]
Andrzej Ruszczyński[c]

RRR 2-2012, January 2012

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]

[b]

[c]RUTCOR Rutgers, the State University of New Jersey, 640 Bartholomew Road, Piscataway, NJ 08854-8003

Rutcor Research Report
RRR 2-2012, January 2012

Scenario decomposition of risk-averse two stage stochastic programming problems

Ricardo A. Collado   Dávid Papp   Andrzej Ruszczyński

Abstract. We develop methods that exploit the dual representation of coherent risk measures to produce efficient algorithms that can solve nonlinear risk-averse stochastic problems by solving a series of linear sub problems. Our main theoretical tool is the development of duality theory for risk-averse two stage stochastic problems. The basic model that we consider is:

$$\min_{x \in X} \; \rho_1 \left(c^\top x + \rho_2 \left[\mathcal{Q}(x, \xi) \right] \right), \tag{1}$$

where ρ_1, ρ_2 are coherent risk measures and $\mathcal{Q}(x, \xi)$ is the optimal value of a second stage linear problem with a random vector $\xi \in \Omega$. We show methods that solve (1) when the underlying probability space Ω is finite.

We also develop a new type of bundle method, called the *truncated bundle method*, which exploits the topological properties of the domain of the functions to obtain better running time than the classical bundle method. This algorithm solves general two stage stochastic programs and has value and applicability on its own. As a testbed for our methods we consider a problem in manufacturing and transportation and implement in AMPL all of our methods for this problem. The numerical results from problems with hundreds of first stage scenarios as well as hundreds of second stage scenarios are compared across all the methods developed.

1 Introduction

The traditional approach of optimizing the expectation operator in stochastic programs successfully introduce uncertainty of events in the models but might fail to convey the element of risk that certain modeling problems face. During the last decade researchers have developed the coherent risk measures as an alternative to the expectation operator in the traditional stochastic programs. These operators are consistent with a systematized theory of risk as presented in [2, 3] and by substituting the expectation operator give rise to *risk-averse programs*. The coherent risk measures have a rich axiomatic theory including duality and differentiability, thus allowing the development of efficient methods for the solution of risk-averse programs (see for example [1, 4, 5]). In [10, 11, 9, 12] we can find a comprehensive treatment of the coherent risk measures and risk-averse optimization including the development of multi stage risk-averse programs. See [8] for a general development of multi staged stochastic problems.

In this article we will define a two stage risk-averse program utilizing coherent measures of risk and develop its dual and differentiability properties. Then we will apply cutting plane and bundle algorithms for its solution as well as a special version of the bundle method that exploits the geometrical structure of the feasible region of the dual problem. Finally we will consider a concrete application of our techniques and will show a comparison of the performance of the different methods on it.

2 Basics

Let us first recall some basic concepts of the theory of coherent risk measures. We will follow closely the development given in [10, 11, 9, 12].

Let $(\Omega, \mathcal{F}, P)$ be a probability space with sigma algebra $\mathcal{F}$ and probability measure P. Also, let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, where $p \in [1, +\infty)$. Each element $Z := Z(\omega)$ of $\mathcal{Z}$ is viewed as an uncertain outcome on $(\Omega, \mathcal{F})$ and it is by definition a random variable with finite p-th order moment. For $Z, Z' \in \mathcal{Z}$ we denote by $Z \preceq Z'$ the pointwise partial order meaning $Z(\omega) \geq Z'(\omega)$ for a.e. $\omega \in \Omega$. We also assume that the smaller the realizations of Z, the better; for example Z may represent a random cost.

Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. A *coherent risk measure* is a proper function $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ satisfying the following axioms:

(A1) *Convexity*: $\rho\left(tZ + (1-t)\,Z'\right) \leq t\rho(Z) + (1-t)\rho(Z')$, for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0, 1]$;

(A2) *Monotonicity*: If $Z, Z' \in \mathcal{Z}$ and $Z \preceq Z'$, then $\rho(Z) \leq \rho(Z')$;

(A3) *Translation Equivalence*: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$;

(A4) *Positive Homogeneity*: If $t > 0$ and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.

A coherent risk measure of particular interest is the *mean upper semideviation of first order* defined by

$$\rho(Z) = \mathbb{E}[Z] + a\,\mathbb{E}\left[Z - \mathbb{E}[Z]\right]_+ ,$$

for every $Z \in \mathcal{Z}$ and a fixed $a \in [0,1]$. See [12] page 277 for the details showing that the mean upper semideviation is a coherent risk measure and some of its applications.

3 A Two Stage Problem

Most of this article will be devoted to the study of a *two stage risk-averse problem* of the form

$$\min_{x \in X} \; \rho_1 \left(c^\top x + \rho_2 \left[\mathcal{Q}(x,\xi) \right] \right), \tag{2}$$

where ρ_1, ρ_2 are coherent risk measures, $X \subseteq \mathbb{R}^n$ is compact and polyhedral, and $\mathcal{Q}(x,\xi)$ is the optimal value of the *second stage problem*

$$\begin{aligned} &\min_{y \in \mathbb{R}^m} \; q^\top y \\ &\text{s.t. } Tx + Wy = h, \; y \geq 0. \end{aligned} \tag{3}$$

Here $\xi := (q,h,T,W)$ is the data of the second stage problem. We view some or all elements of the vector ξ as random and the ρ_1 operator at the first stage problem (2) is taken with respect to the probability distribution of $c^\top x + \rho_2\left[\mathcal{Q}(x,\xi)\right]$.

If for some x and ξ the second stage problem (3) is infeasible, then by definition $\mathcal{Q}(x,\xi) = +\infty$. It could also happen that the second stage problem is unbounded from below and hence $\mathcal{Q}(x,\xi) = -\infty$. This is somewhat pathological situation, meaning that for some value of the first stage decision vector and a realization of the random data, the value of the second stage problem can be improved indefinitely. Models exhibiting such properties should be avoided.

In order to simplify our exposition, we will assume that the distribution of ξ has finite support. That is, ξ has a finite number of realizations (called *scenarios*) $\xi_k = (q_k, h_k, T_k, W_k)$ with respective probabilities $p_k, k = 1, \dots, N$. In this case we will let $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$ which we will just identify with the space $\mathbb{R}^N$.

The following basic duality result for convex risk measures is a direct consequence of the FenchelMoreau theorem (see [12] Theorem 6.4).

Theorem 1. *Suppose that Ω is a finite probability space as described above and $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, where $[p \in 1, +\infty)$. If $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is a proper, lower semicontinuous, and coherent risk measure then for every random variable $Z = (z_1, \dots, z_N)$:*

$$\rho(Z) = \max_{\mu \in \partial\rho(0)} \sum_{i=1}^{N} \mu_i p_i z_i,$$

where

$$\partial\rho(0) \subseteq \left\{ \mu \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^{N} p_i \mu_i = 1, \mu \geq 0 \right\}.$$

Theorem 1 is known as the *representation theorem of coherent risk measures.*

The representation theorem of coherent risk measures allow us to rewrite problem (2)-(3) in the form

$$\begin{aligned} \min_{x\in X} \max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} \mu_i p_i \left[c^\top x + \min_{y_i\in\mathbb{R}^m} \rho_2[q_i^\top y_i] \right] \\ \text{s.t. } T_i x + W_i y_i = h_i,\ i = 1,\dots,N \\ y_i \geq 0,\ i = 1,\dots,N. \end{aligned} \tag{4}$$

For $i = 1,\dots,N$, let $c_i \in \mathbb{R}^n$ and ρ_{2i} be coherent risk measures. For the remaining of this article we will focus on the following slight generalization of problem (4):

$$\begin{aligned} \min_{x\in X} \max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} \mu_i p_i \left[c_i^\top x + \min_{y_i\in\mathbb{R}^m} \rho_{2i}[q_i^\top y_i] \right] \\ \text{s.t. } T_i x + W_i y_i = h_i \\ y_i \geq 0\ ,\quad i = 1,\dots,N. \end{aligned} \tag{5}$$

4 The Dual

In this section we will obtain a dual formulation of problem (5).

Consider the following problem:

$$\begin{aligned} \min_{(x_1,\dots,x_N)\in X^N} \max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} \mu_i p_i \left[c_i^\top x_i + \min_{y_i\in\mathbb{R}^m} \rho_{2i}[q_i^\top y_i] \right] \\ \text{s.t. } T_i x_i + W_i y_i = h_i \\ y_i \geq 0,\ x_i \in X,\quad i = 1,\dots,N \end{aligned} \tag{6}$$

with the extra *nonanticipativity* constraints

$$x_i = \sum_{k=1}^{N} p_k x_k,\ i = 1,\dots,N. \tag{7}$$

The nonanticipativity constraints (7) play a double role in the development of our theory and methods. First, they ensure that the first decision variables of (6) do not depend on the second stage realization of the random data and thus making problem (6)–(7) equivalent to our main problem (5). Second, these extra constraints will help us obtain a dual formulation of (5) through their appearance in the Lagrangian of (6)–(7). We aim to do this next.

Let $\mathcal{X} = \mathbb{R}^{n\cdot N}$ and $\mathcal{L} = \{x = (x_1,\dots,x_N) \,|\, x_1 = \ldots = x_N\}$. Equip the space $\mathcal{X}$ with the scalar product

$$\langle x, y\rangle = \sum_{i=1}^{N} p_i x_i^\top y_i \tag{8}$$

and define the linear operator $P : \mathcal{X} \to \mathcal{X}$ as

$$Px = \left(\sum_{i=1}^{N} p_i x_i, \dots, \sum_{i=1}^{N} p_i x_i \right).$$

The nonanticipativity constraints (7) can be compactly written as

$$x = Px.$$

It can be verified that P is the orthogonal projection operator of $\mathcal{X}$, equipped with the scalar product (8), onto its subspace $\mathcal{L}$.

For every $x \in \mathbb{R}^N$ and $i \in \{1, \dots, N\}$ define

$$\begin{aligned} F(x, \omega_i) \triangleq \ & c_i^\top x + \inf_{y \in \mathbb{R}^m} \rho_{2i}[q_i^\top y] \\ & \text{s.t. } \ T_i x + W_i y = h_i \\ & \qquad y \geq 0. \end{aligned} \tag{9}$$

The convexity property (A1) of the coherent risk measure ρ_{2i} implies that $F(\cdot, \omega_i)$ is a convex function. Also the assumptions made on the second stage problem $\mathcal{Q}(x, \xi)$ imply that $F(X, \omega_i) \subset \mathbb{R}$. Therefore the compactness of X implies that $F(\cdot, \omega_i)$ is continuous relative to X (see [6] Theorem 10.4).

By assigning Lagrange multipliers $\lambda_k \in \mathbb{R}^n, k = 1, \dots, N$, to the nonanticipativity constraints (7), we obtain that the Lagrangian of problem (6)–(7) is given by:

$$L(x, \lambda) \triangleq \max_{\mu \in \partial \rho_1(0)} \sum_{i=1}^{N} \mu_i p_i F(x_i, \omega_i) + \sum_{j=1}^{N} p_j \lambda_j^\top \left(x_j - \sum_{t=1}^{N} p_t x_t \right),$$

where $x = (x_1, \dots, x_N) \in X^N$ and $\lambda^\top = (\lambda_1^\top, \dots, \lambda_N^\top)$. Note that since P is an orthogonal projection, $I - P$ is also an orthogonal projection (onto the space orthogonal to $\mathcal{L}$), and hence

$$\sum_{j=1}^{N} p_j \lambda_j^\top \left(x_j - \sum_{t=1}^{N} p_t x_t \right) = \langle \lambda, (I - P)x \rangle = \langle (I - P)\lambda, x \rangle.$$

Therefore the above Lagrangian can be written in the following equivalent form

$$L(x, \lambda) = \max_{\mu \in \partial \rho_1(0)} \sum_{i=1}^{N} \mu_i p_i F(x_i, \omega_i) + \sum_{j=1}^{N} p_j \left(\lambda_j - \sum_{t=1}^{N} p_t \lambda_t \right)^\top x_j.$$

Observe that shifting the multipliers $\lambda_j, j = 1, \dots, N$, by a constant vector does not change the value of the Lagrangian, because the expression $\lambda_j - \sum_{t=1}^{N} p_t \lambda_t$ is invariant to such shifts. Therefore, with no loss of generality we can assume that

$$\sum_{j=1}^{N} p_j \lambda_j = 0$$

or equivalently, that $P\lambda = 0$. Under the condition $P\lambda = 0$, the Lagrangian can be written simply as

$$L(x,\lambda) = \max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} p_i \left[\mu_i F(x_i,\omega_i) + \lambda_i^\top x_i\right].$$

Putting everything together we obtain the following dual formulation of problem (6)–(7):

$$\max_{\lambda\in\mathbb{R}^{n\cdot N}} \min_{x\in X^N} \max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} p_i \left[\mu_i F(x_i,\omega_i) + \lambda_i^\top x_i\right] \tag{10}$$
$$\text{s.t.} \quad \sum_{j=1}^{N} p_j\lambda_j = 0.$$

Note that X is polyhedral, $\partial\rho_1(0)$ is convex and compact, and $\mu_i F(x_i,\omega_i) + \lambda_i^\top x_i$ is linear in μ and convex in x. Therefore we can interchange the innermost max by the min in (10) and obtain the following equivalent formulation for the dual of (6)–(7):

$$\max_{\substack{\lambda\in\mathbb{R}^{n\cdot N}\\ \mu\in\partial\rho_1(0)}} \sum_{i=1}^{N} p_i \min_{x_i\in X} \left[\mu_i F(x_i,\omega_i) + \lambda_i^\top x_i\right] \tag{11}$$
$$\text{s.t.} \quad \sum_{j=1}^{N} p_j\lambda_j = 0.$$

We will require that our problem satisfies the Slater's constraint qualification condition and in this way guaranteeing that the duality gap will be zero. We should first point out that the convexity of the functions $F(\cdot,\omega_i)$ imply that the primal objective function

$$\max_{\mu\in\partial\rho_1(0)} \sum_{i=1}^{N} \mu_i p_i\, F(x,\omega_i)$$

is convex too. Due to this fact and the relative to X continuity of $F(\cdot,\omega_i)$, problem (6)–(7) satisfies the Slater's constraint qualification condition if there is $x = (x_1,\ldots,x_N) \in \operatorname{int} X^N$ such that

$$x_i = \sum_{k=1}^{N} p_k x_k,\ i = 1,\ldots,N.$$

In this case the duality gap is zero (see [7] Theorem 4.7) and so, the solution of the primal problem (6)–(7) is the same as the solution of the dual problem (11).

Suppose that once we are in scenario i there are exactly N_i possible sub-scenarios that could occur each with probability $\pi_{ij}, j = 1,\ldots,N_i$ and its own vector q_{ij}. Clearly $\pi_{ij} > 0$ and $\sum_{j=1}^{N_i} \pi_{ij} = 1$. The representation theorem of coherent risk measures shows that

$$\rho_{2i}[q_i^\top y] = \max_{\delta\in\partial\rho_{2i}(0)} \sum_{j=1}^{N_i} \delta_j \pi_{ij} q_{ij}^\top y, \tag{12}$$

where

$$\partial\rho_{2i}(0) \subseteq \left\{ \delta \in \mathbb{R}^{N_i} \;\middle|\; \sum_{j=1}^{N_i} \delta_j \pi_{ij} = 1, \delta \geq 0 \right\}.$$

From (9), (11), and (12) we obtain that the dual of the main problem is equivalent to

$$\begin{aligned} \max_{\substack{\lambda \in \mathbb{R}^{n \cdot N} \\ \mu \in \partial\rho_1(0)}} \min_{x \in X^N} \; & \sum_{i=1}^{N} p_i \left[\mu_i c_i^\top x_i + \mu_i \min_{y_i \in \mathbb{R}^m} \left[\max_{\delta_i \in \partial\rho_{2i}(0)} \sum_{j=1}^{N_i} \delta_{ij} \pi_{ij} q_{ij}^\top y_i \right] + \lambda_i^\top x_i \right] \\ \text{s.t.} \;\; & T_i x_i + W_i y_i = h_i, \;\; i = 1, \ldots, N \\ & y_i \geq 0, \;\; i = 1, \ldots, N \\ & \sum_{j=1}^{N} p_j \lambda_j = 0. \end{aligned} \tag{13}$$

This, in turn, is equivalent to

$$\begin{aligned} \max_{\substack{\lambda \in \mathbb{R}^{n \cdot N} \\ \mu \in \partial\rho_1(0)}} \min_{\substack{x \in X^N \\ y \in \mathbb{R}^{m \cdot N}}} \max_{\delta \in \partial\rho_2(0)} \; & \sum_{i=1}^{N} p_i \left[\mu_i c_i^\top x_i + \mu_i \sum_{j=1}^{N_i} \delta_{ij} \pi_{ij} q_{ij}^\top y_i + \lambda_i^\top x_i \right] \\ \text{s.t.} \;\; & T_i x_i + W_i y_i = h_i, \;\; i = 1, \ldots, N \\ & y_i \geq 0, \;\; i = 1, \ldots, N \\ & \sum_{j=1}^{N} p_j \lambda_j = 0, \end{aligned} \tag{14}$$

where $\partial\rho_2(0) := \partial\rho_{21}(0) \times \cdots \times \partial\rho_{2N}(0)$, $y = (y_1, \ldots, y_N)$, and $\delta = (\delta_1, \ldots, \delta_N)$. As before, we can interchange the innermost max and min and obtain the following equivalent formulation

$$\begin{aligned} \max_{\substack{\lambda \in \mathbb{R}^{n \cdot N} \\ \mu \in \partial\rho_1(0) \\ \delta \in \partial\rho_2(0)}} \min_{\substack{x \in X^N \\ y \in \mathbb{R}^{m \cdot N}}} \; & \sum_{i=1}^{N} p_i \left[\mu_i c_i^\top x_i + \mu_i \sum_{j=1}^{N_i} \delta_{ij} \pi_{ij} q_{ij}^\top y_i + \lambda_i^\top x_i \right] \\ \text{s.t.} \;\; & T_i x_i + W_i y_i = h_i, \;\; i = 1, \ldots, N \\ & y_i \geq 0, \;\; i = 1, \ldots, N \\ & \sum_{j=1}^{N} p_j \lambda_j = 0. \end{aligned} \tag{15}$$

Let

$$S := \left\{ \begin{pmatrix} \pi_{11}\mu_1\delta_{11} \\ \vdots \\ \pi_{ij}\mu_i\delta_{ij} \\ \vdots \\ \pi_{NN_N}\mu_N\delta_{NN_N} \end{pmatrix} \;\middle|\; \mu_i \in [\partial\rho_1(0)]^i, \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{iN_i} \end{pmatrix} \in \partial\rho_{2i}(0), \;\; i = 1, \ldots, N \right\},$$

where $[\partial\rho_1(0)]^i$ is the projection of $\partial\rho_1(0)$ on the ith axis. In other words, $\mathcal{S} = \pi \bullet \partial\rho_2(0) \bullet \bigotimes_{i=1}^{N} [\partial\rho_1(0)]^i$, where $\bullet$ denotes the Hadamard product of vectors. Note that $\mathcal{S}$ is a convex and compact set. Then (15) is equivalent to:

$$
\begin{aligned}
\max_{\substack{\lambda\in\mathbb{R}^{n\cdot N}\\ \mu\in\partial\rho_1(0)\\ \sigma\in\mathcal{S}}} \ \min_{\substack{x\in X^N\\ y\in\mathbb{R}^{m\cdot N}}} \ & \sum_{i=1}^{N} p_i \left[\mu_i c_i^\top x_i + \sum_{j=1}^{N_i} \sigma_{ij} q_{ij}^\top y_i + \lambda_i^\top x_i\right] && (16)\\
\text{s.t.} \quad & T_i x_i + W_i y_i = h_i,\ i = 1,\dots,N\\
& y_i \geq 0,\ i = 1,\dots,N\\
& \sum_{j=1}^{N} p_j \lambda_j = 0.\\
& \sum_{j=1}^{N_i} \sigma_{ij} = \mu_i\,,\ i = 1,\dots,N
\end{aligned}
$$

Let $\mathbb{R}^{\widetilde{N}} := \mathbb{R}^{N_1} \times \cdots \times \mathbb{R}^{N_N}$. For every scenario $i = 1,\dots,N$ define the function $\chi_i : \mathbb{R}^{n\cdot N} \times \mathbb{R}^N \times \mathbb{R}^{\widetilde{N}} \to \overline{\mathbb{R}}$ such that

$$
\begin{aligned}
\chi_i(\lambda,\mu,\sigma) \triangleq \min_{x,y} \ & \mu_i c_i^\top x + \sum_{j=1}^{N_i} \sigma_{ij} q_{ij}^\top y + \lambda_i^\top x && (17)\\
\text{s.t.} \quad & T_i x + W_i y = h_i && (18)\\
& x \in X,\ y \in \mathbb{R}^m,\ y \geq 0.
\end{aligned}
$$

Then the dual of our main problem is given by:

$$
\begin{aligned}
\max_{\lambda,\mu,\sigma} \ & \sum_{i=1}^{N} p_i \, \chi_i(\lambda,\mu,\sigma) && (19)\\
\text{s.t.} \quad & \sum_{j=1}^{N} p_j \lambda_j = 0\\
& \sum_{j=1}^{N_i} \sigma_{ij} = \mu_i\,,\ i = 1,\dots,N\\
& \lambda \in \mathbb{R}^{n\cdot N},\ \mu \in \partial\rho_1(0),\ \sigma \in \mathcal{S}.
\end{aligned}
$$

At this point we add the condition that the sets $Y_i := \{y \in \mathbb{R}^m \mid \exists x \in X \text{ s.t. } T_i x + W_i y = h_i\}$ be compact for every $i = 1,\dots,N$. Then $\chi_i(\lambda,\mu,\sigma) \in \mathbb{R}$, for every $(\lambda,\mu,\sigma) \in \mathbb{R}^{n\cdot N} \times \mathbb{R}^N \times \mathbb{R}^{\widetilde{N}}$ and χ_i is proper, concave, and Lipschitz continuous (see [6] Theorem 10.4). Also the set $\mathcal{A}_i$ of pairs (x,y) satisfying the system of constraints (18) is compact.

5 The Subgradient

Define the sets

$$\Lambda \triangleq \left\{ \lambda \in \mathbb{R}^{n \cdot N} \;\middle|\; \sum_{i=1}^{N} p_i \lambda_i = 0 \right\}, \tag{20}$$

$$\Delta \triangleq \left\{ (\mu, \sigma) \in \partial \rho_1(0) \times \mathcal{S} \;\middle|\; \sum_{j=1}^{N_i} \sigma_{ij} = \mu_i \,,\ i = 1, \ldots, N \right\},$$

and the function $\vartheta_i(\cdot,\cdot) : \left(\mathbb{R}^{n \cdot N} \times \mathbb{R}^N \times \mathbb{R}^{\widetilde{N}} \right) \times \mathcal{A}_i \to \mathbb{R}$ by

$$\vartheta_i \left[(\lambda, \mu, \sigma), (x, y) \right] \triangleq \mu_i c_i^\top x + \sum_{j=1}^{N_i} \sigma_{ij} q_{ij}^\top y + \lambda_i^\top x. \tag{21}$$

The above definition implies that for every $(\lambda, \mu, \sigma) \in \mathbb{R}^{n \cdot N} \times \mathbb{R}^N \times \mathbb{R}^{\widetilde{N}}$

$$\chi_i(\lambda, \mu, \sigma) = \min_{(x,y) \in \mathcal{A}_i} \vartheta_i \left[(\lambda, \mu, \sigma), (x, y) \right]. \tag{22}$$

Let

$$\chi(\lambda, \mu, \sigma) \triangleq \sum_{i=1}^{N} p_i \, \chi_i(\lambda, \mu, \sigma). \tag{23}$$

Then, we can rewrite our main dual problem (19) as

$$\max_{(\lambda,\mu,\sigma) \in \Lambda \times \Delta} \chi(\lambda, \mu, \sigma), \tag{24}$$

where $\lambda \in \lambda$ and $(\mu, \sigma) \in \Delta$.

The main purpose of this section is to calculate the subdifferential of χ, which is key in the development of methods to solve efficiently problem (24). Notice first that by definition $\chi(\cdot,\cdot)$ is a proper, concave, and continuous function. Therefore we can apply the Moreau-Rockafellar theorem and obtain

$$\partial \chi(\lambda, \mu, \sigma) = \sum_{i=1}^{N} p_i \partial \chi_i(\lambda, \mu, \sigma). \tag{25}$$

Because of this we will focus on the subdifferentials of the χ_i's. Definition (21) allows us to see that the following proposition holds.

Proposition 2. *The function $\vartheta_i \left[\,\cdot\, , (x, y) \right]$ is concave for every $(x, y) \in \mathcal{A}_i$. Also, the function $\vartheta_i \left[(\lambda, \mu, \sigma), \,\cdot\, \right]$ is lower semicontinuous for every $(\lambda, \mu, \sigma) \in \mathbb{R}^{n \cdot N} \times \mathbb{R}^N \times \mathbb{R}^{\widetilde{N}}$.*

Let $\partial\vartheta_i\,[(\lambda_0,\mu_0,\sigma_0),(x,y)]$ be the subdifferential of $\partial\vartheta_i\,[\,\cdot\,,(x,y)]$ at the point $(\lambda_0,\mu_0,\sigma_0)$ and let Q_i be the $N_i \times m$ matrix with rows $q_{ij}^\top$. Then

$$\partial\vartheta_i\,[(\lambda_0,\mu_0,\sigma_0),(x,y)]^\top = \left(\hat{x}_i^\top, e_i^\top \cdot c_i^\top x, \overline{Q_i y}^\top\right), \tag{26}$$

where $\hat{x}_i^\top := (0,\dots,0,\, x^\top, 0,\dots,0) \in \mathbb{R}^{n\cdot N}$ such that the x is in the ith position and each 0 is a vector of $\mathbb{R}^n$, and $\overline{Q_i y}^\top := \big(0,\dots,0,\,(Q_i y)^\top, 0,\dots,0\big) \in \mathbb{R}^{\tilde{N}}$ such that the $(Q_i y)^\top$ is in the ith position and a 0 in position t is a vector of $\mathbb{R}^{N_t}$. For every $(\lambda,\mu,\sigma) \in \mathbb{R}^{n\cdot N}\times\mathbb{R}^N\times\mathbb{R}^{\tilde{N}}$ define

$$\mathcal{A}_i(\lambda,\mu,\sigma) \triangleq \arg\min_{(x,y)\in\mathcal{A}_i}\left\{\mu_i c_i^\top x + \sum_{j=1}^{N_i}\sigma_{ij} q_{ij}^\top y + \lambda_i^\top x\right\}.$$

The set $\mathcal{A}_i(\lambda,\mu,\sigma)$ is the set of optimal solutions to the right hand side of (22). Then, since the function χ_i is a minimum function (see (22) and [7] Theorem 2.87), we get

$$\begin{aligned}\partial\chi_i(\lambda_0,\mu_0,\sigma_0)^\top &= \operatorname{conv}\left[\bigcup_{(x,y)\in\mathcal{A}_i(\lambda_0,\mu_0,\sigma_0)}\partial\vartheta_i\,[(\lambda_0,\mu_0,\sigma_0),(x,y)]^\top\right] \\ &= \left\{\left(\hat{x}_i^\top, e_i^\top\cdot c_i^\top x, \overline{Q_i y}^\top\right)\,\middle|\,(x,y)\in\mathcal{A}_i(\lambda_0,\mu_0,\sigma_0)\right\}.\end{aligned} \tag{27}$$

Using (25) and (27), we obtain

$$\partial\chi(\lambda_0,\mu_0,\sigma_0)^\top = \sum_{i=1}^{N} p_i\left\{\left(\hat{x}_i^\top, e_i^\top\cdot c_i^\top x, \overline{Q_i y}^\top\right)\,\middle|\,(x,y)\in\mathcal{A}_i(\lambda_0,\mu_0,\sigma_0)\right\}. \tag{28}$$

Usually we need to find a subgradient of χ at a given point $(\lambda_0,\mu_0,\sigma_0)$. From (28) we can derive a simple procedure to accomplish this:

Step 1. For every $i = 1,\dots,N$, solve the linear program

$$\min_{(x,y)\in\mathcal{A}_i}(\mu_0)_i\, c_i^\top x + \sum_{j=1}^{N_i}(\sigma_0)_{ij}\, q_{ij}^\top\, y + (\lambda_0)_i^\top x$$

and call the obtained optimal solution (x_i, y_i). Note that $(x_i,y_i)\in\mathcal{A}_i(\lambda_0,\mu_0,\sigma_0)$, for every $i = 1,\dots,N$.

Step 2. Compute $\alpha := \sum_{i=1}^{N} p_i\left((\hat{x}_i)_i^\top, e_i^\top\cdot c_i^\top(x_i), \overline{Q_i(y_i)}_i^\top\right)$.

Then (28) implies that $\alpha \in \partial\chi(\lambda_0,\mu_0,\sigma_0)$, as we wanted. Notice that the obtained subgradient α has the following structure:

$$\alpha = \begin{pmatrix} p_1 x_1 \\ \vdots \\ p_N x_N \\ p_1 c_1^\top x_1 \\ \vdots \\ p_N c_N^\top x_N \\ p_1 Q_1 y_1 \\ \vdots \\ p_N Q_n y_n \end{pmatrix}. \tag{29}$$

6 Duality And Differentiability of the Two Stage risk-averse Problem With Mean Upper Semideviation of First Order

It is possible to obtain different formulations of the dual problem when we restrict to consider specific risk measures in formulation (5). In this section we will consider problem (5) where ρ_1 and ρ_2 are both mean upper semideviations of first order. Let $\rho_1(Z) := \mathbb{E}[Z] + a_1 \mathbb{E}\left[Z - \mathbb{E}[Z]\right]_+$ and $\rho_{2i}(Z) := \mathbb{E}[Z] + b_i \mathbb{E}\left[Z - \mathbb{E}[Z]\right]_+$, where $i = 1, \ldots, N$ and $a_1, b_i \in [0, 1]$.

The structure of the subdifferential of the mean upper semideviation of first order is well known (see [12] page 278), namely, for every $Z \in \mathbb{R}^N$

$$\partial\rho_1(Z) = \left\{1 - \mathbb{E}[\xi] + \xi \mid \xi \in \arg\max Y\right\}, \tag{30}$$

where

$$Y = \left\{\langle \xi, Z - \mathbb{E}[Z]\rangle \mid \|\xi\|_\infty \leq a_1,\ \xi \geq 0\right\}.$$

If $Z = 0$ then $Y = \{0\}$ and clearly $\arg\max Y = Y$. Therefore

$$\begin{aligned} \arg\max Y &= \left\{\xi \mid \|\xi\|_\infty \leq a_1,\ \xi \geq 0\right\} \\ &= \left\{\xi \mid 0 \leq \xi_i \leq a_1, \forall\, i = 1, \ldots, N\right\}, \end{aligned} \tag{31}$$

here we are using that $\xi = (\xi_1, \ldots, \xi_N)$. Relations (30) and (31) show that

$$\partial\rho_1(0) = \left\{1 - \sum_{i=1}^{N} p_i \xi_i + \xi \,\middle|\, 0 \leq \xi_i \leq a_1\right\}. \tag{32}$$

Applying the representation theorem of the coherent risk measures (Theorem 1), we obtain that

$$\rho_{2i}(q_i^\top y) = \max_{\delta \in \partial\rho_{2i}(0)} \sum_{j=1}^{N_i} \delta_j \pi_{ij} q_{ij}^\top y, \tag{33}$$

where as before

$$\partial\rho_{2i}(0) = \left\{ 1 - \sum_{j=1}^{N_i} \pi_{ij}\xi_j + \xi \;\middle|\; 0 \le \xi_j \le b_i \right\}. \tag{34}$$

Substituting (34) into(33) gives

$$\rho_{2i}(q_i^\top y) = \max_{\xi\in[0,b_i]^{N_i}} \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} \xi_k \pi_{ik} \left[q_{ik}^\top y - \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y \right]. \tag{35}$$

Since $\pi_{ik} > 0$ and $\xi \in [0, b_i]^{N_i}$, the maximum on the right hand side of (35) is given by ξ such that

$$\xi_k = \begin{cases} b_i & \text{if } \; q_{ik}^\top y - \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y \ge 0 \\ 0 & \text{otherwise} \end{cases},$$

$k = 1, \dots, N_i$. Therefore, we can obtain $\rho_{2i}(q_i^\top y)$ by solving the following linear program

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} d_k \\ \text{subject to:} \quad & d_k \ge 0, \\ & d_k \ge b_i \pi_{ik} \left[q_{ik}^\top y - \mathbb{E}[q_i^\top y] \right], \\ & \text{for all } \; k = 1, \dots, N_i. \end{aligned} \tag{36}$$

where $\mathbb{E}[q_i^\top y] := \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y$.

For every scenario $i = 1, \dots, N$ define the function $\varphi_i : \mathbb{R}^{n\cdot N} \times \mathbb{R}^N \to \overline{\mathbb{R}}$ such that

$$\varphi_i(\lambda, \mu) \triangleq \min_{x\in X} \left[\mu_i F(x, \omega_i) + \lambda_i^\top x \right],$$

where $\lambda = (\lambda_1, \dots, \lambda_N)$. Then by (11), the dual of our problem is given by:

$$\begin{aligned} \max_{\substack{\lambda\in\mathbb{R}^{n\cdot N} \\ \mu\in\partial\rho_1(0)}} \quad & \sum_{i=1}^{N} p_i \, \varphi_i(\lambda, \mu) \\ \text{s.t.} \quad & \sum_{j=1}^{N} p_j \lambda_j = 0. \end{aligned} \tag{37}$$

Using the definition of $F(x, \omega_i)$ and the characterization of $\rho_{2i}(q_i^\top y)$ obtained in (36), we get

$$\varphi_i(\lambda, \mu) = \min_{x,y,d} \mu_i \left[c_i^\top x + \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} d_k \right] + \lambda_i^\top x \tag{38}$$

$$\begin{aligned}
&\text{subject to:} \\
&d_k \geq c_k \pi_{ik} \left[q_{ik}^\top y - \mathbb{E}[q_i^\top y] \right], \;\; k = 1, \ldots, N_i \\
&T_i x + W_i y = h_i \\
&x \in X, \;\; y \geq 0, \;\; d \geq 0.
\end{aligned} \tag{39}$$

Note that the compactness of X implies that $\varphi_i : \mathbb{R}^{n \cdot N} \times \mathbb{R}^N \to \mathbb{R}$ is a proper concave function and therefore φ_i is Lipschitz continuous (see [6] Theorem 10.4). Formulation (37)–(39) is practical for the application of cutting plane methods.

For every $i = 1, \ldots, N$, let $\mathcal{B}_i$ be the set of triples (x, y, d) satisfying the system of inequalities (39). The set $\mathcal{B}_i$ is closed but not bounded. Nevertheless, our assumption of the second stage problem $\mathcal{Q}(x, \xi)$ having always a real optimal solution guarantees that we could assume (by adding extra constraints) that $\mathcal{B}_i$ is bounded. So, from now on we will assume that $\mathcal{B}_i$ is closed and bounded.

Define the function $\psi_i(\cdot, \cdot) : \left(\mathbb{R}^{n \cdot N} \times \mathbb{R}^N\right) \times \mathcal{B}_i \to \mathbb{R}$ by

$$\psi_i \left[(\lambda, \mu), (x, y, d)\right] \triangleq \mu_i \left[c_i^\top x + \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} d_k \right] + \lambda_i^\top x. \tag{40}$$

The above definition implies that for every $(\lambda, \mu) \in \mathbb{R}^{n \cdot N} \times \mathbb{R}^N$

$$\varphi_i(\lambda, \mu) = \min_{(x,y,d) \in \mathcal{B}_i} \psi_i \left[(\lambda, \mu), (x, y, d)\right]. \tag{41}$$

Let

$$\varphi(\lambda, \mu) \triangleq \sum_{i=1}^{N} p_i \, \varphi_i(\lambda, \mu). \tag{42}$$

Then, we can rewrite the dual problem (37) as

$$\max_{(\lambda,\mu) \in \Lambda \times \partial\rho_1(0)} \varphi(\lambda, \mu). \tag{43}$$

By definition $\varphi(\cdot, \cdot)$ is proper, concave, and continuous, so applying the Moreau-Rockafellar theorem we obtain:

$$\partial\varphi(\lambda, \mu) = \sum_{i=1}^{N} p_i \partial\varphi_i(\lambda, \mu). \tag{44}$$

By definition (40) we can easily see that the following proposition holds.

Proposition 3. *The function $\psi_i \left[\,\cdot\,, (x, y, d)\right]$ is concave for every $(x, y, d) \in \mathcal{B}_i$. Also, the function $\psi_i \left[(\lambda, \mu), \,\cdot\,\right]$ is upper semicontinuous for every $(\lambda, \mu) \in \mathbb{R}^{n \cdot N} \times \mathbb{R}^N$.*

For every (x, y, d) let

$$\phi_i(x, y, d) = c_i^\top x + \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} d_k$$

and let $\partial\psi_i\left[(\lambda_0,\mu_0),(x,y,d)\right]$ be the subdifferential of $\partial\psi_i\left[\,\cdot\,,(x,y,d)\right]$ at the point (λ_0,μ_0). Then

$$\partial\psi_i\left[(\lambda_0,\mu_0),(x,y,d)\right]^\top = \left(\hat{x}_i^\top, e_i^\top\cdot\phi_i(x,y,d)\right), \tag{45}$$

recall that $\hat{x}_i^\top = (0,\ldots,0,\, x^\top, 0,\ldots,0)\in\mathbb{R}^{n\cdot N}$ has the x is in the ith position and each 0 is a vector of $\mathbb{R}^n$. Also, for every $(\lambda,\mu)\in\mathbb{R}^{n\cdot N}\times\mathbb{R}^N$ define

$$\mathcal{B}_i(\lambda,\mu) \triangleq \underset{(x,y,d)\in\mathcal{B}_i}{\arg\min}\left\{\mu_i\phi_i(x,y,d)+\lambda_i^\top x\right\}.$$

Clearly the set $\mathcal{B}_i(\lambda,\mu)$ is the set of optimal solutions to the right hand side of (41).

Then, as before,

$$\partial\varphi_i(\lambda_0,\mu_0)^\top = \left\{\left(\hat{x}_i^\top, e_i^\top\cdot\phi_i(x,y,d)\right)\,\middle|\,(x,y,d)\in\mathcal{B}_i(\lambda_0,\mu_0)\right\}, \tag{46}$$

and

$$\partial\varphi(\lambda_0,\mu_0)^\top = \sum_{i=1}^N p_i\left\{\left(\hat{x}_i^\top, e_i^\top\cdot\phi_i(x,y,d)\right)\,\middle|\,(x,y,d)\in\mathcal{B}_i(\lambda_0,\mu_0)\right\}. \tag{47}$$

Just as before, (47) give us a simple procedure to obtain a subgradient of φ at (λ_0,μ_0):

Step 1. For every $i=1,\ldots,N$, solve the linear program

$$\min_{(x,y,d)\in\mathcal{B}_i}(\mu_0)_i\,\phi_i(x,y,d)+(\lambda_0)_i^\top x$$

and call the obtained optimal solution (x_i,y_i,d_i). Note that $(x_i,y_i,d_i)\in\mathcal{B}_i(\lambda_0,\mu_0)$, for every $i=1,\ldots,N$.

Step 2. Compute $\alpha := \sum_{i=1}^N p_i\left((\widehat{x}_i)_i^\top, e_i^\top\cdot\phi_i(x_i,y_i,d_i)\right)$.

Then (47) implies that $\alpha\in\partial\varphi(\lambda_0,\mu_0)$, as we wanted. Notice that the obtained subdifferential α is very simple:

$$\alpha = \begin{pmatrix} p_1\,x_1 \\ \vdots \\ p_N\,x_N \\ p_1\,\phi(x_1,y_1,d_1) \\ \vdots \\ p_N\,\phi(x_N,y_N,d_N)\end{pmatrix}. \tag{48}$$

From now on we will concentrate on developing methods specific for the two stage risk-averse problem with mean upper semideviation of the first order.

7 The Dual Cutting Plane Method

We will apply the cutting plane method to the main dual problem:

$$\max_{(\lambda,\mu)\in\Lambda\times\partial\rho_1(0)} \varphi(\lambda,\mu), \tag{49}$$

where

$$\varphi(\lambda,\mu) = \sum_{i=1}^{N} p_i\,\varphi_i(\lambda,\mu), \tag{50}$$

and for every $i = 1,\ldots,N$ and $(\lambda,\mu) \in \Lambda \times \partial\rho_1(0)$,

$$\varphi_i(\lambda,\mu) = \min_{x,y,d} \mu_i \left[c_i^\top x + \sum_{j=1}^{N_i} \pi_{ij} q_{ij}^\top y + \sum_{k=1}^{N_i} d_k \right] + \lambda_i^\top x \tag{51}$$

$$\begin{aligned}
&\text{subject to:}\\
&d_k \geq c_k \pi_{ik} \left[q_{ik}^\top y - \mathbb{E}[q_i^\top y] \right], \;\; k = 1,\ldots,N_i\\
&T_i x + W_i y = h_i\\
&x \in X, \;\; y \geq 0, \;\; d \geq 0.
\end{aligned}$$

As we stated before the functions $\varphi_i : \mathbb{R}^{n\cdot N} \times \mathbb{R}^N \to \overline{\mathbb{R}}$ are proper, concave and Lipschitz continuous. The set $\Lambda \times \partial\rho_1(0)$ is closed but might not be bounded. In order to be assured of finding a solution with the cutting plane method we should add, if necessary, artificial constraints to Λ in such a way that $\Lambda \times \partial\rho_1(0)$ is compact. These constraints should be large enough so that the new problem contains an optimal solution to the original dual problem. Because of this, from now on we will assume that $\Lambda \times \partial\rho_1(0)$ is a compact set.

All the properties mentioned in the previous paragraph are the theoretical requirements that guarantee the convergence to an optimal solution of the cutting plane method when applied to problem (49) (see [7] page 357). The idea behind the *cutting plane method* is to use the subgradient inequality,

$$\varphi(\lambda,\mu) \leq \varphi(\lambda_0,\mu_0) + \langle g, (\lambda,\mu) - (\lambda_0,\mu_0) \rangle \,,$$

which holds true for every $(\lambda,\mu) \in \mathbb{R}^{n\cdot N} \times \mathbb{R}^N$ and each subgradient $g \in \partial\varphi(\lambda_0,\mu_0)$, for constructing upper approximations of $\varphi(\cdot)$ (remember that $\varphi(\cdot)$ is concave). At each step the method refines the approximation to $\varphi(\cdot)$ and selects point which is the "best so far" approximation to an optimal solution of (49).

The method starts at a given point $(\lambda^1,\mu^1) \in \Lambda \times \partial\rho_1(0)$, calculates $g^1 \in \partial\varphi(\lambda^1,\mu^1)$, and constructs a linear approximation of $\varphi(\cdot)$:

$$\varphi^1(\lambda,\mu) \triangleq \varphi(\lambda^1,\mu^1) + \left\langle g^1, (\lambda,\mu) - (\lambda^1,\mu^1) \right\rangle .$$

In a general iteration k, having already generated points $(\lambda^1, \mu^1), \dots, (\lambda^k, \mu^k)$, values of the function $\varphi(\lambda^1, \mu^1), \dots, \varphi(\lambda^k, \mu^k)$, and corresponding subgradients $g^1, \dots, g^k$, the method construct an upper approximation of the function $\varphi(\cdot)$

$$\varphi^k(\lambda, \mu) \triangleq \min_{1 \le j \le k} \left[\varphi(\lambda^j, \mu^j) + \left\langle g^j, (\lambda, \mu) - (\lambda^j, \mu^j) \right\rangle \right]. \tag{52}$$

Then it solves the *master problem*:

$$\underset{(\lambda,\mu) \in \Lambda \times \partial \rho_1(0)}{\text{maximize}} \; \varphi^k(\lambda, \mu), \tag{53}$$

and add its solution $(\lambda^{k+1}, \mu^{k+1})$ to the set of points. After evaluating $\varphi(\lambda^{k+1}, \mu^{k+1})$ and $g^{k+1} \in \partial \varphi(\lambda^{k+1}, \mu^{k+1})$, it increases k by one, and continue the calculations. If

$$\varphi(\lambda^{k+1}, \mu^{k+1}) = \varphi^k(\lambda^{k+1}, \mu^{k+1}),$$

then the method stops; at this moment the point $\varphi(\lambda^{k+1}, \mu^{k+1})$ is optimal ([7] section 7.2).

The master problem (53) is equivalent to the following constrained optimization problem:

$$\begin{aligned} \text{maximize} \quad & z \\ \text{subject to} \quad & z \le \varphi(\lambda^j, \mu^j) + \left\langle g^j, (\lambda, \mu) - (\lambda^j, \mu^j) \right\rangle, \quad j = 1, \dots, k, \\ & (\lambda, \mu) \in \Lambda \times \partial \rho_1(0), \end{aligned} \tag{54}$$

whose solution $\left[(\lambda^{k+1}, \mu^{k+1}), z^{k+1}\right]$ is the next approximation to the solution of (49) and an upper bound for $\varphi(\cdot)$ on $\Lambda \times \partial \rho_1(0)$. This new formulation of the master problem has the advantage that, after passing to iteration $k + 1$, just one constraint is added to this problem, and re-optimization by a dual method is an attractive option. This is particularly useful since the set $\Lambda \times \partial \rho_1(0)$ is polyhedral and problem (54) is a linear program, for which efficient linear programming techniques can be employed.

Now we will show an explicit reformulation of (54). At each iteration k, solve for every $i = 1, \dots, N$ the problem (38)-(39) with $(\lambda, \mu) := (\lambda^k, \mu^k)$. Denote correspondingly by β_i^k and (x_i^k, y_i^k, d_i^k) the obtained optimal value and optimal solution. Then $\phi(\lambda^k, \mu^k) = \sum_{i=1}^N p_i \, \beta_i^k$ and by equation (48) we obtain that

$$\begin{pmatrix} p_1 \, x_1^k \\ \vdots \\ p_N \, x_N^k \\ p_1 \, \phi(x_1^k, y_1^k, d_1^k) \\ \vdots \\ p_N \, \phi(x_N^k, y_N^k, d_N^k) \end{pmatrix} \in \partial \varphi(\lambda^k, \mu^k).$$

This, the definition of Λ (see 20), and the characterization of $\partial \rho_1(0)$ (see (32)) gives the following reformulation of the master problem:

$$
\begin{aligned}
&\text{maximize} \quad z \\
&\text{subject to:} \\
&z \leq \varphi(\lambda^j, \mu^j) + \left\langle \begin{pmatrix} p_1\, x_1^j \\ \vdots \\ p_N\, x_N^j \end{pmatrix}, \lambda - \lambda^j \right\rangle + \left\langle \begin{pmatrix} p_1\, \phi(x_1^j, y_1^j, d_1^j) \\ \vdots \\ p_N\, \phi(x_N^j, y_N^j, d_N^j) \end{pmatrix}, \mu - \mu^j \right\rangle, \\
&\qquad \text{for all } j = 1, \ldots, k, \\
&\qquad \sum_{i=i}^{N} p_i \lambda_i = 0, \\
&\qquad \mu_i = 1 - \sum_{i=1}^{N} p_i \xi_i + \xi_i,\ i = 1, \ldots, N, \\
&\qquad 0 \leq \xi_i \leq a_1,\ i = 1, \ldots, N,
\end{aligned}
\tag{55}
$$

where $\mu = (\mu_1, \ldots, \mu_N)$. This formulation is concise and practical for implementations of the method.

8 New Methods

The dual cutting plane method took advantage of all the particular calculations that we developed for problem (5), its dual, and their restatements. Many of these calculations are economical and have a positive impact in the running time and performance of the dual cut method. There are, however, some more possible routes of optimization that we have not explored yet. First, we could apply a more sophisticated "cutting plane type" methods such at the bundle method to the dual problem (49). Second, we could exploit the geometrical properties of the feasible region of the problem to simplify the required calculations on the selected method. This is exactly what we set to do in the following sections. These new methods exploit the features in the domain of the objective function to reduce the number of variables that will be involved in the quadratic master problem. We call these new methods the *truncated proximal point* and *truncated bundle* methods.

The methods developed in this section follow the literature and act upon convex functions. Despite this, our main goal is to apply these methods on problem (49) which is concave. This should not present any problem since most of the "convex" results could be easily translated into "concave" and later we will do so without explicitly stating it.

8.1 Truncated Proximal Point Method

8.1.1 Truncated Moreau-Yoshida regularization

Consider a convex function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ that is proper, and lower semicontinuous.

For a fixed number $\varrho > 0$, we define the function $f_\varrho : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ by

$$f_\varrho(v, w) \triangleq \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \left\{ \frac{\varrho}{2}\|y - w\|^2 + f(x, y) \right\}. \tag{56}$$

The function f_ϱ is called the *truncated Moreau-Yosida regularization* of f. Since $f(x, y)$ is convex and lower semicontinuous the function

$$F(y) \triangleq \inf_{x\in\mathbb{R}^n} f(x, y)$$

is also convex and lower semicontinuous.

Unfortunately the properties of f do not imply that F is a proper function. For example the function $f(x, y) = x$ satisfies all the properties stated above but $F(y) = \inf_{x\in\mathbb{R}^m} x$ is not proper. A proper function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is *x-bounded* if for every bounded $Y \subset \mathbb{R}^m$ the set $X := \{(x, y) \in \mathbb{R}^n \times Y \mid f(x, y) \in \mathbb{R}\}$ is bounded. Notice that if f is x-bounded then the corresponding function F is proper. From now on we will assume that f is x-bounded.

Let $\overline{f}_\varrho : \mathbb{R}^m \to \overline{\mathbb{R}}$ be defined by

$$\overline{f}_\varrho(w) \triangleq \min_{y\in\mathbb{R}^m} \left\{ \frac{\varrho}{2}\|y - w\|^2 + F(y) \right\}. \tag{57}$$

Then $\overline{f}_\varrho$ is the Moreau-Yoshida regularization of F and it is well known (see [7] Lemma 7.10) that $\overline{f}_\varrho$ is a real-valued, convex and continuously differentiable function with $\nabla\overline{f}_\varrho(w) = \varrho\,(w - y_\varrho(w))$, where $y_\varrho(w)$ is the solution of (57). It is not difficult to see that from the properties of f it follows that $f_\varrho(v, w) = \overline{f}_\varrho(w)$ for all $(v, w) \in \mathbb{R}^n \times \mathbb{R}^m$. Therefore we can conclude the following about the truncated Moreau-Yoshida regularization of x-bounded functions:

Theorem 4. *For every $\varrho > 0$, the function f_ϱ is real-valued, convex and continuously differentiable with $\nabla f_\varrho(v, w) = [0, \varrho\,(w - y_\varrho(w))]$, where $(x_\varrho(v), y_\varrho(w))$ is any solution of* (56).

8.1.2 Application to convex optimization

Let us consider the convex optimization problem

$$\operatorname*{minimize}_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \; f(x, y), \tag{58}$$

in which $f : \mathbb{R}^n\times\mathbb{R}^m \to \overline{\mathbb{R}}$ is convex, proper, lower semicontinuous and x-bounded. Using the truncated Moreau-Yoshida regularization of f, we construct the following iterative process. At iteration k, given the current approximation (v^k, w^k) to the solution of (58), we find a point $(x^k, y^k) = (x_\varrho(v^k), y_\varrho(w^k))$, which is a solution of the problem

$$\operatorname*{minimize}_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \; \frac{\varrho}{2}\|y - w^k\|^2 + f(x, y). \tag{59}$$

The next approximation is defined according to the formula:

$$\left(v^{k+1}, w^{k+1}\right) = \left(x_\varrho(v^k), y_\varrho(w^k)\right), \; k = 1, 2, \ldots \tag{60}$$

The iterative method (60) is called the *truncated proximal point method.* Although we will not directly apply this method, it is of theoretical importance and a natural progression in the development of the truncated bundle method in the following section.

Let us recall that it follows form Theorem 4 that if f is x-bounded then problem (59) has a solution. Thus the truncated proximal point method is well defined. Since $f_\varrho(v^k, w^k) \leq f(v^k, w^k)$ by construction, we have $f(v^{k+1}, w^{k+1}) \leq f(v^k, w^k)$, $k = 1, 2 \ldots$ Actually, the progress made at each iteration can be estimated.

Lemma 5. *Assume that there exists $(\tilde{x}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $f(\tilde{x}, \tilde{y}) < f(v, w)$. Then if $\tilde{y} = w$,*

$$f_\varrho(v, w) \leq f(v, w) - \left(f(v, w) - f(\tilde{x}, \tilde{y})\right),$$

else

$$f_\varrho(v, w) \leq f(v, w) - \varrho\|\tilde{y} - w\|^2 \varphi\left(\frac{f(v, w) - f(\tilde{x}, \tilde{y})}{\varrho\|\tilde{y} - w\|^2}\right),$$

where

$$\varphi(\tau) = \begin{cases} 0 & \text{if } \tau < 0, \\ \tau^2 & \text{if } 0 \leq \tau \leq 1, \\ -1 + 2\tau & \text{if } \tau > 1. \end{cases}$$

We conclude that in any case $f_\varrho(v, w) < f(v, w)$.

Proof. Consider the segment containing points $(x, y) = (v, w) + t\left((\tilde{x}, \tilde{y}) - (v, w)\right)$ with $0 \leq t \leq 1$. Restricting the minimization in (59) to these (x, y)'s provides an upper bound for the optimal value:

$$\begin{aligned} f_\varrho(v, w) &\leq \min_{0 \leq t \leq 1} \left[f\left((1 - t)\,(v, w) + t(\tilde{x}, \tilde{y})\right) + \frac{\varrho t^2}{2}\|\tilde{y} - w\|^2 \right] \\ &\leq f(v, w) + \min_{0 \leq t \leq 1} \left[t\left(f(\tilde{x}, \tilde{y}) - f(v, w)\right) + \frac{\varrho t^2}{2}\|\tilde{y} - w\|^2 \right]. \end{aligned} \tag{61}$$

In the last estimate we also used the convexity of f.

If $\tilde{y} = w$ then (61) implies that $f_\varrho(v, w) \leq f(v, w) - (f(v, w) - f(\tilde{x}, \tilde{y}))$. Else, $\tilde{y} \neq w$ and the value of t that minimizes (61) is equal to

$$\hat{t} = \min\left(1, \frac{f(v, w) - f(\tilde{x}, \tilde{y})}{\varrho\|\tilde{y} - w\|^2}\right).$$

Our assertion follows now from a straightforward calculation. □

At the solution $(x_\varrho(w), y_\varrho(w))$ of problem (59), Lemma 5 shows that $f\left(x_\varrho(w), y_\varrho(w)\right) \leq f_\varrho(v,w) < f(v,w)$. Therefore problem (59) will always find a better point if it exists. Consequently, $(x,y) = (v,w)$ is the minimizer in (59) if and only if (v,w) is a minimizer of f.

We say that a sequence $\left\{(x^k, y^k)\right\} \subset \mathbb{R}^n \times \mathbb{R}^m$ *approximates* an optimal solution (x^*, y^*) of (58) if $\lim_{k\to\infty} f(x^k, y^k) = f(x^*, y^*)$. In fact, the truncated proximal point method must approximate an optimal solution, if an optimal solution exists.

Theorem 6. *Assume that problem (58) has an optimal solution. Then the following holds.*

1. *The sequence $\left\{(v^k, w^k)\right\}$ generated by the truncated proximal point method approximates and optimal solution of (58).*
2. *The sequence $\left\{w^k\right\}$ converges to a point $\tilde{y}$ such that there is an optimal solution of (58) of the form $(\tilde{x}, \tilde{y})$.*

Proof. Let $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ be an optimal solution. We have the identity

$$\|w^{k+1} - y^*\|^2 = \|w^k - y^*\|^2 + 2\left\langle w^{k+1} - w^k, w^{k+1} - y^*\right\rangle - \|w^{k+1} - w^k\|^2. \tag{62}$$

Theorem (4) implies that:

$$\left[0, \varrho\left({}_\lambda w^k - {}_\lambda w^{k+1}\right)\right] \in \partial f(v^{k+1}, w^{k+1}). \tag{63}$$

By the definition of the subgradient,

$$f(x^*, y^*) \geq f(v^{k+1}, w^{k+1}) + \varrho\left\langle w^{k+1} - w^k, w^{k+1} - y^*\right\rangle. \tag{64}$$

Using this inequality in (62) (and skipping the last term) we obtain

$$\|w^{k+1} - y^*\|^2 \leq \|w^k - y^*\|^2 - \frac{2}{\varrho}\left[f(v^{k+1}, w^{k+1}) - f(x^*, y^*)\right] \tag{65}$$

Several conclusions follow from this estimate. First, summing up (65) from $k = 1$ to ∞, we get

$$\sum_{k=2}^{\infty}\left(f(v^k, w^k) - f(x^*, y^*)\right) \leq \frac{\varrho}{2}\|w^1 - y^*\|^2,$$

so $f(v^k, w^k) \to f(x^*, y^*)$ as $k \to \infty$. Therefore the sequence $\left\{(v^k, w^k)\right\}$ approximates and optimal solution of (58).

Second, the sequence $\left\{w^k\right\}$ is bounded and so it has accumulation points. Similarly, the x-boundedness of f implies that $\left\{v^k\right\}$ also has accumulation points. Consequently the lower semicontinuity of f implies that for every accumulation point $(\tilde{x}, \tilde{y})$ of $\left\{(v^k, w^k)\right\}$ we have $f(\tilde{x}, \tilde{y}) = f(x^*, y^*)$. We choose one such $(\tilde{x}, \tilde{y})$, substitute it for (x^*, y^*) in (65), and conclude that the sequence $\left\{w^k\right\}$ is convergent to $\tilde{y}$. □

8.2 The Truncated Bundle Method

8.2.1 The method

We consider the problem

$$\underset{(x,y)\in A}{\text{minimize}}\, f(x,y), \tag{66}$$

in which the set $A \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is closed convex and the function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is convex, proper, lower semicontinuous, and x-bounded. We obtain the following *regularized master problem*:

$$\underset{(x,y)\in A}{\text{minimize}}\, \frac{\varrho}{2}\|y - w^k\|^2 + f^k(x,y), \tag{67}$$

where the model f^k is defined by:

$$f^k(x,y) \triangleq \max_{j\in J_k} \left[f(x^j,y^j) + \left\langle g^j, (x,y) - (x^j,y^j) \right\rangle \right], \tag{68}$$

with $g^j \in \partial f(x^j,y^j)$, $j \in J_k$. The set J_k is a subset of $\{1,\ldots,k\}$ determined by a procedure for selecting cuts. At this moment we may think of J_k as being equal to $\{1,\ldots,k\}$.

In the proximal term $(\varrho/2)\|y-w^k\|^2$, where $\varrho > 0$, the center (v^k,w^k) is updated depending on the relations between the value of $f(x^{k+1},y^{k+1})$ at the mater's solution, (x^{k+1},y^{k+1}), and its prediction provided be the current model, $f^k(x^{k+1},y^{k+1})$. If these values are equal or close, we set $(v^{k+1},w^{k+1}) := (x^{k+1},y^{k+1})$ (*descent step*); otherwise $(v^{k+1},w^{k+1}) := (v^k,w^k)$ (*null step*). In any case, the collection of cuts is updated, and the iteration continues.

If the model f^k were exact, that is $f^k = f$, then problem (67) would be identical to the subproblem (59) solved at each iteration of the truncated proximal point method and we could just set $(v^{k+1},w^{k+1}) := (x^{k+1},y^{k+1})$. All steps would be descent steps. However, due to the approximate character of f^k, the solution of (67) is different than the solution of (59). It may not even be better than (v^k,w^k) (in terms of the value of the objective function). This is the reason for introducing conditional rules for updating the center (v^k,w^k).

On the other hand, the regularized master problem can be equivalently written as a problem with a quadratic objective function and linear constraints:

$$\begin{aligned} \text{minimize}\quad & z + \frac{\varrho}{2}\|y - w^k\|^2 \\ \text{subject to}\quad & z \geq f(x^j,y^j) + \left\langle g^j, (x,y) - (x^j,y^j) \right\rangle,\ j \in J_k, \\ & (x,y) \in A. \end{aligned} \tag{69}$$

If the set A is a convex polyhedron, the master problem can be readily solved by specialized techniques, enjoying the finite termination property.

Let us observe that problem (69) satisfies Slater's constraint qualification condition. Indeed, for every $(x_S,y_S) \in A$ we can choose z_s so large that all constraints are satisfied as strict inequalities. Therefore the optimal solution of the master problem satisfies the necessary and sufficient conditions of optimality with Lagrange multipliers (see [7] Theorem 3.34).

We denote by $\lambda_j^k \in J_k$, the Lagrange multipliers associated with the constraints of problem (69).

The detailed algorithm is stated in bellow. The parameter $\gamma \in (0,1)$ is a fixed constant used to compare the observed improvement in the objective value to the predicted improvement.

Step 0. Set $k := 1$, $J_0 := \emptyset$, and $z^1 := -\infty$.

Step 1. Calculate $f(x^k, y^k)$ and $g^k \in \partial f(x^k, y^k)$. If $f(x^k, y^k) > z^k$ then set $J_k := J_{k-1} \cup \{k\}$; otherwise set $J_k := J_{k-1}$.

Step 2. If $k = 1$ or if

$$f(x^k, y^k) \leq (1-\gamma) f(v^{k-1}, w^{k-1}) + \gamma f^{k-1}(x^k, y^k),$$

then set $(v^k, w^k) := (x^k, y^k)$; otherwise Step 2 is a null step and we set $(v^k, w^k) := (v^{k-1}, w^{k-1})$.

Step 3. Solve the master problem (69). Denote by (x^{k+1}, y^{k+1}) and z^{k+1} its solution and set $f^k(x^{k+1}, y^{k+1}) := z^{k+1}$.

Step 4. If $f^k(x^{k+1}, y^{k+1}) = f(v^k, w^k)$ then stop (the point (v^k, w^k) is an optimal solution); otherwise continue.

Step 5. If Step 2 was a null step then go to Step 6. Else (Step 2 was a descent step) remove from the set of cuts J_k some (or all) cuts whose Lagrange multipliers λ_j^k at the solution of (69) are 0.

Step 6. Increase k by one, and go to Step 1.

8.2.2 Convergence

First we prove that if the algorithm gets stuck at a w-center then it will approximate an optimal solution.

Lemma 7. *Let f^* be an optimal solution to (66) and suppose that the sequence, $\{(x^k, y^k)\}$, obtained by the truncated bundle method consists of only null steps from iteration t on. Then*

$$\lim_{k\to\infty} f^{k-1}(x^k, y^k) = f^* = \lim_{k\to\infty} f(x^k, y^k).$$

Proof. For any $\epsilon > 0$, let $\mathcal{K}_\epsilon := \{k \,|\, k > t \text{ and } f^{k-1}(x^k, y^k) + \epsilon < f(x^k, y^k)\}$ and let $k_1, k_2 \in \mathcal{K}_\epsilon$ with $t < k_1 < k_2$.

Since we only have null steps we get that for every $k > t$, $(v^k, w^k) = (x^t, y^t)$ and the cutting plane generated at k will remain on the master problem from k on. This implies that the sequence $\left\{f^{k-1}(x^k, y^k)\right\}$ is non-decreasing from $t+1$ on. Also, since the cutting plane generated at (x^{k_1}, y^{k_1}) will remain in the master problem at iteration $k_2 - 1$, we get:

$$f(x^{k_1}, y^{k_1}) + \left\langle g^{k_1}, (x^{k_2}, y^{k_2}) - (x^{k_1}, y^{k_1}) \right\rangle \leq f^{k_2-1}(x^{k_2}, y^{k_2}).$$

On the other hand, $\epsilon < f(x^{k_2}, y^{k_2}) - f^{k_2-1}(x^{k_2}, y^{k_2})$ which combined with the last inequality yields

$$\epsilon < f(x^{k_2}, y^{k_2}) - f(x^{k_1}, y^{k_1}) + \left\langle g^{k_1}, (x^{k_1}, y^{k_1}) - (x^{k_2}, y^{k_2}) \right\rangle.$$

Since all the steps made are null, the points y^k, with $k > t$, are contained in a bounded neighborhood of $w^k = y^t$. This and the x-boundedness of f guarantee us that $B := \operatorname{Conv}\{(x^j, y^j) \mid j \in \mathcal{K}_\epsilon\}$ is bounded. The function f is subdifferentiable in $\overline{B}$, so there exists a constant C such that $f(x_1, y_1) - f(x_2, y_2) \leq C\|(x_1, y_1) - (x_2, y_2)\|$, for all $x_1, x_2 \in \overline{B}$. Subgradients on bounded sets are bounded, and thus we can choose C large enough so that $\|g^j\| \leq C$, for all $j \in \mathcal{K}_\epsilon$. It follows from the last displayed inequality that

$$\epsilon < 2C\|(x^{k_1}, y^{k_1}) - (x^{k_2}, y^{k_2})\| \text{ for all } k_1, k_2 \in \mathcal{K}_\epsilon,\ k_1 \neq k_2.$$

As the set $\overline{B}$ is compact, there can exist only finitely many points in $\mathcal{K}_\epsilon \subset \overline{B}$ having distance at least $\epsilon/(2C)$ from each other. Thus the last inequality implies that the set $\mathcal{K}_\epsilon$ is finite for each $\epsilon > 0$. This means that

$$\lim_{k\to\infty} f(x^k) - f^{k-1}(x^k) = 0. \tag{70}$$

By construction the sequences $\left\{f^{k-1}(x^k)\right\}$ and $\left\{f(x^k)\right\}$ satisfy the relation

$$f^{k-1}(x^k) \leq f^* \leq f(x^k), \text{ for every } k \in \mathbb{N}.$$

Therefore the eventual monotonicity of $\left\{f^{k-1}(x^k)\right\}$ and (70) imply that

$$\lim_{k\to\infty} f^{k-1}(x^k, y^k) = f^* = \lim_{k\to\infty} f(x^k, y^k).$$

□

Next we prove another intermediate step towards convergence.

Lemma 8. *Assume that problem (66) has an optimal solution and suppose that the sequence,* $\left\{(x^k, y^k)\right\}$*, obtained by the truncated bundle method has infinitely many descent steps. Then the following holds.*

1. *The sequence* $\{(v^k, w^k)\}$ *generated by the truncated bundle method approximates an optimal solution of (66).*

2. *The sequence* $\left\{w^k\right\}$ *converges to a point* $\tilde{y}$ *such that there is an optimal solution of (66) of the form* $(\tilde{x}, \tilde{y})$*.*

Proof. Let us denote by $\mathcal{K}$ the set of iterations at which descent steps occur. If $(v^{k+1}, w^{k+1}) = (x^{k+1}, y^{k+1})$ is the optimal solution of the master problem (67), we have the necessary condition of optimality

$$0 \in \partial \left[\frac{\varrho}{2} \|y - w^k\|^2 + f^k(x, y) \right] + N_A(x, y) \;\; \text{at} \;\; (x, y) = (v^{k+1}, w^{k+1}).$$

Hence

$$- \left[0, \varrho(w^{k+1} - w^k)\right] \in \partial f^k(v^{k+1}, w^{k+1}) + N_A(v^{k+1}, w^{k+1}).$$

Let (x^*, y^*) be an optimal solution of (66). By the subgradient inequality for f^k we get (for some $h \in N_A(v^{k+1}, w^{k+1})$) the estimate

$$\begin{aligned} f^k(x^*, y^*) \geq\; & f^k(v^{k+1}, w^{k+1}) - \left\langle \left[0, \varrho \left(w^{k+1} - w^k\right)\right], (x^*, y^*) - (v^{k+1}, w^{k+1}) \right\rangle \\ & - \left\langle h, (x^*, y^*) - (v^{k+1}, w^{k+1}) \right\rangle \\ \geq\; & f^k(v^{k+1}, w^{k+1}) - \varrho \left\langle w^{k+1} - w^k, y^* - w^{k+1} \right\rangle . \end{aligned} \tag{71}$$

A descent step from (v^k, w^k) to (v^{k+1}, w^{k+1}) occurs, so the test of Step 2 is satisfied (for $k+1$):

$$f(v^{k+1}, w^{k+1}) \leq (1-\gamma) f(v^k, w^k) + \gamma f^k(v^{k+1}, w^{k+1}).$$

After elementary manipulations we can rewrite it as

$$f^k(v^{k+1}, w^{k+1}) \geq f(v^{k+1}, w^{k+1}) - \frac{1-\gamma}{\gamma} \left[f(v^k, w^k) - f(v^{k+1}, w^{k+1}) \right]. \tag{72}$$

Combining the last inequality with (71) and using the relation $f(x^*, y^*) \geq f^k(x^*, y^*)$ we obtain

$$f(x^*, y^*) \geq f(v^{k+1}, w^{k+1}) + \frac{1-\gamma}{\gamma} \left[f(v^{k+1}, w^{k+1}) - f(v^k, w^k) \right] - \varrho \left\langle w^{k+1} - w^k, y^* - w^{k+1} \right\rangle .$$

This can be substituted to the identity (62) which, after skipping the last term yields

$$\begin{aligned} \|w^{k+1} - y^*\|^2 \leq\; & \|w^k - y^*\|^2 - \frac{\varrho}{2} \left[f(v^{k+1}, w^{k+1}) - f(x^*, y^*) \right] \\ & + \frac{2(1-\gamma)}{\gamma \varrho} \left[f(v^k, w^k) - f(v^{k+1}, w^{k+1}) \right] \;\; \text{for all} \;\; k \in \mathcal{K}. \end{aligned} \tag{73}$$

The series $\sum_{k=1}^{\infty} [f(v^k, w^k) - f(v^{k+1}, w^{k+1})]$ is convergent, because $\{f(v^k, w^k)\}$ is non-increasing and bounded from below by $f(x^*, y^*)$. Therefore we obtain from (73) that the distance $\|w^{k+1} - y^*\|$ is uniformly bounded, and so $\{w^k\}$ must have accumulation points. This and the x-boundedness of f imply that the sequence $\{v^k, w^k\}$ has accumulation points.

Summing up (73) for $k \in \mathcal{K}$ we get

$$\sum_{k \in \mathcal{K}} \left(f(v^{k+1}, w^{k+1}) - f(x^*, y^*) \right) \leq \frac{\varrho}{2} \|w^1 - y^*\|^2 + \frac{1-\gamma}{\gamma} \left[f(v^1, w^1) - \lim_{k \to \infty} f(v^k, w^k) \right],$$

so $f(v^{k+1}, w^{k+1}) \to f(x^*, y^*)$, $k \in \mathcal{K}$. Consequently, at every accumulation point $(\tilde{x}, \tilde{y})$ of $\{(v^k, w^k)\}$ one has $f(\tilde{x}, \tilde{y}) = f(x^*, y^*)$. Since $(\tilde{x}, \tilde{y})$ is optimal, we can substitute it for (x^*, y^*) in (73). Skipping the negative term we get

$$\|w^{k+1} - \tilde{y}\|^2 \le \|w^k - \tilde{y}\|^2 + \frac{2(1-\gamma)}{\gamma\varrho}\left[f(v^k, w^k) - f(v^{k+1}, w^{k+1})\right].$$

It is true not only for $k \in \mathcal{K}$ but for all k, because at $k \notin \mathcal{K}$ we have a trivial equality here. Summing these inequalities from $k = l$ to $k = q > l$ we get

$$\|w^{q+1} - \tilde{y}\|^2 \le \|w^l - \tilde{y}\|^2 + \frac{2(1-\gamma)}{\gamma\varrho}\left[f(v^l, w^l) - f(v^{q+1}, w^{q+1})\right].$$

Since $\tilde{y}$ is an accumulation point, for $\epsilon > 0$ we can find l such that $\|w^l - \tilde{y}\| \le \epsilon$. Also, if l is large enough, $f(v^l, w^l) - f(v^{q+1}, w^{q+1}) \le \epsilon$ for all $q > l$, because $\{f(v^k, w^k)\}$ is convergent. Then $\|w^{q+1} - \tilde{y}\|^2 \le \epsilon^2 + 2\epsilon(1-\gamma)/(\gamma\varrho)$ for all $q > l$, so the sequence $\{w^k\}$ is convergent to $\tilde{y}$. □

Now we are ready to prove convergence of the truncated bundle method.

Theorem 9. *Assume that problem (66) has an optimal solution, f^*, and let $\{(x^k, y^k)\}$ be the sequence obtained by the truncated bundle method. Then*

$$\liminf_{k\to\infty} f(x^k, y^k) = f^*.$$

Proof. If there are only finitely many descent steps then Lemma 7 gives the desired result. Thus we assume that the number of descent steps is infinite and by Lemma 8, $\lim_{k\to\infty} f(v^k, w^k) = f^*$. Clearly the sequence $\{f(v^k, w^k)\}$ is an infinite subsequence of $\{f(x^k, y^k)\}$. Then, since $f(x^k, y^k) \ge f^*$ for every k, we obtain that

$$\liminf_{k\to\infty} f(x^k, y^k) = f^*.$$

□

9 Applying the Truncated Bundle Method

We are ready now to apply the truncated bundle method to the dual problem (49)-(51). Remember that $\Lambda \times \partial\rho_1(0)$ is concave, proper, compact and the function φ is proper, concave, continuous, and x-bounded. Therefore, we can apply the truncated bundle method modified for concave functions. The method starts at a given point $(\lambda^1, \mu^1) \in \Lambda \times \partial\rho_1(0)$ and continues as stated below:

Step 0. Set $k := 1$, $J_0 := \emptyset$, and $z^1 := -\infty$.

Step 1. For every $i = 1, \ldots, N$ solve the problem (38)-(39) with $(\lambda, \mu) := (\lambda^k, \mu^k)$. Denote correspondingly by β_i^k and (x_i^k, y_i^k, d_i^k) the obtained optimal value and optimal solution. Then

$$\varphi(\lambda^k, \mu^k) := \sum_{i=1}^{N} p_i \, \beta_i^k$$

and

$$g^k := \begin{pmatrix} p_1 \, x_1^k \\ \vdots \\ p_N \, x_N^k \\ p_1 \, \phi(x_1^k, y_1^k, d_1^k) \\ \vdots \\ p_N \, \phi(x_N^k, y_N^k, d_N^k) \end{pmatrix}.$$

If $\varphi(\lambda^k, \mu^k) < z^k$ then set $J_k := J_{k-1} \cup \{k\}$; otherwise set $J_k := J_{k-1}$.

Step 2. If $k = 1$ or if

$$\varphi(\lambda^k, \mu^k) \geq (1 - \gamma)\varphi(v^{k-1}, w^{k-1}) + \gamma\varphi^{k-1}(\lambda^k, \mu^k),$$

then set $(v^k, w^k) := (\lambda^k, \mu^k)$; otherwise Step 2 is a null step and we set $(v^k, w^k) := (v^{k-1}, w^{k-1})$.

Step 3. Solve the *regularized master problem*:

$$\begin{aligned} &\text{maximize } z - \frac{\varrho}{2}\|\lambda - w^k\|^2 \\ &\text{subject to:} \\ &z \leq \varphi(\lambda^j, \mu^j) + \left\langle \begin{pmatrix} p_1 \, x_1^j \\ \vdots \\ p_N \, x_N^j \end{pmatrix}, \lambda - \lambda^j \right\rangle + \left\langle \begin{pmatrix} p_1 \, \phi(x_1^j, y_1^j, d_1^j) \\ \vdots \\ p_N \, \phi(x_N^j, y_N^j, d_N^j) \end{pmatrix}, \mu - \mu^j \right\rangle, \\ &\qquad \text{for all } j = 1, \ldots, k, \\ &\qquad \sum_{i=i}^{N} p_i \lambda_i = 0, \\ &\qquad \mu_i = 1 - \sum_{i=1}^{N} p_i \xi_i + \xi_i, \ i = 1, \ldots, N, \\ &\qquad 0 \leq \xi_i \leq a_1, \ i = 1, \ldots, N, \end{aligned} \quad . \tag{74}$$

where $\mu = (\mu_1, \ldots, \mu_N)$. Denote by $(\lambda^{k+1}, \mu^{k+1})$ and z^{k+1} its solution and set $\varphi^k(\lambda^{k+1}, \mu^{k+1}) := z^{k+1}$.

Step 4. If $\varphi^k(\lambda^{k+1}, \mu^{k+1}) = \varphi(v^k, w^k)$ then stop (the point (v^k, w^k) is an optimal solution); otherwise continue.

Step 5. If Step 2 was a null step then go to Step 6. Else (Step 2 was a descent step) remove from the set of cuts J_k some (or all) cuts whose Lagrange multipliers λ_j^k at the solution of (74) are 0.

Step 6. Increase k by one, and go to Step 1.

To see the correctness of this method just notice that the vector $g^k \in \partial\varphi(\lambda^k, \mu^k)$ and problem (74) is a reformulation of the regularized master problem (67) with $A := \Lambda \times \partial\rho_1(0)$, $(x, y) := (\lambda^k, \mu^k)$, and $f := \varphi$. Then by Theorem (9) the solution obtained by the method approaches the optimal solution of the dual problem (49)-(51).

10 Application of the Methods and Results

Our aim is to apply the methods developed in previous sections to the following production problem.

Manufacturing and Transportation Problem. *There is a product line consisting of a few different models to be produced. Each model has its own list of parts and different models may have parts in common. First we decide how many units of each part will be bought. After the purchase is done the actual demand for the different models will be revealed. Then we decide how many units of each model will be produced. This basically amounts to choose if the demand of each particular model will be under-satisfied, satisfied, or over-satisfied while keeping within the constraints defined by the number of purchased parts.*

There is a penalty for each unit that we fail to sell due to unsatisfied demand and there is a storage cost associated to each excedent unit that is produced. Since the storage cost depends on the current market, it is variable and will be known only after the second decision has been made. It is assumed that all the products will eventually be sold and the storage cost is paid only once.

Let x_i be the number of parts of type i that will be purchased and let y_j be the number of units of model j that will be produced. Also let

$$\begin{array}{ll} r_j := \textit{selling price of product } j & c_i := \textit{cost of part } i \\ d_j := \textit{penalty for uncovered demand} & e_j := \textit{cost of storage of product } j \quad . \\ D^j := \textit{demand of product } j & \end{array}$$

Our goal is to maximize the profit given by the following formula

$$\sum_{\textit{Product } j} r_j y_j - \sum_{\textit{Part } i} c_i x_i - \sum_{\substack{\textit{Product } j \\ \textit{is under} \\ \textit{produced}}} d_j(D_j - y_j) - \sum_{\substack{\textit{Product } j \\ \textit{is over} \\ \textit{produced}}} e_j(y_j - D_j). \tag{75}$$

First we express the production problem as a two stage risk-averse problem (2)-(3). Suppose there are m product models with a total of n different parts. Let $c := (c_1, \dots, c_n)$ and $D := (D^1, \dots, D^m)$. Define the binary matrix $W \in \mathbb{R}^{n\times m}$ such that $W_{ji} = 1$ if and only if part j is used by model i. Given a vector $x = (x_1, \ldots, x_n)$, let $\mathcal{Q}(x, \xi)$ be the optimal value of the second stage problem

$$\begin{aligned} \min_{y,w,v \in \mathbb{R}^m} \quad & \sum_{i=1}^{m} d_i w_i + \sum_{i=1}^{m} e_i v_i - \sum_{i=1}^{m} r_i y_i \\ \text{s.t.} \quad & Wy - x \leq 0,\ y \geq 0 \\ & w \geq D - y,\ w \geq 0 \\ & v \geq y - D,\ v \geq 0 \end{aligned} \qquad . \tag{76}$$

We interpret the function $\mathcal{Q}(x, \xi)$ as a random variable on the vector of storage cost $e = (e_1, \ldots, e_m)$ and the function $c^\top x + \mathbb{E}[\mathcal{Q}(x, \xi)]$ as a random variable on the vector of model demand D. The goal of the production problem is achieved by finding

$$\min_{x \in \mathbb{R}^n_+} \mathbb{E}\left(c^\top x + \mathbb{E}[\mathcal{Q}(x, \xi)]\right).$$

Since we are interested in a risk-averse model of the production problem, we will instead consider

$$\min_{x \in \mathbb{R}^n_+} \rho_1\left(c^\top x + \rho_2[\mathcal{Q}(x, \xi)]\right), \tag{77}$$

where $\rho_i(Z) = \mathbb{E}[Z] + a_i\, \mathbb{E}[Z - \mathbb{E}[Z]]_+$, $a_i \in [0, 1]$.

Assume now that there are $\{1, \ldots, N\}$ possible demand scenarios each occurring with corresponding probability p_j. Moreover, suppose that each demand scenario j there are $\{1, \ldots, N_j\}$ possible storage cost scenarios each occurring with corresponding with probability s_{ji}. In this case a straight forward linear programming formulation of (77) is given by:

$$
\begin{aligned}
\min_{\substack{x,y_j,w_j,v_j\\A_j,B_j,Q_{jt},R_{jt}}} \quad & \sum_{j=1}^{N} p_j A_j + a_1 \sum_{j=1}^{N} p_j B_j \\
\text{s.t.} \quad & A_j \geq c^\top x + C_j \\
& B_j \geq A_j - \sum_{k=1}^{N} p_k A_k, \quad B_j \geq 0 \\
& C_j = \sum_{t=1}^{N_j} s_{jt} Q_{jt} + a_2 \sum_{t=1}^{N_j} s_{jt} R_{jt} \\
& R_{jt} \geq Q_{jt} - \sum_{k=1}^{N_j} s_{jk} Q_{jk}, \quad R_{jt} \geq 0 \\
& Q_{jt} \geq \sum_{i=1}^{m} d_i w_{ji} + \sum_{i=1}^{m} e^t_{ji} v_{ji} - \sum_{i=1}^{m} r_i y_{ji} \\
& W y_j - x \leq 0, \ y_j \geq 0 \\
& w_j \geq D_j - y_j, \ w_j \geq 0 \\
& v_j \geq y_j - D_j, \ v_j \geq 0 \\
& \text{for all } j = 1, \ldots, N \text{ and} \\
& \text{for each } j \text{ the corresponding} \quad t = 1, \ldots, N_j,
\end{aligned} \tag{78}
$$

where c_j, d_j, r_j are as defined before, e^t_{ji} is the storage cost of product i under demand scenario j and storage scenario t. Also $D_j := (D^1_j, \ldots, D^m_j)$ is the vector of product demands under demand scenario j.

The size of the linear programming representation of the production problem shows the importance of developing efficient methods to solve multi stage risk-averse problems. We applied to (77) the cutting plane, the classical bundle, and the truncated bundle method. Whenever possible, we compared the results obtained by these methods with the result of solving directly the linear program (78) with a simplex algorithm. Following this we compared the total running time, total number of iterations, and the average time per iteration of each method. Table 1 shows the comparison of all the methods.

We can see from the results that even in small problems the bundle and truncated bundle methods outperform the cutting plane method. On the other hand for small problems it is much better to solve directly the linear program (78). The usefulness of the bundle and truncated bundle method shows when considering large problems. Here the meager memory requirements of these methods allow us to obtain a solution even when the linear program (78) is too big for our computer memory. In general we saw the truncated bundle method outperforming the classical bundle method but this might be problem specific.

Size	**LP**	**Cut**			**Bundle**			**Truncated**		
S_1 x S_2	Time	Time	Iterations	T/I	Time	Iterations	T/I	Time	Iterations	T/I
6 x 3	0	2083	2905	0.717	106	476	0.223	15	97	0.155
5 x 5	0	508	1417	0.359	95	451	0.211	36	194	0.186
5 x 6	0	660	1621	0.407	75	388	0.193	13	86	0.151
6 x 6	0	-	-	-	134	574	0.233	133	441	0.302
10 x 10	0	-	-	-	313	435	0.720	287	419	0.685
50 x 50	5	-	-	-	1381	510	2.708	1652	485	3.406
100 x 100	98	-	-	-	5570	660	8.439	1547	300	5.157
200 x 200	5767	-	-	-	5975	240	24.896	4722	200	23.610
300 x 300	-	-	-	-	19910	255	78.078	20622	255	80.871

Table 1: Tests were performed with 10 parts, 5 products, S_1 first stage scenarios, and S_2 second stage scenarios.

11 Conclusions

We defined a two stage risk-averse program and used Lagrangian duality to formulate a dual representation of it. Along the way we developed the tools necessary for the development of dual cutting plane methods.

Later we developed a dual cutting plane method and the truncated bundle method. A lot of time was devoted to show the correctness of the truncated bundle method. In doing so we developed a robust theoretical foundation of the method giving to it the possibility of future applications. Lastly we applied our methods to a numerical example and compared their performance.

References

[1] Philippe Artzner, *Application of coherent risk measures to capital requirements in insurance*, N. Am. Actuar. J. **3** (1999), no. 2, 11–25, SOA Seminar: Integrated Approaches to Risk Measurement in the Financial Services Industry (Atlanta, GA, 1997).

[2] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath, *Coherent measures of risk*, Math. Finance **9** (1999), no. 3, 203–228.

[3] ———, *Coherent measures of risk*, Risk management: value at risk and beyond (Cambridge, 1998), Cambridge Univ. Press, Cambridge, 2002, pp. 145–175.

[4] Sungyong Choi and Andrzej Ruszczyński, *A risk-averse newsvendor with law invariant coherent measures of risk*, Oper. Res. Lett. **36** (2008), no. 1, 77–82.

[5] Naomi Miller and Andrzej Ruszczyński, *Risk-adjusted probability measures in portfolio optimization with coherent measures of risk*, European J. Oper. Res. **191** (2008), no. 1, 193–206.

[6] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970.

[7] Andrzej Ruszczynski, *Nonlinear optimization*, Princeton University Press, Princeton, NJ, USA, 2006.

[8] Andrzej Ruszczyński and Alexander Shapiro, *Stochastic programming models*, Stochastic programming, Handbooks Oper. Res. Management Sci., vol. 10, Elsevier, Amsterdam, 2003, pp. 1–64.

[9] Andrzej Ruszczyński and Alexander Shapiro, *Conditional risk mappings*, Math. Oper. Res. **31** (2006), no. 3, 544–561.

[10] ———, *Optimization of convex risk functions*, Math. Oper. Res. **31** (2006), no. 3, 433–452.

[11] ———, *Corrigendum to: "optimization of convex risk functions," mathematics of operations research31 (2006) 433–452*, Math. Oper. Res. **32** (2007), no. 2, 496–496.

[12] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, *Lectures on stochastic programming: Modeling and theory*, MPS-SIAM series on optimization, no. 9, MPS-SIAM, Philadelphia, 2009.