Resource Allocation for Contingency Planning: An Inexact Bundle Method for Stochastic Optimization

Ricardo A. Collado, Somayeh Moazeni*
School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA
rcollado@stevens.edu, smoazeni@stevens.edu

Resource contingency planning aims to mitigate the effects of unexpected disruptions in supply chains. While these failures occur infrequently, they often have disastrous consequences. This paper formulates the resource allocation problem in contingency planning as a two-stage stochastic optimization problem with a risk-averse recourse function. Furthermore, the paper proposes a novel computationally tractable solution approach. The proposed algorithm relies on an inexact bundle method with subgradient approximations through a scenario reduction mechanism. We prove that our scenario reduction and function approximations satisfy the requirements of the oracle in the inexact bundle method, ensuring convergence to an optimal solution. The practical performance of the developed inexact bundle method under risk aversion is investigated for our resource allocation problem. We create a library of test problems and obtain their optimal values by applying the exact bundle method. The computed solutions from the developed inexact bundle method are compared against these optimal values, under different risk measures. Our analysis indicates that our inexact bundle method significantly reduces the computational time of solving the resource allocation problem in comparison to the exact bundle method, and is capable of achieving a high percentage of optimality within a much shorter time.

Key words: Logistics, risk-averse optimization, stochastic programming
History: submitted in October 2017

1. Introduction

This paper studies the optimal allocation of resources to reduce the risk of demand unfulfilment due to demand spikes, supply interruptions, or tie-line disruptions. Such network disruptions can arise due to various man-made and natural disasters, such as severe weather or storms. Resource contingency planning is one of the proactive strategies to mitigate such uncertainties and to be prepared to withstand disruptions in supply chains (Tomlin 2006, Snyder et al. 2006). An optimal allocation of these resources to different areas in a network is critical to achieve lower costs of failure and higher reliability. While the frequency of network disruptions due to disasters can be rare, they can lead to severe supply chain interruptions. Hence, decision makers should take into account such risks when allocating additional resources.

* Corresponding author. Tel.: +1(201)216-8723

For various disaster management strategies, the reader is referred to Gupta et al. (2016). Matta (2016) studies contingency planning by the addition of reserve capacity in the supply chain through network optimization tools. Grass and Fischer (2016) provides a recent literature review on contingency planning in disaster management by two-stage stochastic programming. The existing literature primarily focuses on minimizing the expected failure cost, e.g., see Cui et al. (2010), Alem et al. (2016) and the references therein. Noyan (2012) considers the risk-averse two-stage stochastic optimization model for disaster management and discusses the importance of incorporating a risk measure to derive optimal decisions computed from the Benders-decomposition method. A risk-averse model is studied in Alem et al. (2016) where a heuristic solution approach is proposed. As it is pointed out in Alem et al. (2016), computational challenges are the primary barrier in the risk-averse models for such two-stage logistic problems since the number of decision variables would depend on the number of scenarios, which is potentially large in the presence of down-side risk measures. The present paper aims to address this challenge by proposing a computationally tractable approach for the resource location problems arising in contingency planning.

This resource allocation problem lends itself to the class of two-stage stochastic optimization problems with a risk-averse recourse function. Consider a stochastic optimization problem of the form:

$$\min_{x\in\mathcal{X}} \quad \varphi(x) := c^\top x + \rho\left[\mathcal{Q}(x,\omega)\right], \tag{1}$$

where ρ is a risk measure, and for any $\omega\in\Omega$, $\mathcal{Q}(x,\omega)$ is the optimal value of the second-stage problem

$$\begin{aligned} \mathcal{Q}(x,\omega) := \min_{y\in\mathcal{Y}} \quad & q(y,\omega) \\ \text{s.t.} \quad & t_i(x,\omega) + r_i(y,\omega) \le 0, \quad i=1,\ldots,m. \end{aligned} \tag{2}$$

Here, Ω is the support of the probability distribution of ω. For any $\omega\in\Omega$, the feasible set $\mathcal{Y}\subseteq\mathbb{R}^s$ is convex and compact, and the real valued functions $q(\cdot,\omega)$, $r_i(\cdot,\omega)$, $t_i(\cdot,\omega)$, $i=1,\ldots,m$ are proper convex almost everywhere; consequently, they are continuous. The function $\mathcal{Q}(x,\omega)$ is assumed to be finite for all $x\in\mathcal{X}$ and all $\omega\in\Omega$, which implies that the two-stage risk-averse stochastic problem has complete recourse. The functions $t_i(\cdot,\omega)$, $i=1,\cdots,m$, are assumed to be continuously differentiable. Hence, $T(x,\omega) := (t_1(x,\omega),\cdots,t_m(x,\omega))^\top$ is differentiable in x. We assume that $0\in \operatorname{int}\{T(x,\omega)+\nabla_x T(x,\omega)\mathbb{R}^s - \operatorname{dom}\vartheta(\cdot,\omega)\}$, where $\vartheta(T(x,\omega),\omega) := \mathcal{Q}(x,\omega)$ for all $x\in\mathcal{X}$ and $\omega\in\Omega$. The feasible set $\mathcal{X}\subseteq\mathbb{R}^n$ is nonempty and compact.

In this paper, we focus on coherent risk measures (Artzner et al. 1999). The convexity of the second-stage problem together with the convexity and monotonicity of coherent risk measures ρ imply that φ is a proper convex function, e.g., see Ruszczyński and Shapiro (2003). In addition, $\varphi(\cdot)$ is subdifferentiable over the interior of its domain (Ruszczynski 2006, p. 59-61). However, the function $\varphi(x)$ is nonsmooth in general for nondifferentiable coherent risk measures ρ.

The bundle method is an approach for solving problems of the form (1), which is capable of handling the non-smoothness in the objective function $\phi(x)$. For details see Hiriart-Urruty and Lemarechal (1993), Ruszczynski (2006), Teo et al. (2010). This approach iteratively builds linearizations for $\varphi(x)$ around a projection point and includes a cutting-plane model using the piecewise maximum of linearizations. At iteration k, given the finite set of information $\mathcal{J}_k = \{\hat{x}^j,\ \varphi(\hat{x}^j),\ g^j \in \partial\varphi(\hat{x}^j)\}_{j\in J_k}$ for $J_k \subseteq \{1,\cdots,k\}$, the bundle method constructs a piecewise-linear approximation of φ obtained from finitely many linear constraints

$$\varphi^k(x) := \max_{j\in J_k} \left\{\varphi(\hat{x}^j) + \left\langle g^j,\, x - \hat{x}^j \right\rangle\right\}. \tag{3}$$

The function φ^k is then used to approximate $\varphi(x)$ and compute its gradients.

This process requires evaluations of the objective function $\varphi(\hat{x}^j)$ and consequently computing $\mathcal{Q}(\hat{x}^j,\omega)$ for all $\omega \in \Omega$. This step involves solving $|\Omega|$ problems of the form (2), that can be computationally intensive particularly when the size of the scenario space grows or a large number of decision variables and constraints are present.

While this computational challenge persists when the expected value of the recourse function $\mathbb{E}\left[\mathcal{Q}(x,\omega)\right]$ is considered in $\varphi(x)$, the computational demand increases in the presence of a downside risk measure. Obtaining a reliable estimation of the probability distribution of $\mathcal{Q}(x,\omega)$, and hence an accurate evaluation of $\rho\left[\mathcal{Q}(x,\omega)\right]$, often relies on a large number of scenarios.

To alleviate the computational cost when the first stage objective function includes the expectation of the recourse function $\mathbb{E}\left[\mathcal{Q}(x,\omega)\right]$, a number of extensions to bundle methods capable of working with less accurate function evaluations have been developed, see Oliveira and Sagastizabal (2014). These methods replace the function values and subgradients by their approximations, for all or just a subset of the iterations.

Suppose $\widetilde{\varphi}_{\hat{x}_j}$ and $\widetilde{g}_{\hat{x}_j}$ are estimates obtained from an oracle for the projection point $\hat{x}^j$. Then, at iteration k the inexact bundle method uses the approximate linearization φ^k as follows,

$$\varphi^k(x) = \max_{j\in\mathcal{J}_k} \left\{\widetilde{\varphi}_{\hat{x}_j} + \left\langle \widetilde{g}_{\hat{x}_j},\, x - \hat{x}^j \right\rangle\right\}. \tag{4}$$

When $\widetilde{\varphi}_{\hat{x}_j} = \varphi(\hat{x}^j)$ and $\widetilde{g}_{\hat{x}_j} = g^j$, equation (4) is reduced to that of the exact bundle method.

To achieve convergence in the inexact bundle method, the estimates $\widetilde{\varphi}_x$ and $\widetilde{g}_x$, which are the outputs of an oracle, should satisfy some conditions. At each iteration for a given point x, the inexact bundle method (Kiwiel 2006, Oliveira et al. 2011) requires a function estimate $\widetilde{\varphi}_x$ and a subgradient estimate $\widetilde{g}_x$ satisfying

$$\widetilde{\varphi}_x \in [\varphi(x) - \epsilon_1,\ \varphi(x) + \epsilon_2], \tag{5}$$

$$\widetilde{g}_x \in \partial_{\epsilon_0}\varphi(x). \tag{6}$$

Here, $\epsilon_1, \epsilon_2 \geq 0$ are unknown but bounded, $\epsilon_0 = \epsilon_1 + \epsilon_2$, and the ϵ_0-approximate subdifferential $\partial_{\epsilon_0}\varphi(x)$ in (6) is given by

$$\partial_{\epsilon_0}\varphi(x) := \{g \in \mathbb{R}^n \mid \varphi(z) \geq \varphi(x) + \langle g, z - x\rangle - \epsilon_0, \forall z \in \mathcal{X}\}. \tag{7}$$

Therefore, to achieve convergence in this risk-neutral inexact bundle method, one needs to compute approximations $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ which satisfy equations (5) and (6). In addition, to address the original motivation of achieving a computationally efficient approach, these approximations must be easily computable.

This paper extends this inexact bundle method for problems of the form (1) when the aversion to risk in the second-stage optimal value $\rho[\mathcal{Q}(x,\omega)]$ appears in the objective function of the first stage problem. We achieve this by describing appropriate oracles capable of generating estimates $\widetilde{\varphi}_{x^j}$ and $\widetilde{g}_{x^j}$, which guarantee convergence in the developed inexact bundle method. A scenario reduction method is also proposed to further mitigate the computational cost of evaluating (the tail of) the distribution of $\mathcal{Q}(x,\omega)$ and consequently $\rho[\mathcal{Q}(x,\omega)]$. The convergence of the approach to a risk-averse optimal solution is established.

The developed inexact bundle method for the resource allocation model is studied using an extensive computational investigation. In particular, we focus on two coherent risk measures and compare their outcomes with those for the case of minimizing the resource costs and the risk of unfulfilled demand.

The contributions of the paper can be summarized as follows.

- An oracle needed to implement the framework of inexact bundle method is introduced for a class of risk-averse two-stage stochastic optimization problems.
- We prove that the objective function and subgradient approximations from this oracle meet the requirements stated in (5)-(6).
- The inexact approach with the introduced risk-averse oracle is applied to a resource allocation problem arising in contingency planning.
- We perform the benchmarking of the algorithm against the exact bundle method for problem instances with different sizes, and demonstrate the computational benefits of the developed approach.

This paper is organized as follows. Section 2 provides background on risk-averse two-stage optimization. Section 3 explains the bundle method for the risk-averse optimization problem, where the details on evaluating the objective function in (1) and its subgradient are also discussed. Section 4 presents an overview of the inexact bundle methods. Section 5 introduces the risk-averse oracle and proves its correctness. The modeling details for a resource allocation problem in contingency planning are explained in Section 6. Section 7 reports the results of the numerical experiments and the comparisons on the benchmark problems. We list our conclusions in Section 8.

2. Coherent Risk Measures and Risk-Averse Stochastic Optimization

Risk-averse bundle methods reduce the risk-averse problem to iteratively constructing a family of risk-neutral approximations. This section details the evaluation of the first-stage objective function with risk-averse recourse and the computation of its subgradients. First, we briefly discuss coherent risk measures and its representation theorem. For an in-depth treatment see Ruszczyński and Shapiro (2006b, 2007, 2006a), Shapiro et al. (2014).

Let $(\Omega, \mathcal{F}, P)$ be a probability space with sigma-algebra $\mathcal{F}$ and probability measure P. Let $p \in [1, +\infty)$ and $q \in (1, +\infty]$ be such that $1/p + 1/q = 1$. Define $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$ to be a pair of conjugate dual spaces with the scalar product

$$\langle \mu, Z \rangle := \int_\Omega \mu(\omega) Z(\omega) dP(\omega).$$

Each element $Z := Z(\omega)$ of $\mathcal{Z}$ is viewed as an uncertain outcome on $(\Omega, \mathcal{F})$ and is by definition a random variable whose p-th order moment is finite.

For $Z, Z' \in \mathcal{Z}$, let $Z \preceq Z'$ denote the pointwise partial order, i.e., $Z(\omega) \leq Z'(\omega)$ for all $\omega \in \Omega$. In our model Z represents a random cost and as such smaller realizations are preferred.

Definition 2.1 *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*:* $\rho(\alpha Z + (1-\alpha) Z') \leq \alpha \rho(Z) + (1-\alpha)\rho(Z')$*, for all* $Z, Z' \in \mathcal{Z}$ *and all* $\alpha \in [0, 1]$.

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

(**A3**) Translation Equivariance*: If* $\alpha \in \mathbb{R}$ *and* $Z \in \mathcal{Z}$*, then* $\rho(Z + \alpha) = \rho(Z) + \alpha$.

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

The following theorem is a fundamental result employed in the evaluation of coherent measures and risk-averse stochastic optimization (Shapiro et al. 2014):

Theorem 2.1 (Representation Theorem of Coherent Risk Measures) *Let* $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ *be a lower semicontinuous coherent risk measure. Then the function* ρ *is subdifferentiable at* 0 *and*

$$\rho(Z) = \sup_{\mu \in \partial\rho(0)} \langle \mu, Z \rangle, \tag{8}$$

where $\partial\rho(0) \subseteq \left\{ \mu \in \mathcal{Z}^* \,\middle|\, \mu \succeq 0 \text{ and } \int_\Omega \mu dP = 1 \right\}$. *When* Ω *is finite with* N *elements and the probability measure* $P = (p_1, \ldots, p_N)$*, equation* (8) *takes the form*

$$\rho(Z) = \max_{\mu \in \partial\rho(0)} \sum_{i=1}^{N} \mu_i z_i p_i := \max_{\mu \in \partial\rho(0)} \mathbb{E}_\mu[Z], \tag{9}$$

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

Theorem 2.1 implies that problem (1) can be reformulated as

$$\min_{x\in\mathcal{X}} \; c^\top x+\rho(\mathcal{Q}(x,\omega)) = \min_{x\in\mathcal{X}} \; \left\{ c^\top x + \sup_{\mu\in\partial\rho(0)} \langle \mu, Q(x)\rangle \right\} = \min_{x\in\mathcal{X}} \sup_{\mu\in\partial\rho(0)} \; c^\top x + \langle \mu, Q(x)\rangle \;,$$

where $Q(x) := \mathcal{Q}(x,\cdot)$ denotes the random variable on Ω described by problem (2). For $x\in\mathcal{X}$ and $\mu\in\mathcal{Z}^*$, define

$$\phi(x) := \sup_{\mu\in\partial\rho(0)} \langle \mu, Q(x)\rangle \,. \tag{10}$$

For the finite sample space Ω, we have $\langle \mu, Q(x)\rangle = \sum_{\omega\in\Omega} \mathcal{Q}(x,\omega)\mu_\omega P_\omega$, where P_ω is the probability of sample $\omega\in\Omega$. Consequently

$$\phi(x) = \sup_{\mu\in\partial\rho(0)} \sum_{\omega\in\Omega} \mathcal{Q}(x,\omega)\mu_\omega P_\omega. \tag{11}$$

Each $\mathcal{Q}(x,\omega)$ is computed by solving an optimization problem of type (2). Given $x\in\mathcal{X}$ and $\omega\in\Omega$, let y_ω be an optimal solution of problem (2), i.e., $\mathcal{Q}(x,\omega) = q(y_\omega,\omega)$. Using this equality in equation (11) yields

$$\phi(x) = \sup_{\mu\in\partial\rho(0)} \sum_{\omega\in\Omega} q(y_\omega,\omega)\mu_\omega P_\omega. \tag{12}$$

Problem (11) is a convex optimization problem and for common coherent risk measures ρ this problem can be reformulated as a linear optimization problem. Subsequently, let μ^* denote an optimal solution for the supremum problem in (12).

The rest of this section aims to specify the subdifferential $\partial\phi(x)$. Let $\widetilde{\mathcal{Q}} : \mathbb{R}^n \times \mathcal{Z}^* \to \overline{\mathbb{R}}$ where $\widetilde{\mathcal{Q}}(x,\mu) := \langle \mu, Q(x)\rangle$. The subdifferential of ϕ (e.g., see Theorem 2.87 in Ruszczynski (2006)) is given by

$$\partial\phi(x) = \operatorname{conv}\left[\bigcup_{\mu\in\widehat{\partial\rho}(0)[x]} \partial_x \widetilde{\mathcal{Q}}(x,\mu) \right], \tag{13}$$

where $\widehat{\partial\rho}(0)[x] := \{\mu\in\partial\rho(0) \mid \widetilde{\mathcal{Q}}(x,\mu) = \phi(x)\}$ and $\partial_x\widetilde{\mathcal{Q}}(x,\mu)$ is the subdifferential of $\widetilde{\mathcal{Q}}(\cdot,\mu)$ evaluated at x. The subdifferential set $\partial\rho(0)$ is convex, and for most popular coherent risk measures $\partial\rho(0)$ is a compact set, e.g., see Shapiro et al. (2014). The compactness of $\partial\rho(0)$ implies that $\widehat{\partial\rho}(0)[x]$ is nonempty and compact, for every $x\in\mathcal{X}$.

It follows from optimality of μ^* in (12) that $\mu^*\in\widehat{\partial\rho}(0)[x]$. Therefore, from equation (13)

$$\partial_x\widetilde{\mathcal{Q}}(x,\mu^*) \subseteq \partial\phi(x). \tag{14}$$

Next, we further characterize $\partial_x\widetilde{\mathcal{Q}}(x,\mu^*)$. For the finite set Ω, it follows from the Moureau-Rockafeller Theorem (see Theorem 6 in Ruszczyński and Shapiro (2003)) that

$$\partial_x\widetilde{\mathcal{Q}}(x,\mu) = \sum_{\omega\in\Omega} \mu_\omega P_\omega \partial_x \mathcal{Q}(x,\omega), \tag{15}$$

where $\partial_x \mathcal{Q}(x,\omega)$ is the subdifferential of $\mathcal{Q}(\cdot,\omega)$ evaluated at x. Denote $\chi_i = t_i(x,\omega)$ and $\chi = T(x,\omega)$. Then the problem (2) achieves strong primal-dual optimality (see Proposition 25 in Ruszczyński and Shapiro (2003)) and is given by

$$\mathcal{Q}(x,\omega) = \max_{\pi \geq 0} \left\{ \pi^\top \chi + \inf_{y \in \mathcal{Y}} \mathcal{L}(y,\pi,\omega) \right\}, \tag{16}$$

where $\mathcal{L}(y,\pi,\omega)$ is the Lagrangian of problem (2), i.e., $\mathcal{L}(y,\pi,\omega) := q(y,\omega) + \sum_{i=1}^m \pi_i r_i(y,\omega)$.

Let $D(\chi,\omega)$ denote the set of optimal solutions of problem (16) with the optimal value $\vartheta(\chi,\omega)$, i.e., $\mathcal{Q}(x,\omega) = \vartheta(\chi,\omega)$. It follows from the convexity of problem (2), the differentiability property of $t_i(\cdot,\omega)$s, the assumption $0 \in \operatorname{int}\{T(x,\omega) + \nabla_x T(x,\omega)\mathbb{R}^s - \operatorname{dom}\vartheta(\cdot,\omega)\}$, and Proposition 26 in (Ruszczyński and Shapiro 2003) that

$$\partial_x \mathcal{Q}(x,\omega) = \nabla_x T(x,\omega)^\top D(\chi,\omega). \tag{17}$$

This result along with equation (15) imply that

$$\partial_x \widetilde{\mathcal{Q}}(x,\mu) = \sum_{\omega \in \Omega} \mu_\omega P_\omega \nabla_x T(x,\omega)^\top D(\chi,\omega). \tag{18}$$

For a given $x \in \mathcal{X}$ and $\omega \in \Omega$, let π_ω be a dual optimal solution corresponding to y_ω of problem (2). Then equation (18) at μ^* together with equation (14) yields

$$\zeta_x = \sum_{\omega \in \Omega} \mu^*_\omega P_\omega \nabla_x T(x,\omega)^\top \pi_\omega \in \partial\phi(x). \tag{19}$$

Consider the risk-averse two-stage stochastic programming problem (1)–(2) with the objective function $\varphi(x)$. Let $X \subset \mathbb{R}^n$ be convex and compact. By Theorem 2.1, the function ρ is subdifferentiable at zero and satisfies equation (8).

Remark 2.1 *This paper considers finite sample spaces Ω potentially of very large cardinality. While the methods presented here can be applied to problems with infinite sample space using numerical integration and approximation, this is not within the scope of the present paper.*

3. Bundle Method

For risk-neutral multistage stochastic optimization problems, the family of decomposition methods constitutes an established and efficient approach (see Birge and Louveaux (1997), Kall and Mayer (2005), Prékopa (1995), Ruszczyński (2003) and the references therein). However, it cannot be directly applied to problem (1) where the risk aversion to the recourse objective function is present. With coherent risk measures, the main feature facilitating decomposition, the integral form of the objective function, is absent. The class of cutting plane methods, in particular bundle methods, proved to be a useful approach to solve risk-averse optimization problems. This approach has been

successfully applied to multistage risk-averse optimization problems (Miller and Ruszczyński 2008, Choi and Ruszczyński 2008, Miller and Ruszczyski 2011, Collado et al. 2012).

The essence of the bundle method includes the application of Moreau-Yosida regularization with respect to the first-stage decision variable $x \in \mathcal{X}$ and solving a sequence of quadratic optimization problems. Localizing the iterations through regularization improves linear approximations, and makes the bundle method more reliable for problems of higher dimension, where simpler methods, such as the cutting plane method, become too slow to reach convergence in a reasonable time.

We apply the bundle method directly to the first-stage problem (1). Consider

$$\min_{x \in \mathcal{X}} \varphi(x).$$

At each iteration of the bundle method, given $\mathcal{J}_k$, the piecewise-linear approximation of φ as in equation (3) denoted by φ^k is constructed. For a regularization parameter $\gamma > 0$ and the *stability center* $\beta^k \in \mathcal{X}$, the following *master problem* is solved,

$$\min_{x \in \mathcal{X}} \ \varphi^k(x) + \frac{\gamma}{2} \left\| x - \beta^k \right\|^2. \tag{20}$$

Given the definition of φ^k in equation (3), problem (20) is expressed as

$$\begin{aligned} \min_{x \in \mathcal{X},\ v \in \mathbb{R}} \quad & v + \frac{\gamma}{2} \left\| x - \beta^k \right\|^2 \\ \text{s.t.} \quad & \varphi(\hat{x}^j) + \left\langle g^j,\, x - \hat{x}^j \right\rangle \leq v, \qquad \forall\ j \in J_k. \end{aligned} \tag{21}$$

Iterations in the bundle method require to evaluate $\varphi(x)$ and obtain subgradients $g \in \partial\varphi(x)$, for any $x \in \mathcal{X}$. These tasks are carried out by means of equations (12) and (19). The details of the steps in the (exact) bundle method are presented in Algorithm 1. See section 7.4 in (Ruszczynski 2006) for an in-depth discussion on the details and convergence of the bundle method for a generic nondifferentiable optimization problem.

Algorithm 1 *Bundle Method*

Inputs: regularization parameter $\gamma > 0$, tolerance level $\delta > 0$.

Step 0: *(Initialization).*

[*i*] *Set $k := 1$, $J_0 := \emptyset$, $v^1 := -\infty$.*

[*ii*] *Let $\hat{x}^1 \in \mathcal{X}$ be a given initial feasible point and set $\beta^0 := \hat{x}^1$.*

Step 1: *For all $\omega \in \Omega$, compute $\mathcal{Q}(\hat{x}^k, \omega)$ by solving problem (2). Let (y_ω, π_ω) be its optimal primal-dual pair.*

Step 2: *Compute $\phi(\hat{x}^k)$ using equation (12). Let μ^* be an optimal solution for problem (12).*

Step 3: *Compute:*

[*i*] $\varphi(\hat{x}^k) = c^\top \hat{x}^k + \rho\left[\mathcal{Q}(\hat{x}^k, \omega)\right] = c^\top \hat{x}^k + \phi(\hat{x}^k)$.

[*ii*] $g^k := c + \zeta_x$, *where* ζ_x *is given in equation* (19). *Notice that* $g^k \in \partial\varphi(\hat{x}^k)$.

Step 4: *If* $\varphi(\hat{x}^k) > v^k$, *then* $J_k := J_{k-1} \cup \{k\}$. *Otherwise,* $J_k := J_{k-1}$.

Step 5: *If* $k = 1$ *or if* $k \geq 2$ *and* $\varphi(\hat{x}^k) \leq (1-\gamma)\varphi(\beta^{k-1}) + \gamma\varphi^{k-1}(\hat{x}^k)$, *then* $\beta^k := \hat{x}^k$. *Otherwise,* $\beta^k := \beta^{k-1}$.

Step 6: *Solve problem* (21). *Let* $(\hat{x}^{k+1}, v^{k+1})$ *be an optimal solution.*

Step 7: *If* $\varphi(\beta^k) - v^{k+1} < \delta\,(1 + v^{k+1})$, *then stop and return* $\hat{x}^{k+1}$ *as a solution with tolerance* δ. *Otherwise, set* $k := k+1$ *and continue to* Step 1.

Notice that here we forgo the typical bundle "pruning" of $\mathcal{J}_k$ based on Lagrangian multipliers. This guarantees optimality but has the disadvantage of increasing the size of the master problem at every iteration.

4. Inexact Bundle Method

Computing $\varphi(\hat{x}_k)$ and g^k in Step 3 of the bundle method involves solving the second-stage problem for all $\omega \in \Omega$. This is a computationally expensive task. To mitigate this computational challenge, inexact bundle methods, which rely on only approximations of these values, have been proposed. Here, we follow the inexact bundle method appearing in Kiwiel (2006), Oliveira et al. (2011).

Suppose, at a point x, approximations $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ as in equations (5) and (6) are available. At iteration k, problem (20) with the regularization parameter $\frac{1}{t_k}$ is solved, where here $\varphi^k(x)$ is given by (4). Therefore, this problem can be written as

$$\begin{aligned} \min_{x \in \mathcal{X},\ v \in \mathbb{R}}\ & v + \frac{1}{2t_k}\left\|x - \beta^k\right\|^2 \\ \text{s.t. } & \widetilde{\varphi}_{\hat{x}^j} + \left\langle \widetilde{g}_{\hat{x}^j},\, x - \hat{x}^j \right\rangle \leq v, \qquad \forall\ j \in J_k. \end{aligned} \tag{22}$$

The parameter $t_k > 0$, referred to as the *stepsize*, controls the quadratic penalty $\left\|x - \beta^k\right\|^2$ and is adjusted during iterations (Kiwiel 1990, 2006). Let $\hat{x}_{k+1}$ be an optimal solution to this problem. It will be a candidate point to be included to achieve an improved linear approximation. This method is described in Algorithm 2.

Algorithm 2 *Inexact Bundle Method.*

Inputs: descent parameter $\kappa \in (0,1)$, *stepsize bound* $T_1 > 0$, *stepsize* $t_1 \in (0, T_1]$, *tolerance level* $\delta > 0$.

Step 0: *(Initialization).*

[*i*] *Set* $k := k(0) := 1$ *and* $\ell := 0$. *Here* $k(\ell) - 1$ *denotes the iteration of the* ℓ*th descent step.*

[*ii*] *Let* $\hat{x}^1 \in \mathcal{X}$ *be a given initial feasible point with inexact oracle approximations* $\widetilde{\varphi}_{\hat{x}^1}$ *and* $\widetilde{g}_{\hat{x}^1}$.

[*iii*] *Set* $\beta^1 := \hat{x}^1$, $J_1 := \{1\}$, *and* $i_1 := 0$.

Step 1: *(Trial point finding). Let* $(\hat{x}^{k+1}, v^{k+1})$ *be an optimal solution of problem* (22), *and* $\{\lambda_j^k\}_{j \in J_k}$ *be the Lagrangian multipliers.*

Compute $p^k := \dfrac{1}{t_k}(\beta^k - \hat{x}^{k+1})$, *the nonnegative predicted descent* $w^k := \widetilde{\varphi}_{\hat{x}^k} - v^{k+1} > 0$, *and the aggregate linearization error* $\alpha_k := w^k - t_k\|p^k\|^2$.

Step 2: *(Stopping criterion). Compute optimality measure* $W_k := \max\{\|p^k\|,\ \alpha_k\}$. *If* $W_k \le \delta$, *stop.*

Step 3: *(Noise attenuation & stepsize correction). If* $w^k < -\alpha_k$, *set* $t_k := 10\,t_k$, $T_k := \max\{T_k, t_k\}$, $i_k := k$ *and loop back to* Step 1*; else set* $T_{k+1} := T_k$.

Step 4: *(Descent test). Obtain* $\widetilde{\varphi}_{\hat{x}^{k+1}}$ *and* $\widetilde{g}_{\hat{x}^{k+1}}$ *from the inexact oracle satisfying* (5) *and* (6). *If*

$$\widetilde{\varphi}_{\hat{x}^{k+1}} \le \widetilde{\varphi}_{\beta^k} - \kappa w^k,$$

declare a descent step, *set* $\beta^{k+1} := \hat{x}^{k+1}$, $\ell := \ell + 1$, $i_{k+1} := 0$, $k(\ell) := k+1$.
Otherwise (if $\widetilde{\varphi}_{\hat{x}^{k+1}} > \widetilde{\varphi}_{\beta^k} - \kappa w^k$*), declare a* null step, *and set* $\beta^{k+1} := \beta^k$ *and* $i_{k+1} = i_k$.

Step 5: *(Bundle management). Choose* $J_{k+1} \supseteq \left\{j \in J_k \,\middle|\, \lambda_j^k \neq 0\right\} \cup \{k+1\}$.

Step 6: *(Stepsize updating & looping). If* $k(\ell) = k+1$ *(i.e., after a descent step), select* $t_{k+1} \in [t_k, T_{k+1}]$. *Otherwise declare a null step and set* $t_{k+1} := t_k$. *If* $i_{k+1} = 0$ *and*

$$W_k \le \widetilde{\varphi}_{\beta^k} - \left(\widetilde{\varphi}_{\hat{x}^{k+1}} + \left\langle \widetilde{g}_{\hat{x}^{k+1}},\, \beta^k - \hat{x}^{k+1} \right\rangle\right),$$

choose $t_{k+1} \in [0.1\,t_k, t_k]$. *Set* $k := k+1$ *and go to* Step 1.

Due to errors in the approximations obtained from the oracle, the model might not approximate φ from below. If this is the case then we increase the stepsize t_k and loop over a *noise attenuation* step in Step 3 until corrected. In the descent test in Step 4, a null step improves the model approximation (4) through the addition of an extra constraint.

The method uses the Lagrange multipliers λ_j^k in Step 5 to reduce the number of cuts used to construct the model approximation. Namely, those cuts j corresponding to inactive Lagrange multipliers $\lambda_j^k = 0$ that do not contribute to the new trial point x^{k+1} can be eliminated in equation (4).

We refer the reader to Kiwiel (2006) for an in-depth discussion on the convergence of the inexact bundle method and some of its applications. In order to guarantee convergence of Algorithm 2, it is sufficient to require that $\mathcal{X} \neq \emptyset$ is closed convex, φ is finite convex on a neighborhood of $\mathcal{X}$, and estimates $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ satisfy (5) and (6). With the tolerance level $\delta = 0$, the inexact bundle method has the following two possible outcomes, see Theorem 9 in Oliveira et al. (2011):

I. The method loops forever at noise attenuation (Step 3), in which case the last generated β^k is $2(\epsilon_1 + \epsilon_2)$-optimal.
II. The method generates and infinite sequence of either descent or null steps. In this case due to the compactness of $\mathcal{X}$, the method generates a sequence $\{\beta^k\}_{k=1}^{\infty}$ for which each cluster point β^* satisfies $\beta^* \in \mathcal{X}$ and β^* is $2(\epsilon_1 + \epsilon_2)$-optimal.

Lagrangian multipliers λ_j^k are only necessary for bundle management (Step 5). For the feasible set $\mathcal{X} = \{x \in X \mid Ax \le b,\ x \ge 0,\ x \text{ is integral}\}$ with integrality constraints we forgo Step 5. As with the bundle method, this has the effect of increasing the size of the master problem with each iteration.

5. Defining a Risk-Averse Inexact Oracle

A key component in Algorithm 2 is the definition of an inexact oracle capable of providing estimates $\widetilde{\varphi}_x,\ \widetilde{g}_x$ satisfying (5) and (6). In this section, we define an inexact oracle specialized to work on our two-stage risk-averse stochastic optimization problem (1)–(2). We refer to this oracle as the *inexact risk-averse oracle* which together with Algorithm 2 comprises the *risk-averse inexact bundle method*. The construction of our risk-averse inexact oracle is motivated by the approach in (Oliveira et al. 2011) for the risk-neutral case and linear two-stage models.

The analysis in this section relies on the assumption that for every $x \in \mathcal{X}$, the optimal value of the second-stage problem takes the form

$$\mathcal{Q}(x,\omega) = \max_{\pi} \left\{ \pi^\top \eta(x,\omega) \mid \pi \in \Pi(\omega) \right\}, \tag{23}$$

for some differentiable function $\eta : \mathcal{X} \times \Omega \to \mathbb{R}^s$ and a nonempty convex set $\Pi(\omega) \subseteq \mathbb{R}^s$. The function $\mathcal{Q}(x,\omega)$ attains the structure in (23), when the optimal value $\inf_{y\in\mathcal{Y}} \mathcal{L}(y,\pi,\omega)$ in (16) is linear in π. The feasible region $\Pi(\omega)$ can be expressed in the general form $\{\pi \in \mathbb{R}^s \mid c_i(\pi,\omega) \le 0,\ i = 1,\cdots,l\}$. Denote an optimal solution of the maximization problem (23) by π_ω. Given the optimality of π_ω and expression (23), we have $\mathcal{Q}(x,\omega) = \pi_\omega^\top \eta(x,\omega)$.

The main idea to estimate $\{\pi_\omega\}_{\omega\in\Omega}$, without computing π_ω for every single $\omega \in \Omega$, is to select a subset $\mathcal{I} \subseteq \Omega$, and then assign to every scenario $\omega \in \Omega \setminus \mathcal{I}$ an estimate for π_ω based on an element of $\mathcal{I}$ in the same cluster.

Suppose $\mathcal{I}$ is a subset of scenarios $\mathcal{I} \subseteq \Omega$ obtained through a clustering procedure for the scenario set Ω. Therefore, $\Omega = \cup_{\omega\in\mathcal{I}} \mathcal{J}_\omega$, where $\mathcal{J}_\omega$ denotes a cluster around ω. Given $\mathcal{I}$ and for any $\omega \in \mathcal{I}$, define

$$\widehat{\Pi}_\omega := \{\pi \in \mathbb{R}^s \mid \mathbb{E}_\xi \left[c_i(\pi,\xi) | \xi \in \mathcal{J}_\omega\right] \le 0,\ i = 1,\cdots,l\}. \tag{24}$$

For the finite scenario space, we have

$$\mathbb{E}_\xi \left[c_i(\pi,\xi) | \xi \in \mathcal{J}_\omega\right] = \frac{\sum_{\xi\in\mathcal{J}_\omega} P_\xi c_i(\pi,\xi)}{\sum_{\xi\in\mathcal{J}_\omega} P_\xi}.$$

Consider the optimization problem corresponding to the feasible set $\widehat{\Pi}_\omega$, defined in (24), and denote its optimal solution by $\widetilde{\pi}_\omega$:

$$\widetilde{\pi}_\omega \in \arg\max_{\pi} \left\{ \pi^\top \eta(x,\omega) \mid \pi \in \widehat{\Pi}_\omega \right\}. \tag{25}$$

The approximate oracle then computes the exact optimal values $\mathcal{Q}(x,\omega)$ for every $\omega \in \mathcal{I}$ and derives approximate optimal values for the remaining scenarios $\omega \in \Omega \setminus \mathcal{I}$. This is carried out as follows. For each element in $\mathcal{I}$, we compute $\mathcal{Q}(x,\omega)$, by first obtaining an exact solution π_ω,

and thus $\mathcal{Q}(x,\omega) = \pi_\omega^\top \eta(x,\omega)$. Then, for each remaining scenario $\omega \in \Omega \setminus \mathcal{I}$, we obtain, without solving the second-stage optimization problem, an approximation to $\mathcal{Q}(x,\omega)$. More precisely, the corresponding cluster $\mathcal{J}_\omega$ is identified and an appropriate $\widetilde{\pi}_\omega$ as in (25) is computed. Hence, the algorithm adopts $\mathcal{Q}^{approx}(x,\omega) = \widetilde{\pi}_\omega^\top \eta(x,\omega)$ as an approximation of $\mathcal{Q}(x,\omega)$. These computed values are then aggregated to form the objective function approximation $\widetilde{\varphi}_x$,

$$\widetilde{\varphi}_x := c^\top x + \sup_{\mu \in \partial\rho(0)} \left[\sum_{\omega \in \mathcal{I}} \mu_\omega P_\omega \widetilde{\pi}_\omega^\top \eta(x,\omega) + \sum_{\psi \in \Omega \setminus \mathcal{I}} \mu_\psi P_\psi \widetilde{\varpi}_\psi^\top \eta(x,\psi) \right]. \tag{26}$$

and the approximate subgradient $\widetilde{g}_x$,

$$\widetilde{g}_x := c + \left[\sum_{\omega \in \mathcal{I}} \mu_\omega^* P_\omega \nabla_x \eta(x,\omega)^\top \widetilde{\pi}_\omega + \sum_{\psi \in \Omega \setminus \mathcal{I}} \mu_\psi^* P_\psi \nabla_x \eta(x,\psi)^\top \widetilde{\varpi}_\psi \right], \tag{27}$$

where, $\mu^* := (\mu_\omega^*)_{\omega \in \Omega}$ is an optimal solution to the supremum problem in (26). In both equations (26) and (27), $\widetilde{\varpi}_\psi = \widetilde{\pi}_{\bar{\omega}}$, where $\bar{\omega}$ is associated to the cluster $\mathcal{J}_{\bar{\omega}}$ that contains $\psi \in \Omega \setminus \mathcal{I}$. Later in Theorem 5.1, we prove that under some conditions on structure (23) and $\mathcal{I}$, the computed function and gradient approximations satisfy conditions (5) and (6).

We select the subset $\mathcal{I}$ such that the corresponding set of vectors $\{\eta(x,\psi) \,|\, \psi \in \mathcal{I}\}$ sufficiently deviates from collinearity. We measure collinearity of two scenarios ω and ψ by the cosine of the angle $\theta_{\omega,\psi}$ between the two vectors $\eta(x,\omega)$ and $\eta(x,\psi)$, namely

$$\theta_{\omega,\psi} := \cos^{-1}\left(\frac{\eta(x,\omega)^\top \eta(x,\psi)}{\|\eta(x,\omega)\| \, \|\eta(x,\psi)\|} \right). \tag{28}$$

Hence, two scenarios ω and ψ are collinear if $\cos\theta_{\omega,\psi} = 1$. For any given $x \in \mathcal{X}$ and a given *collinearity parameter* $\epsilon_{\cos} \in (0,1)$, we consider a maximal subset $\mathcal{I} \subseteq \Omega$ such that for every $\omega, \psi \in \mathcal{I}$ we have $\cos\theta_{\omega,\psi} \leq 1 - \epsilon_{\cos}$. Algorithm 3 formally states the risk-averse inexact oracle.

Algorithm 3 *Risk-Averse Inexact Oracle*

Inputs: collinearity parameter $\epsilon_{cos} \in (0,1)$ *and* $x \in \mathcal{X}$.

Step 0: *(Initialization). Select a maximal subset* $\mathcal{I} \subseteq \Omega$ *such that* $\cos\theta_{\omega,\psi} \leq (1 - \epsilon_{cos})$, *for every* $\omega, \psi \in \mathcal{I}$.

Step 1: *(Collinearity clustering). For each* $\omega \in \mathcal{I}$, *set* $\mathcal{J}_\omega := \{\omega\} \cup \{\psi \notin \mathcal{I} \mid \cos\theta_{\omega,\psi} > 1 - \epsilon_{cos}\}$.

Step 2: *(Estimates in* $\mathcal{I}$*). For each* $\omega \in \mathcal{I}$, *specify* $\widehat{\Pi}_\omega$ *in* (24) *and obtain an optimal solution* $\widetilde{\pi}_\omega$ *as in* (25).

Step 3: *(Estimates in* $\Omega \setminus \mathcal{I}$*). For each* $\psi \in \Omega \setminus \mathcal{I}$, *let* $\mathcal{J}_{\bar{\omega}}$ *be its containing set in the collinearity partition obtained in* Step 1, *i.e.* $\psi \in \mathcal{J}_{\bar{\omega}}$. *Then set* $\widetilde{\varpi}_\psi := \widetilde{\pi}_{\bar{\omega}}$.

Step 4: *(Risk-averse oracle estimates). Use the approximate solutions obtained in* Step 3 *to compute* $\widetilde{\varphi}_x$ *using equation* (26). *Compute an optimal solution to the optimization problem in equation* (26) *and denote by* $\mu^* := (\mu_\omega^*)_{\omega \in \Omega}$. *Derive the subgradient approximation* $\widetilde{g}_x$ *by equation* (27).

For each $\omega \in \mathcal{I}$, we first construct the set of all ω-almost collinear scenario set $\mathcal{J}_\omega \subseteq \Omega$. This task is carried out in Step 1. Note that maximality of $\mathcal{I}$ implies that $\cup_{\omega\in\mathcal{I}}\mathcal{J}_\omega = \Omega$. We then in Step 2 compute a $\widetilde{\pi}_\omega \in \mathbb{R}^s$ by solving version of problem (23) with the feasible set $\widehat{\Pi}(\omega)$, in which q_ω by the corresponding average over the set $\mathcal{J}_\omega$. Step 3 makes the approximation that $\pi_\psi \approx \widetilde{\pi}_\omega$ for all $\psi \in \mathcal{J}_\omega$. Note that to clearly express that it is just an approximation we use the notation $\widetilde{\varpi}_\psi$ to refer to $\widetilde{\pi}_\omega$ when assigned to the scenario ψ.

Each call to Algorithm 3 requires the calculation of $\mathcal{I} \subseteq \Omega$. We carry out this step via a combinatorial method given in Oliveira et al. (2011). This method is presented in Algorithm 4. This algorithm completes Steps 0 and 1 in Algorithm 3.

Algorithm 4 *Selection of* $\mathcal{I} \subseteq \Omega$ *and* $\{\mathcal{J}_\omega\}_{\omega\in\Omega}$

Inputs: $\Omega = \{\omega_1, \ldots, \omega_N\}$, x, *and* $\epsilon_{cos} \in (0, 1)$.

Step 0: *(Initialization). Set* $\mathcal{I} := \Omega$.

Step 1: *(Main procedure).*

for every $i \in \{1, \cdots, |\Omega|\}$ such that $\omega_i \in \mathcal{I}$: **do**
  define $\mathcal{J}_{\omega_i} := \{\omega_i\}$
  for every $j \in \{i+1, \cdots, |\Omega|\}$ such that $\omega_j \in \mathcal{I}$ **do**
    compute $\cos\theta_{\omega_i,\omega_j}$ using equation (28).
    If $\cos\theta_{\omega_i,\omega_j} > 1 - \epsilon_{\cos}$ then $\mathcal{I} := \mathcal{I} \setminus \{\omega_j\}$ and $\mathcal{J}_{\omega_i} := \mathcal{J}_{\omega_i} \cup \{\omega_j\}$.
  end for
end for

The generated subset $\mathcal{I} \subseteq \Omega$ from Algorithm 4 depends on the permutation of elements of Ω fixed in Step 0. In particular, the initial elements of the permutation of Ω are favored to be part of $\mathcal{I}$. We consider a random permutation of elements of Ω on each call to Algorithm 4.

5.1. Proof of Results

In this section, we prove that the risk-averse inexact oracle (Algorithm 3) offers $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ satisfying (5) and (6); hence ensuring optimality of the risk-averse inexact bundle method. This property is referred to as the correctness of the risk-averse inexact oracle.

For a given convex set $\mathcal{S}$, denote the support function of $\mathcal{S}$ evaluated at d by $s_{\mathcal{S}}(d)$, i.e.,

$$s_{\mathcal{S}}(d) := \max\left\{\pi^\top d \;\middle|\; \pi \in \mathcal{S}\right\}. \tag{29}$$

Denote $d_\omega^x = \eta(x, \omega)$. The following assumptions are made on the structure in (23):

[A] For every $x \in \mathcal{X}$ and $\omega \in \Omega$, there exists a constant $\Gamma_\omega^x > 0$ such that $|s_{\Pi(\omega)}(d) - s_{\widehat{\Pi}_\psi}(d)| \leq \Gamma_\omega^x$ holds, for every $d \in \mathbb{R}^s \cap \{\eta(u,\omega)\}_{u\in\mathcal{X}}$. Here, $\psi = \omega$ if $\omega \in \mathcal{I}$, otherwise if $\omega \notin \mathcal{I}$, it is the $\psi \in \mathcal{I}$ where $\omega \in \mathcal{J}_\psi$.

[B] For every $x \in \mathcal{X}$ and $\omega \in \Omega$, there exists a constant κ_ω^x such that $s_{\widehat{\Pi}_\psi}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x \leq \kappa_\omega^x \|d_\omega^x\|$ holds. Here, $\psi = \omega$ if $\omega \in \mathcal{I}$, otherwise if $\omega \notin \mathcal{I}$, it is the $\psi \in \mathcal{I}$ where $\omega \in \mathcal{J}_\psi$.

Notice that the parameters Γ_ω^x and κ_ω^x, and consequently ϵ^* defined below, do depend on the set of clusters $\mathcal{I}$ and consequently the clustering mechanism through the collinearity measure (28).

Theorem 5.1 *Consider a two-stage risk-averse stochastic optimization problem with the first-stage problem* (1) *with the nonempty compact feasible set $\mathcal{X}$, and the second-stage problem* (2) *which has fixed and complete recourse. Suppose the optimal value of the second-stage problem $\mathcal{Q}(x,\omega)$ can be expressed as in equation* (23) *for some function η. Then, for every $x \in \mathcal{X}$ and $\epsilon_{cos} \in (0,1)$,* Algorithm 3 *along with the $\mathcal{I}$–selection method described in* Algorithm 4 *provides outputs $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ satisfying equations* (5) *and* (6) *with $\epsilon_1 = \epsilon_2 = \epsilon^* > 0$, where $\epsilon^* := \max_{x\in\mathcal{X},\omega\in\Omega}\{\Gamma_\omega^x + \kappa_\omega^x \|d_\omega^x\|\}$.*

Proof of Theorem 5.1 Fix $x \in \mathcal{X}$. Let $\mathcal{Q}^{approx}(x,\omega)$ be the approximation of $\mathcal{Q}(x,\omega)$ computed from Algorithm 3. According to equation (23) and using the notation in (29), we have $\mathcal{Q}(x,\omega) = s_{\Pi(\omega)}(d_\omega^x)$. If $\omega \in \mathcal{I}$, Step 2 in Algorithm 3 yields $\mathcal{Q}^{approx}(x,\omega) := s_{\widehat{\Pi}_\omega}(d_\omega^x)$. If $\omega \notin \mathcal{I}$, there exists some $\psi \in \mathcal{I}$ such that $\omega \in \mathcal{J}_\psi$. Hence, Algorithm 3 results in $\mathcal{Q}^{approx}(x,\omega) := \widetilde{\varpi}_\omega^\top d_\omega^x = \widetilde{\pi}_\psi^\top d_\omega^x$, where $\widetilde{\pi}_\psi$ solves problem $\max_{\pi\in\widehat{\Pi}_\psi} \pi^\top \eta(x,\psi)$. Therefore, the approximation error in the second-stage value function evaluation is given by

$$\epsilon_\omega := \mathcal{Q}(x,\omega) - \mathcal{Q}^{approx}(x,\omega) = \begin{cases} s_{\Pi(\omega)}(d_\omega^x) - s_{\widehat{\Pi}_\omega}(d_\omega^x), & \text{if } \omega \in \mathcal{I} \\ s_{\Pi(\omega)}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x, & \text{if } \omega \notin \mathcal{I} \text{ and } \omega \in \mathcal{J}_\psi. \end{cases} \tag{30}$$

For any $\omega \in \mathcal{I}$, it follows from assumption [A] that

$$\epsilon_\omega \leq \left| s_{\Pi(\omega)}(d_\omega^x) - s_{\widehat{\Pi}_\omega}(d_\omega^x) \right| \leq \Gamma_\omega^x. \tag{31}$$

For any $\omega \notin \mathcal{I}$ and $\omega \in \mathcal{J}_\psi$, from equation (30) we have

$$\begin{aligned} \epsilon_\omega \leq |s_{\Pi(\omega)}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x| &= |s_{\Pi(\omega)}(d_\omega^x) - s_{\widehat{\Pi}_\psi}(d_\omega^x) + s_{\widehat{\Pi}_\psi}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x| \\ &\leq |s_{\Pi(\omega)}(d_\omega^x) - s_{\widehat{\Pi}_\psi}(d_\omega^x)| + |s_{\widehat{\Pi}_\psi}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x| \\ &\leq \Gamma_\omega^x + \kappa_\omega^x \|d_\omega^x\|, \end{aligned} \tag{32}$$

where the inequality (32) follows from assumptions [A] and [B]. Hence,

$$\epsilon_\omega \leq \Gamma_\omega^x + \kappa_\omega^x \left\|d_\omega^x\right\|. \tag{33}$$

Therefore, from equations (31) and (33) we see that for any $\omega \in \Omega$,

$$\epsilon_\omega \leq \max\{\Gamma^x_\omega,\ \Gamma^x_\omega + \kappa^x_\omega \|d^x_\omega\|\} = (\Gamma^x_\omega + \kappa^x_\omega \|d^x_\omega\|) \leq \epsilon^*. \tag{34}$$

Note that given the compactness of $\mathcal{X}$ and finiteness of Ω, ϵ^* is well-defined and $\epsilon^* < \infty$.

With this bound (34) on estimation errors, we can focus on proving that the inexact risk-averse oracle satisfies requirements (5) and (6). We complete this step in two parts focusing on $\widetilde{\varphi}_x$ and $\widetilde{g}_x$, respectively.

Part 1: *Correctness of $\widetilde{\varphi}_x$:* It follows from the definition of $\varphi(x)$ in (1), $\varphi(x) = c^\top x + \sup_{\mu\in\partial\rho(0)} \sum_{\omega\in\Omega} \mathcal{Q}(x,\omega)\mu_\omega P_\omega$, and the expression (23) that

$$\begin{aligned}
\varphi(x) &= c^\top x + \sup_{\mu\in\partial\rho(0)} \left[\sum_{\omega\in\mathcal{I}} \mu^*_\omega P_\omega \mathcal{Q}(x,\omega) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \mathcal{Q}(x,\psi)\right] \\
&= c^\top x + \sup_{\mu\in\partial\rho(0)} \left[\sum_{\omega\in\mathcal{I}} \mu_\omega P_\omega \left(\widetilde{\pi}^\top_\omega \eta(x,\omega) + \epsilon_\omega\right) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \left(\widetilde{\varpi}^\top_\psi \eta(x,\psi) + \epsilon_\psi\right)\right] \\
&\leq c^\top x + \sup_{\mu\in\partial\rho(0)} \left[\sum_{\omega\in\mathcal{I}} \mu_\omega P_\omega \widetilde{\pi}^\top_\omega \eta(x,\omega) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \widetilde{\varpi}^\top_\psi \eta(x,\psi)\right] + \sup_{\mu\in\partial\rho(0)} \sum_{\omega\in\Omega} \mu_\omega P_\omega \epsilon^* \\
&= \widetilde{\varphi}_x + \epsilon^*.
\end{aligned} \tag{35}$$

The last equality comes from the definition of $\widetilde{\varphi}_x$ in equation (26), and the equality $\sup_{\mu\in\partial\rho(0)} \sum_{\omega\in\Omega} \mu_\omega P_\omega \epsilon^* = \epsilon^*$. This equality holds since by Theorem 2.1, for every $\mu \in \partial\rho(0)$, we have $\sum_{\omega\in\Omega} \mu_\omega P_\omega = 1$.

Similarly, we have

$$\begin{aligned}
\varphi(x) &= c^\top x + \sup_{\mu\in\partial\rho(0)} \left[\sum_{\omega\in\mathcal{I}} \mu_\omega P_\omega \widetilde{\pi}^\top_\omega \eta(x,\omega) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \widetilde{\varpi}^\top_\psi \eta(x,\psi) + \sum_{\omega\in\Omega} \mu_\omega P_\omega \epsilon_\omega\right] \\
&\geq c^\top x + \sup_{\mu\in\partial\rho(0)} \left[\sum_{\omega\in\mathcal{I}} \mu_\omega P_\omega \widetilde{\pi}^\top_\omega \eta(x,\omega) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \widetilde{\varpi}^\top_\psi \eta(x,\psi) - \sum_{\omega\in\Omega} \mu_\omega P_\omega \epsilon^*\right] \\
&= \widetilde{\varphi}_x - \epsilon^*.
\end{aligned} \tag{36}$$

Inequalities (35) and (36) imply that $\widetilde{\varphi}_x - \epsilon^* \leq \varphi(x) \leq \widetilde{\varphi}_x + \epsilon^*$, thus satisfying the requirement (5), $\widetilde{\varphi}_x \in [\varphi(x) - \epsilon^*,\ \varphi(x) + \epsilon^*]$, with $\epsilon_1 := \epsilon_2 := \epsilon^*$.

Part 2: *Correctness of $\widetilde{g}_x$.* Let $z \in \mathcal{X}$ such that $z \neq x$. Let π^z_ω be an optimal solution for problem (23), i.e.,

$$\mathcal{Q}(z,\omega) = \max_{\pi\in\Pi(\omega)} \pi^\top \eta(z,\omega) = (\pi^z_\omega)^\top \eta(z,\omega) = s_{\Pi(\omega)}(d^z_\omega), \tag{37}$$

where $d_\omega^z := \eta(z,\omega)$. From Algorithm 3, $\mathcal{Q}^{approx}(x,\omega) = \eta(x,\omega)^\top \vartheta_\omega^x$ for $\vartheta_\omega^x \in \mathcal{S}$, where $\mathcal{S} := \widehat{\Pi}_\omega$ (if $\omega \in \mathcal{I}$) and $\mathcal{S} := \widehat{\Pi}_\psi$ (if $\omega \notin \mathcal{I}$ and $\omega \in \mathcal{J}_\psi$). Using assumptions [A] at $d = d_\omega^z$, we get

$$\epsilon^* \geq \Gamma \geq |s_{\Pi(\omega)}(d_\omega^z) - s_{\widehat{\Pi}_\psi}(d_\omega^z)| \geq s_{\widehat{\Pi}_\psi}(d_\omega^z) - s_{\Pi(\omega)}(d_\omega^z) = \max_{\pi \in \widehat{\Pi}_\psi} \pi^\top d_\omega^z - s_{\Pi(\omega)}(d_\omega^z).$$

Since $\vartheta_\omega^x \in \widehat{\Pi}_\psi$, we have $\max_{\pi \in \widehat{\Pi}_\psi} \pi^\top d_\omega^z \geq (\vartheta_\omega^x)^\top d_\omega^z$. Applying this inequality in the above statement yields

$$s_{\Pi(\omega)}(d_\omega^z) \geq (\vartheta_\omega^x)^\top d_\omega^z - \epsilon^*. \tag{38}$$

Using (37) and (38) along with the equality $\mathcal{Q}^{approx}(x,\omega) = \eta(x,\omega)^\top \vartheta_\omega^x$ implies that

$$\begin{aligned} \mathcal{Q}(z,\omega) &\geq \left(\mathcal{Q}^{approx}(x,\omega) - \eta(x,\omega)^\top \vartheta_\omega^x\right) + \left((\vartheta_\omega^x)^\top d_\omega^z - \epsilon^*\right) \\ &= \mathcal{Q}^{approx}(x,\omega) - \epsilon^* + \left(\eta(z,\omega) - \eta(x,\omega)\right)^\top \vartheta_\omega^x. \end{aligned} \tag{39}$$

From the differentiability of $\eta(\cdot,\omega)$ at x, we have $\eta(z,\omega) - \eta(x,\omega) \geq \langle \nabla_x \eta(x,\omega),\, z - x\rangle = \nabla_x \eta(x,\omega)(z-x)$. Here, $\nabla_x \eta(x,\omega)$ is the $s \times n$ Jacobian matrix of η. Using this inequality in (39) along with $\vartheta_\omega^x \geq 0$ (problem (16) includes $\pi \geq 0$ as the constraints), we arrive at

$$\begin{aligned} \mathcal{Q}(z,\omega) &\geq \mathcal{Q}^{approx}(x,\omega) - \epsilon^* + \left(\langle \nabla_x \eta(x,\omega),\, z - x\rangle\right)^\top \vartheta_\omega^x \\ &= \mathcal{Q}^{approx}(x,\omega) - \epsilon^* + \left(\vartheta_\omega^x\right)^\top \nabla_x \eta(x,\omega)(z-x). \end{aligned} \tag{40}$$

From the definition of $\varphi(z)$ we have

$$\begin{aligned} \varphi(z) &= c^\top z + \sup_{\mu \in \partial\rho(0)} \left[\sum_{\omega \in \Omega} \mu_\omega P_\omega \mathcal{Q}(z,\omega)\right] = c^\top x + \sup_{\mu \in \partial\rho(0)} \left[\sum_{\omega \in \Omega} \mu_\omega P_\omega \mathcal{Q}(z,\omega) + c^\top (z-x)\right] \\ &\geq c^\top x + \sum_{\omega \in \Omega} \mu_\omega^* P_\omega \mathcal{Q}(z,\omega) + c^\top (z-x), \end{aligned}$$

where μ_ω^* is an optimal solution to the optimization problem in equation (26) for x. Using inequality (40) in the above equality yields

$$\begin{aligned} \varphi(z) &\geq c^\top x + \sum_{\omega \in \Omega} \mu_\omega^* P_\omega \left(\mathcal{Q}^{approx}(x,\omega) + \left(\vartheta_\omega^x\right)^\top \nabla_x \eta(x,\omega)(z-x) + c^\top (z-x) - \epsilon^*\right) \\ &= \left[c^\top x + \sum_{\omega \in \Omega} \mu_\omega^* P_\omega \mathcal{Q}^{approx}(x,\omega)\right] + \left[\sum_{\omega \in \Omega} \mu_\omega^* P_\omega \left(\vartheta_\omega^x\right)^\top \nabla_x \eta(x,\omega)(z-x) + c^\top (z-x)\right] - \epsilon^* \\ &= \varphi_x + \left[\sum_{\omega \in \Omega} \mu_\omega^* P_\omega \left(\nabla_x \eta(x,\omega)\right)^\top \vartheta_\omega^x + c\right]^\top (z-x) - \epsilon^* \\ &= \varphi_x + \langle \widetilde{g}_x,\, z - x\rangle - \epsilon^* \\ &\geq \varphi(x) + \langle \widetilde{g}_x,\, z - x\rangle - 2\epsilon^*. \end{aligned}$$

Here, the first equality uses $\sum_{\omega \in \Omega} \mu_\omega^* P_\omega \epsilon^* = \epsilon^*$. The last inequality comes from inequality (35). Thus, $\widetilde{g}_x \in \partial_{2\epsilon^*} \varphi(x)$ satisfies (6) with $\epsilon_1 := \epsilon_2 := \epsilon^*$. This completes the proof. □

5.2. The Case of Linear Second-Stage Problem

The modeling assumption (23) holds, for example, for linear second-stage problems, $q(y,\omega) = q_\omega^\top y$, $t_i^\top(x,\omega) = t_{i,\omega}x$, and $r_i(y,\omega) = r_i^\top y - h_\omega$, and $\mathcal{Y} = \mathbb{R}^s_+$, i.e.,

$$\begin{aligned} \mathcal{Q}(x,\omega) = \min_{y\in\mathbb{R}^s} \quad & q_\omega^\top y \\ \text{s.t.} \quad & Ry + T_\omega x \le h_\omega, \qquad y \ge 0, \end{aligned} \tag{41}$$

where q_ω, h_ω are vectors, T_ω is a stochastic matrix, and R is a fixed recourse matrix, i.e., $\omega := (q_\omega, T_\omega, h_\omega)$. The full recourse assumption on the second-stage problem implies that the strong duality for problem (41) holds and we have

$$\begin{aligned} \mathcal{Q}(x,\omega) = \max_{\pi\in\mathbb{R}^s} \quad & \pi^\top (T_\omega x - h_\omega) \\ \text{s.t.} \quad & R^\top \pi \ge -q_\omega, \qquad \pi \ge 0. \end{aligned} \tag{42}$$

Hence, $\mathcal{Q}(x,\omega)$ takes the form of (23) with $\Pi(\omega) := \left\{\pi \in \mathbb{R}^s_+ \middle| R^\top\pi \ge -q_\omega\right\}$ and $\eta(x,\omega) = T_\omega x - h_\omega$. Whence, $\nabla_x \eta(x,\omega)^\top \pi_\omega = T_\omega^\top \pi_\omega \in \partial_x \mathcal{Q}(x,\omega)$.

For the linear second-stage model (41), the expression (24) implies that $\widehat{\Pi}_\omega = \left\{\pi \in \mathbb{R}^s_+ \;\middle|\; R^\top\pi \le \hat{q}_\omega\right\}$, where $\hat{q}_\omega$ is the average of q_ψ over the collinear set $\mathcal{J}_\omega$, i.e.,

$$\hat{q}_\omega := \frac{\sum_{\xi\in\mathcal{J}_\omega} P_\xi q_\xi}{\sum_{\xi\in\mathcal{J}_\omega} P_\xi}. \tag{43}$$

For the linear model (41) both assumptions [A] and [B] stated in Subsection 5.1 hold. Applying the result that for the linear optimization problem (42) its feasible set is Lipschitz with respect to perturbations in the right-hand-side vector q_ω (e.g., see Theorem 2.4 of Mangasarian and Shiau (1987)), we have $\left|s_{\Pi(\omega)}(d_\omega^x) - s_{\Pi(\psi)}(d_\omega^x)\right| \le l_\omega \left\|q_\omega - q_\psi\right\|$, for some constant $l_\omega > 0$. Therefore, $\Gamma_\omega^x := \max_{\psi\in\omega}\{l_\psi \left\|q_\omega - q_\psi\right\|\}$ satisfies in assumption [A]. This establishes the validity of assumption [A] for linear second-stage models.

Assumption [B] also holds when the second-stage problem is linear, as in (42). To see this, suppose that $\bar{\pi}$ is an optimal solution of problem $\max_{\pi\in\widehat{\Pi}_\psi} \pi^\top d_\omega^x$, i.e., $s_{\widehat{\Pi}_\psi}(d_\omega^x) = \bar{\pi}^\top d_\omega^x$. Since $\widetilde{\pi}_\psi \in \widehat{\Pi}_\psi$, we have $\widetilde{\pi}_\psi^\top d_\omega^x < s_{\widehat{\Pi}_\psi}(d_\omega^x)$. Hence,

$$s_{\widehat{\Pi}_\psi}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x = \left|s_{\widehat{\Pi}_\psi}(d_\omega^x) - \widetilde{\pi}_\psi^\top d_\omega^x\right| = \left|\bar{\pi}^\top d_\omega^x - \widetilde{\pi}_\psi^\top d_\omega^x\right| \le \left\|\bar{\pi} - \widetilde{\pi}_\psi\right\| \left\|d_\omega^x\right\|.$$

Thus, assumption [B] holds for $\kappa_\omega^x := \left\|\bar{\pi} - \widetilde{\pi}_\psi\right\|$. Note that $\bar{\pi}$ and $\widetilde{\pi}_\psi$ depends on x and ω.

6. Resource Allocation for Contingency Planning

This section presents the details of our resource allocation problem and its formulation as a risk-averse two-stage stochastic optimization problem. For a detailed review on facility location problems, the reader is referred to (Daskin 1995, Drezner 1995). This problem aims to allocate a set

of reserve resources to the nodes in a network in order to achieve an optimal risk-adjusted level of cost versus reliability in the network. This problem arises for example in the optimal allocation of a finite number of energy storage facilities to different areas in an electricity grid, given area generation, area demand, and tie-line connections between areas. For details on this problem see see Jirutitijaroen and Singh (2006, 2008), Chowdhury et al. (2004). We formulate the optimal allocation problem as a two-stage risk-averse stochastic optimization problem, where the second-stage problem is modeled as a capacity network flow problem.

Note that in this problem, integer constraints only appear in the first-stage problem. This ensures the convergence of the inexact bundle method and particularly the results in Kiwiel (2006) remain valid. We start by describing the network components in the model.

6.1. The Model

Consider a multi-area network with the set of nodes $I := \{1, \ldots, n\}$ and edges $E \subseteq \{\{i, j\} \,|\, i, j \in I\}$. We let the network be directed and define $\tilde{E} := \{(i, j), (j, i) \,|\, \{i, j\} \in E\}$. Elements of randomness in the network are driven by a finite probability space $(\Omega, \mathcal{F}, P)$, where each $\omega \in \Omega$ represents an outcome of the system via the following given functions:

$t_{i,j}(\omega)$: Tie-line capacity between areas i and j under scenario ω: $t_{i,j}(\omega) \colon \tilde{E} \times \Omega \to \mathbb{R}$

$c_i^l(\omega)$: Cost of demand unfulfillment in area i under scenario ω: $c_i^l(\omega) \colon I \times \Omega \to \mathbb{R}$

$g_i(\omega)$: Production capacity of area i under scenario ω: $g_i(\omega) \colon I \times \Omega \to \mathbb{R}$

$l_i(\omega)$: Demand in area i under scenario ω: $l_i(\omega) \colon I \times \Omega \to \mathbb{R}$

In addition, the following parameters are provided:

c_i^b: Cost of an additional reserve resource in area i

G_i^b: Capacity of the reserve facility if installed in area i

B: Maximum capacity of reserve resources for contingency planning

The network $\left(I \cup \{G, L\}, \tilde{E} \cup \{(G, i)\}_{i \in I} \cup \{(i, L)\}_{i \in I}\right)$ with supply, demand, tie-line capacity, and cost demand functions encode the full state of the network under scenario $\omega \in \Omega$. This paper assumes that all described functions of scenario ω are known and given. We refer the reader to Jirutitijaroen and Singh (2006, 2008), Lago-conzalez and Singh (1989), Garver (1966), Mitra and Singh (1999), Lawton et al. (2003) for more details in the generation of state-dependent functions from available data. Our model can cast as a generalization of the model in Jirutitijaroen and Singh (2006, 2008). Our model does not assume that the events of network disruptions are independent. Our computational studies let the tie-line distributions be dependent.

The main objective of the problem is to efficiently allocate the given set of external reserve resources in terms of cost versus reliability, in the presence of uncertainties and state dependent functions $t_{i,j}$, c_i, g_i, and l_i.

The first stage decision variables, x_i's, are the number of reserve facilities to be allocated to each area. These integer decision variables must be determined before the realization of a random scenario $\omega \in \Omega$ for demands, generations, and congestions. Given an allocation $\{x_i\}_{i\in I}$, flows in the network for each scenario constitute the second-stage decision variables. Denote the flow from arc i to j for system state ω by $y_{ij}(w)$. The precise formulation of this two-stage problem is as follows:

$$\min_{x\in\mathbb{R}^n} \quad \textstyle\sum_{i=1}^n c_i^b x_i + \rho\left[\mathcal{Q}(x,\omega)\right] \tag{44}$$

$$\text{s.t.} \quad \textstyle\sum_{i=1}^n x_i \le B \tag{45}$$

$$x \ge 0,\ x \text{ is integral}, \tag{46}$$

where $\rho : \mathcal{Z} \to \mathbb{R}$ is a coherent risk measure, and

$$\mathcal{Q}(x,\omega) = \min_{\substack{y\in\mathbb{R}^{|\tilde{E}|}\\ y_G, y_L\in\mathbb{R}^n}} \quad \sum_{i=1}^n c_i^l(\omega)\left(l_i(\omega) - y_{L,i}(\omega)\right) \tag{47}$$

$$\text{s.t.} \quad y_{G,i}(\omega) \le g_i(\omega) + G_i^b x_i, \quad i \in \{1,2,\cdots,n\}, \tag{48}$$

$$y_{L,i}(\omega) \le l_i(\omega), \quad i \in \{1,2,\cdots,n\}, \tag{49}$$

$$|y_{ji}(\omega) - y_{ij}(\omega)| \le t_{i,j}(\omega), \quad i,j \in \{1,2,\cdots,n\},\ i \ne j, \tag{50}$$

$$\sum_{\substack{j\in I\\ j\ne i}} y_{ji}(\omega) - \sum_{\substack{j\in I\\ j\ne i}} y_{ij}(\omega) + y_{G,i}(\omega) - y_{L,i}(\omega) = 0, \quad i \in \{1,2,\cdots,n\}, \tag{51}$$

$$y_{ij}(\omega), y_{G,i}(\omega), y_{L,i}(\omega) \ge 0, \quad i,j \in \{1,2,\cdots,n\}. \tag{52}$$

Function $\mathcal{Q}(x,\omega)$ is the optimal objective value of the second-stage problem of minimizing the cost of demand unfulfillment under scenario $\omega \in \Omega$. Constraints (48)–(50) correspond to the maximum capacity flow in the network. These constraints consider generation, demand, and tie-line capacity, respectively. Equation (51) is the flow conservation constraint. The cost $c_i^l(\omega)$ is the penalty cost of not serving the customer per unit of missed demand. Hence, the objective function (48), $c_i^l(\omega)(l_i(\omega) - y_{L,i}(\omega))$, serves as an unreliability index, which we aim to minimize. The failure cost to measure reliability is typical in the facility location literature, e.g., see Cui et al. (2010).

In the first-stage problem, $\mathcal{Q}(x,\omega)$ is a random variable on Ω. Equation (45) is the first-stage bound on the total number of reserve facilities to be allocated. This bound is based on our resource availability on such components. The objective function (44) adds the cost of additional reserve capacities to the risk-averse evaluation of the second-stage cost of demand loss under uncertainty.

The risk measure ρ as included in the first-stage problem makes our model fundamentally different from the standard literature on related problems, e.g. Jirutitijaroen and Singh (2006, 2008), Ruszczyński and Shapiro (2006b,a, 2007), Shapiro et al. (2014). More specifically, instead of considering as our recourse the (risk-neutral) expectation of the cost of demand unfulfillment, we use

a risk measure of the second-stage objective value. This enables the decision maker to incorporate his risk preferences in the reliability management while allocating the reserve resources.Under this paradigm, the decision maker is capable of placing more attention on particular scenarios based on his risk preferences.

6.2. Coherent Risk Measures for Network Reliability Assessment

We illustrate the inexact bundle method with the risk averse oracle for two important coherent risk measures: *Mean-Upper Semideviation* and *Coditional-Value-at-Risk CVaR*. For further review on these risk measures, see Shapiro et al. (2014). Below, $[a]_+ := \max\{0, a\}$.

Definition 6.1 *Let $\alpha \in [0,1]$. The* mean-upper semideviation *measures the risk of our loses exceeding the expectation and is defined by*

$$\rho(Z) = \mathbb{E}[Z] + \alpha \mathbb{E}\left[Z - \mathbb{E}[Z]\right]_+, \quad \forall\ Z \in \mathcal{Z}.$$

For a finite probability space Ω and probability density function $p = (p_1, \ldots, p_{|\Omega|})$, we have that ρ's subdifferential at zero is given by

$$\partial\rho(0) = \left\{ \mathbb{1} - \mathbb{1}\sum_{i=1}^{|\Omega|} p_i\tau_i + \tau \;\middle|\; \tau = (\tau_i)_{i=1}^{|\Omega|} \text{ and } 0 \le \tau_i \le \alpha,\ \forall i = 1, \ldots, |\Omega| \right\}, \tag{53}$$

where $\mathbb{1}$ is the vector with all entries equal to 1.

Definition 6.2 *Let $\alpha \in [0,1]$ and $VaR_\alpha[Z] := \inf\{t \,|\, Pr(Z \le t) \ge 1-\alpha\}$, $\forall z \in \mathcal{Z}$, be the* value-at-risk *(at level α) operator. The* conditional value-at-risk *at level α is defined by*

$$CVaR_\alpha(Z) := \inf_{t\in\mathbb{R}} \left\{ t + \alpha^{-1}\mathbb{E}[Z-t]_+ \right\} = \frac{1}{\alpha}\int_{1-\alpha}^{1} VaR_{1-\tau}(Z)d\tau, \quad \forall\ Z \in \mathcal{Z}.$$

For a finite probability space Ω and probability density function $p = (p_1, \ldots, p_{|\Omega|})$, we have that $CVaR_\alpha$'s subdifferential at zero is given by

$$\partial(CVaR_\alpha)(0) = \left\{ \tau \in \mathbb{R}^{|\Omega|} \;\middle|\; \mathbb{E}[\tau] = 1 \text{ and } 0 \le \tau_i \le 1/\alpha,\ \forall i = 1, \ldots, |\Omega| \right\}. \tag{54}$$

The subdifferentials in (53) and (54) can be expressed by linear programming formulations. For example, (53) imply that under the mean-upper semideviation risk measure we obtain $\phi(x)$ in (12) by solving the following linear optimization problem:

$$\begin{aligned} \sup_{\mu,\tau\in\mathbb{R}^{|\Omega|}} \quad & \sum_{\omega\in\Omega} \bar{q}_\omega \mu_\omega \\ \text{s.t.} \quad & \mu_\omega = 1 - \sum_{\nu\in\Omega} P_\nu \tau_\nu + \tau, \quad \omega \in \Omega \\ & 0 \le \tau_\omega \le \kappa, \quad \omega \in \Omega, \end{aligned}$$

where $\bar{q}_\omega = q(y_\omega, \omega)P_\omega$, for every $\omega \in \Omega$.

Continuing with the example of mean-upper semideviation, we show below how to compute the essential components needed in Algorithm 1, namely $\nabla T(x,\omega)$ and $\zeta \in \partial\phi(x)$. It follows from the rules of subdifferential calculus that for $x \in X$ and $\omega \in \Omega$ the gradient $\nabla T(x,\omega)$ corresponding to the two-stage model (44)-(52) is given by

$$(\nabla T(x,\omega))^\top = \begin{bmatrix} -G_1^b(e_n^1) & \cdots & -G_n^b(e_n^n) & 0_n & \cdots & 0_n \end{bmatrix} := (\nabla T)^\top, \tag{55}$$

where $\mathbf{0}_n \in \mathbb{R}^n$ is the zero vector of size n, and e_n^i, $i = 1, \cdots, n$, is the canonical basis of $\mathbb{R}^n$. There are $2n^2 + n$ consecutive transposed zero vectors at the bottom of the vector in (55). Since $\nabla T(x,\omega)$ does not depend on x or ω, we simplify the notation and denote it by ∇T.

Applying equations (19) and (55), given μ^*, optimal solution to equation (12), and primal-dual optimal pairs (y_ω, π_ω), $\forall \omega \in \Omega$, we obtain $\zeta \in \partial\phi(x)$ by

$$\zeta = \sum_{\Omega} \mu^*(\omega) P_\omega \begin{bmatrix} -G_1^b(e_n^1)^\top \\ \vdots \\ -G_n^b(e_n^n)^\top \\ 0_n^\top \\ \vdots \\ 0_n^\top \end{bmatrix}^\top \pi_\omega = \begin{bmatrix} \sum_{\omega\in\Omega} \mu^*(\omega) P_\omega \left(-G_1^b \pi_{\omega,1}\right) \\ \vdots \\ \sum_{\omega\in\Omega} \mu^*(\omega) P_\omega \left(-G_n^b \pi_{\omega,n}\right) \end{bmatrix}, \tag{56}$$

where $\pi_{\omega,i}$ is the i-th coordinate of vector π_ω.

6.3. Application of Risk-Averse Inexact Bundle Method

This section presents the details when applying the risk-averse inexact bundle method described in Algorithms 2 to 4 for the two-stage model (44)-(52). This time we use CVaR in our description.

For $x \in \mathcal{X}$ and $\omega \in I$, we consider $\mathcal{Q}(x,\omega)$ as given in (47)–(52) but with cost vector q_ω replaced by the corresponding average cost $\tilde{q}_\omega$ vector given in eq. (43). We solve the resulting linear optimization problem (via for example the dual simplex method) and obtain the pair of primal–dual solutions $(\widetilde{y}_\omega, \widetilde{\pi}_\omega)$. The dual solution $\widetilde{\pi}_\omega$ is used to get $\widetilde{\varphi}_x$ as stated in eq. (26), where $\partial(\mathrm{CVaR}_\alpha)(0)$ is formed from finitely many inequalities (54). More explicitly, $\widetilde{\varphi}_x$ is obtained as a solution to

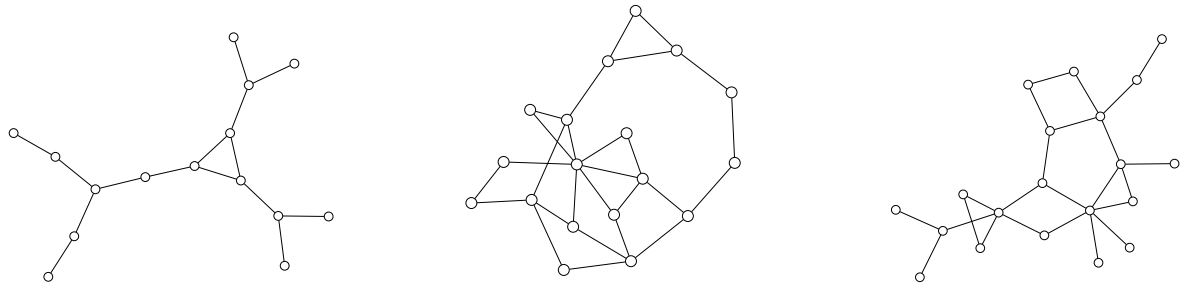
$$\begin{aligned} c^\top x + \sup_{\mu\in\mathbb{R}^{|\Omega|}} & \left[\sum_{\omega\in\mathcal{I}} \mu_\omega P_\omega \widetilde{\pi}_\omega^\top (h_\omega - T_\omega x) + \sum_{\psi\notin\mathcal{I}} \mu_\psi P_\psi \widetilde{\varpi}_\psi^\top (h_\psi - T_\psi x) \right] \\ \text{s.t.} & \sum_{\nu\in\Omega} P_\nu \mu_\nu = 1 \quad \text{and} \quad 0 \le \mu_\nu \le 1/\alpha,\ \nu \in \Omega, \end{aligned} \tag{57}$$

where $\widetilde{\varpi}_\psi$, $\psi \notin \mathcal{I}$, is as described in Step 3 of Algorithm 3.

Equation (27) is then used to obtain the approximate subgradient $\widetilde{g}_x$. To complete this phase, we use the optimal solution μ_ω^* from (57) and T_ω from the linearization of $\mathcal{Q}(x,\omega)$.

7. Computational Results

In our numerical experiments we randomly generate sparse connected networks of different sizes (see Figure 1).



(a) 15 nodes. (b) 18 nodes. (c) 20 nodes.

Figure 1 Examples of randomly generated connected networks of 15, 18, and 20 nodes.

A set of simulations is used to generate the set of scenarios Ω.

- $t_{i,j}$: For each tie-line $(i,j) \in \tilde{E}$, the tie-line capacity is generated as $t_{i,j} = \beta_{i,j}\,(c_{i,j} + \Lambda_{i,j})$, where $\beta_{i,j}$ is a Bernoulli random variable with probability $p_{i,j}$, $c_{i,j}$ is the (i,j) tie-line average capacity, and Λ is a multivariate Poisson random variable with parameter $\lambda = (\lambda_{i,j})_{(i,j)\in\tilde{E}}$. The random variable $\beta_{i,j}$ represents the probability of tie-line (i,j) to be completely cut-off due to an extreme event. This, coupled with the Poisson marginal distribution $\Lambda_{i,j}$ with intensity factor $\lambda_{i,j}$, enables the simulation of failures across multiple lines. In order to obtain a wide range of extreme events we setup the parameters to $p_{i,j} = 0.7$, $c_{i,j} = 900$, and $\lambda_{i,j} = 0.5$.
- c_i^l: The cost of demand loss is simulated independently as $c_i^l = \beta_i^\ell \times 4,000 + 1,000$, where β_i^ℓ is a Bernoulli random variable with probability $p_i = 0.5$.
- g_i: Generation capacity is simulated independently by a discrete random variable under the assumption of equally likely 500, 600, 800, 950, 1,000, and 1,200 MW of generation.
- l_i: Similar to tie-line generation, demand levels l_i is generated as a discrete stochastic process with log-Poisson spikes; hence, capturing demand spikes in the network.

In our simulations, we set $B = 10n$, where n is the number of nodes. Algorithms (1)-(2) are implemented in the Python programming language with calls to Gurobi quadratic, mixed-integer, and dual simplex solvers (Gurobi Optimization 2016). To improve efficiency, we parallelized the calls to multiple scenarios $\omega \in \Omega$. An Ubuntu 14.04 PC with dual Intel Xeon E5-2650 v4 @ 2.20GHz CPU with a total of 48 threads available for computation and 128GB of RAM is used in these numerical experiments.

7.1. Computational Time: Inexact Bundle Method versus Exact Bundle Method

This section reports the run-time of the inexact bundle method and compares it with the run-time of the exact bundle method. The analysis on run-time is conducted by varying the number of nodes in the network, the number of scenarios in the scenario set Ω, and the parameter $\epsilon_{\cos}$ in the inexact method. The results are presented for three risk functionals ρ: expectation, mean-upper semideviation, and CVaR.

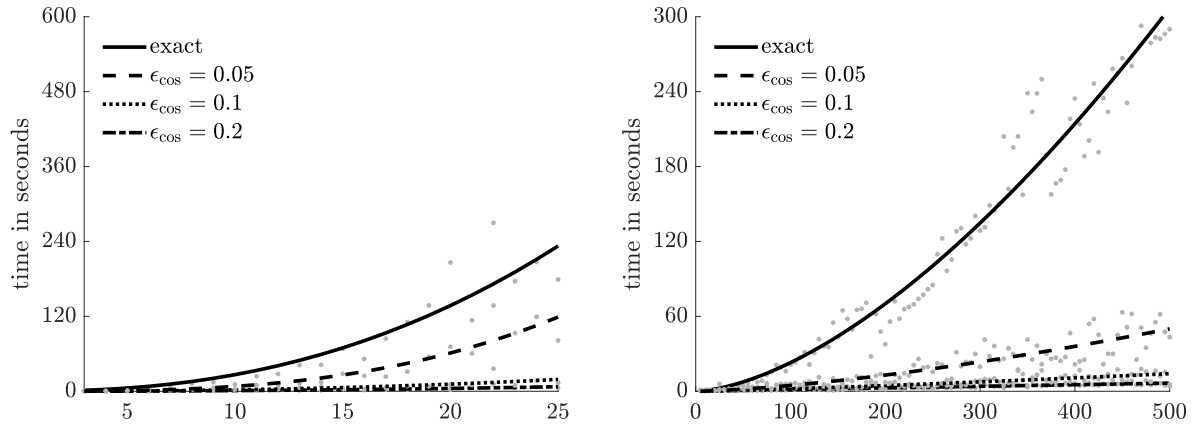
Figure 2 shows the run-times (in seconds) of the exact method and the inexact bundle methods (with different collinearity parameters $\epsilon_{\cos} = 0.2, 0.1, 0.05$) with different risk measures. In Figure 2, left plots illustrate the run-times to networks of 5 to 25 nodes. Each network was randomly generated. In this experiment, the scenario set has $|\Omega| = 100$ scenarios. Every point in these plots represent the average over 5 simulations, as described before, of $c_i^l(\omega)$, $g_i(\omega)$, $l_i(\omega)$, and $t_{i,j}(\omega)$. In the exact method, the regularization parameter is set to $\gamma = 0.31$. In the inexact method, the descent parameter is $\kappa = 0.3$, the initial stepsize bound is $T_1 = 0.05$, and the initial stepsize is $t_1 = 0.1$. In both exact and inexact methods tolerance level $\delta = 10^{-6}$ is used.

Plots in Figure 2 illustrate that significant run-time improvements ($85\% - 95\%$ for $\epsilon_{\cos} = 0.1, 0.2$ and $65\% - 75\%$ for $\epsilon_{\cos} = 0.05$) can be achieved through the inexact method. The run time for both exact and inexact method for the case of CVaR is lower than the other two risk measures, which is more prominent as the number of nodes grows. The time improvement from inexact bundle method is slightly higher for expectation and mean-upper semideviation risk functionals than the CVaR.

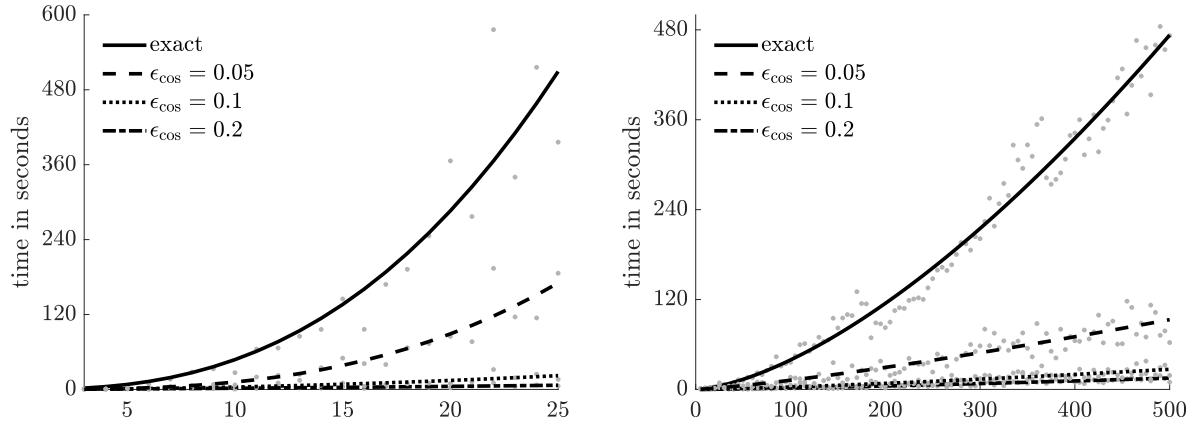
In Figure 2, right plots depict the run-time as the number of scenarios $|\Omega|$ increases. Here, we consider a network with 10 nodes and 20 arcs and vary the number of scenarios. Each point represents the average over 3 simulations. These results indicate that higher run-time improvements are observed as the size of the scenario set increases. The run-time for CVaR is again lower than that of the mean-upper semi deviation risk measure and the expectation.

7.2. Accuracy: Inexact Bundle Method versus Exact Bundle Method

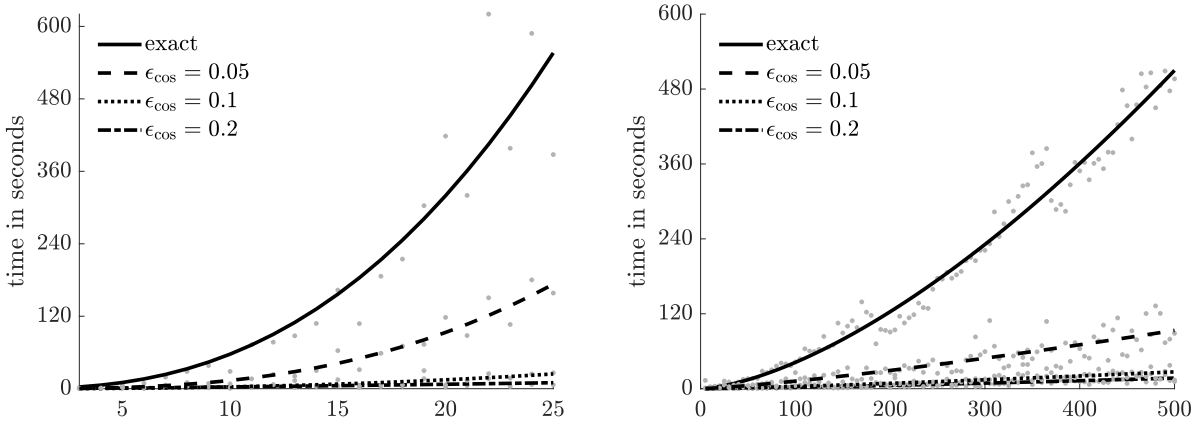
Figure 3 exhibits the percentage of approximation error (suboptimality) of the solution computed from the inexact bundle method. The suboptimality is computed as the absolute value of (optimal value of the inexact method-optimal value of the exact method) divided by the optimal value of the exact method. These plots are obtained from applying the exact bundle method and the inexact bundle methods (with $\epsilon_{\cos} = 0.2, 0.1, 0.05$) under different risk functionals to random networks of 2 to 25 nodes. Each point represents the average over 5 simulations on a fixed network with 100 randomly generated scenarios. Other settings are similar to those explained in Subsection 7.1. The error in approximation improves with smaller values of $\epsilon_{\cos}$. This is expected since a smaller $\epsilon_{\cos}$ creates a finer partition $\mathcal{I}$, thus leading to solving more second-stage scenarios in Algorithm 3.



(a) Conditional Value-at-Risk with $\alpha = 0.40$



(b) Mean-Upper Semideviation Risk Measure with $\alpha = 0.40$



(c) Expectation $\mathbb{E}$

Figure 2 **Run time (in seconds) for exact and inexact bundle methods (with $\epsilon_{\cos} = 0.2, 0.1, 0.05$). Left plots: Run time vs. number of nodes. The x-axis represents the number of nodes in the network. Right plots: Run time vs. size of the scenario set. The x-axis represents the number of scenarios in the simulation.**

The suboptimality level is relatively consistent among different risk measures and different scenario sizes. Right plots in Figure 3 indicate that a solution with an acceptable level of accuracy can be obtained from the inexact method even for a large number of scenarios. This result along with the time improvements achieved in the inexact method makes this approach attractive for solving risk-averse two-stage stochastic problem with downside risk measures. For risk measures such as CVaR, a large number of scenarios must be considered to accurately capture the tail of the recourse function distributions. This leads to high computational time in the exact method for risk-averse two-stage stochastic problems.

Table 1 reports the run-times of the inexact bundle method for different risk measures and the corresponding suboptimality levels against the exact bundle method. All parameters are fixed except $\epsilon_{\cos}$. In this analysis, a network with 20 nodes and 48 arcs are considered. The size of the scenario set is $|\Omega| = 100$, and the risk parameters in CVaR and mean-upper semideviation are set to $\alpha = 0.40$. The results in Table 1 suggest that $\epsilon_{\cos} = 0.1$ offers an acceptable tradeoff between approximation error and run-time.

CVaR

$\epsilon_{\cos}$	Time	Suboptimality
0.9	2.60	80.85%
0.8	2.32	80.85%
0.7	2.34	80.85%
0.6	2.33	80.85%
0.5	2.33	80.85%
0.4	4.68	48.04%
0.3	5.01	44.58%
0.2	8.34	26.85%
0.1	**28.10**	**10.51%**
0.09	20.33	12.42%
0.08	21.77	5.28%
0.07	41.47	4.28%
0.06	41.88	2.95%
0.05	94.53	1.76%
0.04	61.97	0.23%
0.03	114.55	0.00%
0.02	71.03	0.00%
0.01	69.38	0.00%
Exact running time: 80.43		

Semideviation

$\epsilon_{\cos}$	Time	Suboptimality
0.9	2.72	71.87%
0.8	2.44	71.87%
0.7	2.46	71.87%
0.6	2.46	71.87%
0.5	2.47	71.87%
0.4	4.75	36.16%
0.3	5.50	32.65%
0.2	7.23	14.81%
0.1	**15.22**	**0.31%**
0.09	27.55	8.06%
0.08	23.73	2.31%
0.07	31.83	3.14%
0.06	42.43	2.05%
0.05	69.67	1.67%
0.04	79.26	0.12%
0.03	69.62	0.08%
0.02	61.04	0.00%
0.01	61.28	0.00%
Exact running time: 121.44		

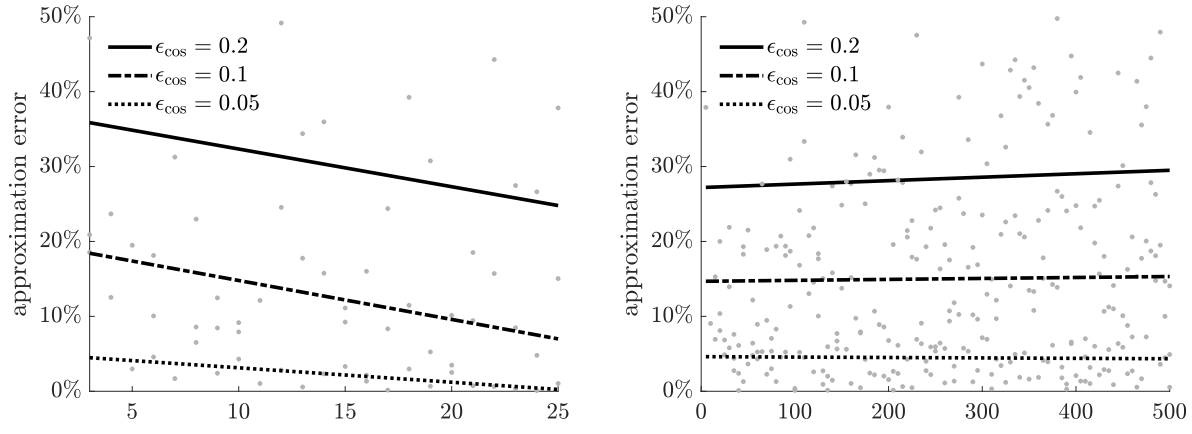
Expectation

$\epsilon_{\cos}$	Time	Suboptimality
0.9	2.52	63.58%
0.8	2.47	63.58%
0.7	2.46	63.58%
0.6	2.47	63.58%
0.5	2.47	63.58%
0.4	4.53	32.60%
0.3	4.85	29.30%
0.2	7.54	11.95%
0.1	**15.60**	**3.30%**
0.09	14.71	7.74%
0.08	22.52	0.84%
0.07	50.80	1.55%
0.06	49.93	1.16%
0.05	103.80	1.27%
0.04	148.38	0.14%
0.03	96.96	0.11%
0.02	124.28	0.00%
0.01	121.38	0.00%
Exact running time: 124.87		

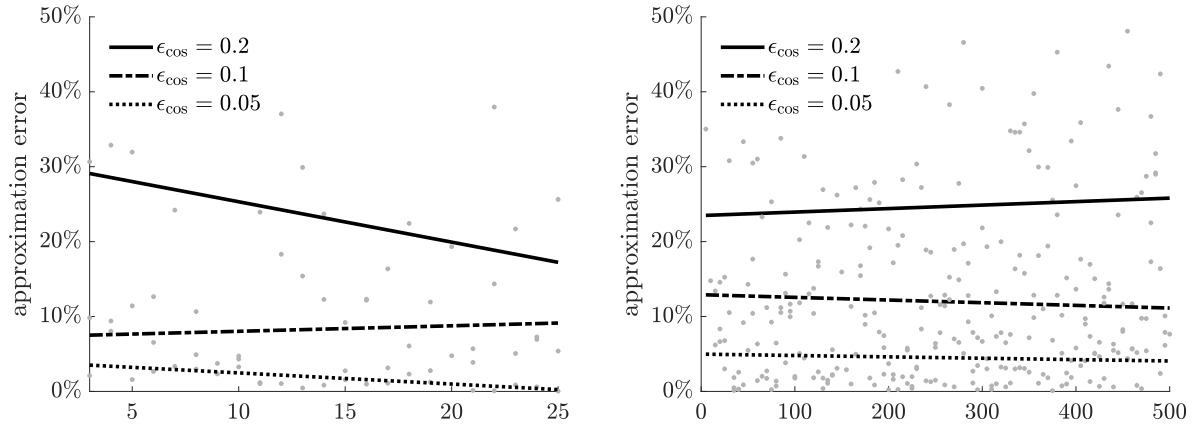
Table 1 Run-time (in seconds) of the inexact bundle methods as a function of $\epsilon_{\cos}$ and the percentage of sub-optimality (approximation errors in %) against the exact bundle method.

8. Conclusion

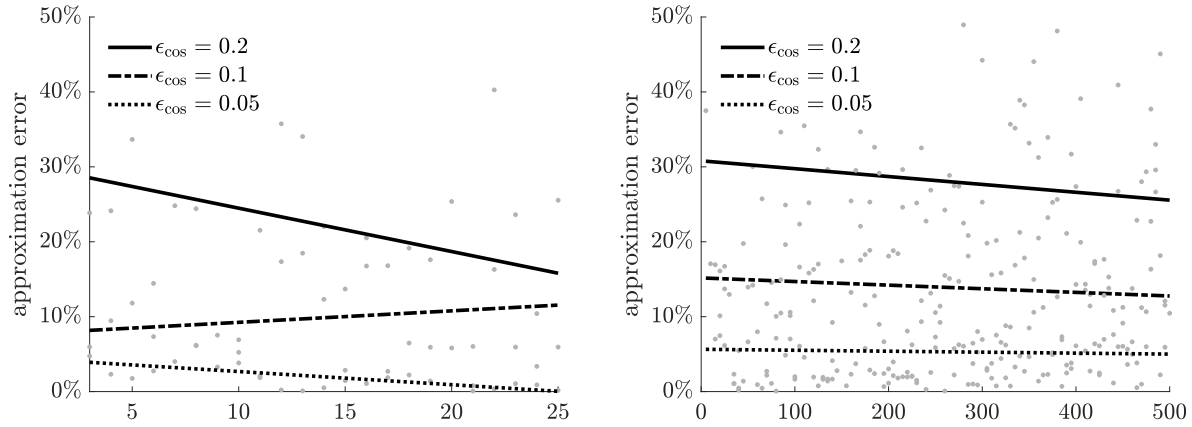
This paper studies the resource allocation problem as a two-stage stochastic optimization problem with risk-averse recourse. An inexact bundle method with a risk-averse oracle for evaluating objective function and a subgradient is developed and the correctness of the risk-averse oracle is theoretically established. The performance of the methodology is investigated using an application in a network reliability management using reserve resources. Our computational experiments



(a) Conditional Value-at-Risk with $\alpha = 0.40$



(b) Mean-Upper Semideviation Risk Measure with $\alpha = 0.40$



(c) Expectation $\mathbb{E}$

Figure 3 **Suboptimality of (linear fits) inexact bundle methods (with $\epsilon_{\cos} = 0.2, 0.1, 0.05$). Left plot: Suboptimality vs. number of nodes. The x-axis represent the number of nodes in the network. Right plot: Suboptimality vs. size of the scenario set. The x-axis represent the number of scenarios in the simulation.**

exhibit that the inexact bundle method can provide a significant improvement in the run-time to achieve a solution for the two-stage problem with a low relative error. A sensitivity analysis on the scenario clustering parameter $\epsilon_{\cos}$ for this two-stage risk averse stochastic problem is carried out, which guides on the selection of an appropriate value for this parameter.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1610302.

References

Alem, Douglas, Alistair Clark, Alfredo Moreno. 2016. Stochastic network models for logistics planning in disaster relief. *European Journal of Operational Research* **255**(1) 187–206.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, David Heath. 1999. Coherent measures of risk. *Mathematical Finance* **9**(3) 203–228.

Birge, J. R., F. V. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.

Choi, Sungyong, Andrzej Ruszczyński. 2008. A risk-averse newsvendor with law invariant coherent measures of risk. *Oper. Res. Lett.* **36**(1) 77–82.

Chowdhury, A. A., B. P. Glover, L. E. Brusseau, S. Hebert, F. Jarvenpaa, A. Jensen, K. Stradley, H. Turanli, G. E. Haringa. 2004. Assessing mid-continent area power pool capacity adequacy incl uding transmission limitations. *2004 International Conference on Probabilistic Methods Applied to Power Systems*. 56–63.

Collado, R., D. Papp, A. Ruszczyński. 2012. Scenario decomposition of risk-averse multistage stochastic programming problems. *Annals of Operations Research* **200**(1) 147–170.

Cui, Tingting, Yanfeng Ouyang, Zuo-Jun Max Shen. 2010. Reliable facility location design under the risk of disruptions. *Operations Research* **58**(4) 998–1011.

Daskin, M. S. 1995. *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley, New York.

Drezner, Z. 1995. *Facility Location: A Survey of Applications and Methods*. Springer, New York.

Garver, L. L. 1966. Effective load carrying capability of generating units. *IEEE Transactions on Power Apparatus and Systems* **PAS-85**(8) 910–919. doi:10.1109/TPAS.1966.291652.

Grass, Emilia, Kathrin Fischer. 2016. Two-stage stochastic programming in disaster management: A literature survey. *Surveys in Operations Research and Management Science* **21** 85–100.

Gupta, Sushil, Martin K. Starr, Reza Zanjirani Farahani, Niki Matinrad. 2016. Disaster management from a pom perspective: Mapping a new domain. *Production and Operations Management* **25**(10) 1611–1637.

Gurobi Optimization, Inc. 2016. Gurobi optimizer reference manual. URL `http://www.gurobi.com`.

Hiriart-Urruty, J-B., C. Lemarechal. 1993. *Convex Analysis and Minimization Algorithms, Volume II: Advanced Theory and Bundle Methods*. Springer-Verlag, Berlin.

Jirutitijaroen, P., C. Singh. 2008. Reliability constrained multi-area adequacy planning using stochastic programming with sample-average approximations. *IEEE Transactions on Power Systems* **23**(2) 504–513. doi:10.1109/TPWRS.2008.919422.

Jirutitijaroen, Panida, Chanan Singh. 2006. Multi-area generation adequacy planning using stochastic programming. *Proceedings of the IEEE Power Systems Conference and Exposition (PSCE)* 1327–1332.

Kall, P., J. Mayer. 2005. *Stochastic Linear Programming*. Springer, New York.

Kiwiel, Krzysztof C. 1990. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming* **46**(1) 105–122.

Kiwiel, Krzysztof C. 2006. A proximal bundle method with approximate subgradient linearizations. *SIAM Journal on Optimization* **16**(4) 1007–1023.

Lago-conzalez, A., C. Singh. 1989. The extended decomposition-simulation approach for multi-area reliability calculations (interconnected power systems). *Conference Papers Power Industry Computer Application Conference*. 66–73. doi:10.1109/PICA.1989.38975.

Lawton, L., M. Sullivan, K. V. Liere, A. Katz, J. Eto. 2003. A framework and review of customer outage costs: Integration and analysis of electric utility outage cost surveys. Tech. Rep. LBNL-54365, Lawrence Berkeley National Laboratory.

Mangasarian, Olvi L, T-H Shiau. 1987. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization* **25**(3) 583–595.

Matta, Renato De. 2016. Contingency planning during the formation of a supply chain. *Annals of Operations Research* 1–31.

Miller, Naomi, Andrzej Ruszczyński. 2008. Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. *European J. Oper. Res.* **191**(1) 193–206.

Miller, Naomi, Andrzej Ruszczyski. 2011. Risk-averse two-stage stochastic linear programming: Modeling and decomposition. *Operations Research* **59**(1) 125–132.

Mitra, J., C. Singh. 1999. Pruning and simulation for determination of frequency and duration indices of composite power systems. *IEEE Transactions on Power Systems* **14**(3) 899–905. doi:10.1109/59.780901.

Noyan, Nilay. 2012. Risk-averse two-stage stochastic programming with an application to disaster management. *Computers and Operations Research* **39** 541–559.

Oliveira, W., C. Sagastizabal. 2014. Level bundle methods for oracles with on-demand accuracy. *Optimization Methods and Software* **29**(6) 1180–1209.

Oliveira, Welington, Claudia Sagastizbal, Susana Scheimberg. 2011. Inexact bundle methods for two-stage stochastic programming. *SIAM Journal on Optimization* **21**(2) 517–544.

Prékopa, A. 1995. *Stochastic Programming*. Kluwer, Dordrecht.

Ruszczyński, A. 2003. Decomposition methods. *Stochastic Programming, Handbooks Oper. Res. Management Sci.* **10**.

Ruszczynski, Andrzej. 2006. *Nonlinear Optimization*. Princeton University Press, Princeton, NJ, USA.

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

Ruszczyński, Andrzej, Alexander Shapiro. 2006a. Conditional risk mappings. *Math. Oper. Res.* **31**(3) 544–561.

Ruszczyński, Andrzej, Alexander Shapiro. 2006b. Optimization of convex risk functions. *Math. Oper. Res.* **31**(3) 433–452.

Ruszczyński, Andrzej, Alexander Shapiro. 2007. Corrigendum to: "Optimization of convex risk functions," Math. Oper. Res. 31 (2006) 433–452. *Math. Oper. Res.* **32**(2) 496–496.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2014. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. MPS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics.

Snyder, L. V., M. P. Scaparra, M. L. Daskin, R. C. Church. 2006. Planning for disruptions in supply chain networks. *Tutorials in Operations Research* 234–257.

Teo, Choon Hui, S.V. N. Vishwanathan, Alex Smola, Quoc V. Le. 2010. Bundle methods for regularized risk minimization. *Journal of Machine Learning Research* **11** 311–365.

Tomlin, B. T. 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science* **52**(5) 639–657.