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

Somayeh Moazeni* , Ricardo A. Collado
School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA
smoazeni@stevens.edu, rcollado@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. The solution method proposed relies on an inexact proximal bundle method with subgradient approximations through a scenario reduction mechanism. The paper extends the inexact oracle to a more general risk-averse setting, and proves that it meets 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 coherent risk measures. Our analyses indicate 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, proximal bundle method

## 1. Introduction

This paper studies the optimal allocation of resources to reduce the risk of demand unfulfillment 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, fires, traffic accidents, or sabotage. 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.

For various disaster management strategies, the reader is referred to Gupta et al. (2016). The addition of reserve capacity in the supply chain for contingency planning is studied in Kleindorfer

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

and Saad (2005), Matta (2016), Parajuli et al. (2017), Avlov et al. (2019), where network optimization tools are used. For a recent review on contingency planning and resilient strategies in supply chains, see Behzadi et al. (2020). Grass and Fischer (2016) reviews the literature 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), Guo et al. (2016), Chen et al. (2017), Moreno et al. (2018), Avlov et al. (2019) 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. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where $\Omega$ is the sample space, $\mathcal{F}$ is a $\sigma$-algebra on $\Omega$, and $P$ is a probability measure on $\Omega$. We consider the case of a finite probability space with a potentially very large number of elementary events $\omega_1, \cdots, \omega_N$ occurring with probabilities $p_1, \cdots, p_N$. 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 $\rho$ is a risk measure and $\mathcal{Q}(x, \omega)$ is the optimal value of the second-stage problem

$$\begin{aligned} \mathcal{Q}(x, \omega) := \min_{y \in \mathcal{Y}} \quad & q(y, \omega) \\ \text{s.t.} \quad & T(x, \omega) + R(y, \omega) \leq 0. \end{aligned} \tag{2}$$

Here, $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{Y} \subseteq \mathbb{R}^m$, and $c \in \mathbb{R}^n$. In the second-stage problem (2), $q$ is a real-valued function on $\mathbb{R}^m \times \Omega$, and $T$ and $R$ are vector-valued functions on $\mathbb{R}^n \times \Omega$ and $\mathbb{R}^m \times \Omega$, respectively.

In this paper, we focus on coherent risk measures (Artzner et al. 1999) and convex finite-valued second-stage optimal value functions. The convexity of $\mathcal{Q}(\cdot, \omega)$ together with the convexity and monotonicity of the coherent risk measures $\rho$ imply that $\varphi$ is a proper convex function, e.g., see Proposition 6.8 in Shapiro et al. (2014). In addition, $\varphi(\cdot)$ is subdifferentiable over the interior of its domain, see Theorem 6.11 in Shapiro et al. (2014). However, the function $\varphi(x)$ is generally nondifferentiable and the problem (1) becomes a nonsmooth optimization problem.

One approach for solving convex minimization problems with a nonsmooth objective function is the bundle method; see Hiriart-Urruty and Lemarechal (1993) and Section 7.4 in Ruszczynski (2006) for details on the bundle method, and see Nesterov (2018) for an overview on nonsmooth convex optimization. For a survey on applications of the bundle method and comparisons to alternative methods, see Mäkelä (2002) and references therein. Mäkelä et al. (2013) compares implementations of bundle-type methods against subgradient methods for nonsmooth optimization problems and have found its efficiency and outperformance. The method has been successfully applied to regularized risk minimization (Teo et al. 2010), machine learning (Le et al. 2008), two-stage stochastic linear problems (Ruszczynski 1986, Ruszczyanski and Swietanowski 1997), two-stage stochastic quadratic programming (Liu and Sen 2020), risk-averse two-stage stochastic linear programming (Miller and Ruszczyski 2011), and risk-averse multistage stochastic optimization (Ruszczyński 2010, Collado et al. 2012). Two recent specialized bundle methods for multistage stochastic programs are given in Asamov and Powell (2018), van Ackooij et al. (2019).

The 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 $\{\widehat{x}^j,\ \varphi(\widehat{x}^j),\ g^j \in \partial\varphi(\widehat{x}^j)\}_{j\in J_k}$ for some index set $J_k \subseteq \{1,\cdots,k\}$, the proximal bundle method constructs a piecewise-linear approximation of $\varphi$, denoted by $\varphi^k(x)$, in terms of $\widehat{x}^j$, $\varphi(\widehat{x}^j)$, and $g^j$. The approximate function $\varphi^k$ is used to construct a master optimization problem. Then, an optimal solution of the master problem $\widehat{x}^{j+1}$ is found, the objective $\varphi(\widehat{x}^{j+1})$ is evaluated, a subgradient in $\partial\varphi(\widehat{x}^{j+1})$ is obtained, and the algorithm continues into the next iteration.

This method requires evaluations of the objective function $\varphi(\widehat{x}^j)$ and consequently computing $\mathcal{Q}(\widehat{x}^j,\omega)$ for all $\omega\in\Omega$. This step involves solving $|\Omega|$ second-stage problems of the primal form (2) or its dual. This task 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 $\mathbb{E}[\mathcal{Q}(x,\omega)]$ 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 $\rho[\mathcal{Q}(x,\omega)]$, 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 $\mathbb{E}[\mathcal{Q}(x,\omega)]$, a number of extensions to bundle methods capable of working with less accurate function evaluations have been developed. For a recent review on algorithms based on the bundle method to handle inexact data, see de Oliveira and Solodov (2020). These methods replace the function values and subgradients by their approximations, for all or just a subset of the iterations. Suppose $\widetilde{\varphi}_{\widehat{x}^j}$ and $\widetilde{g}_{\widehat{x}_j}$ are objective value and subgradient estimates

obtained from an inexact oracle for the projection point $\widehat{x}^j$. Then, the inexact bundle method uses an approximate linearization for the first-stage objective function $\varphi$ constructed by these estimates.

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

$$\widetilde{\varphi}_x \in [\varphi(x) - \epsilon_1,\ \varphi(x) + \epsilon_2], \tag{3}$$

$$\widetilde{g}_x \in \partial_{\epsilon_0}\varphi(x). \tag{4}$$

Here, $\epsilon_1, \epsilon_2 \geq 0$ are unknown but fixed, $\epsilon_0 = \epsilon_1 + \epsilon_2$, and the $\epsilon_0$-approximate subdifferential $\partial_{\epsilon_0}\varphi(x)$ in (4) is given by

$$\partial_{\epsilon_0}\varphi(x) := \{g \in \mathbb{R}^n \mid \varphi(z) \geq \varphi(x) + \langle g, z - x\rangle - \epsilon_0, \forall z \in \mathcal{X}\}. \tag{5}$$

Therefore, to achieve convergence in this inexact bundle method, one needs to compute approximations $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ which satisfy equations (3) and (4). In addition, to address the original motivation of achieving a computationally efficient approach, these approximations must be easily computable.

This paper extends this inexact proximal bundle method (Kiwiel 2006, Oliveira et al. 2011) to the risk-averse two-stage stochastic optimization problems of the form (1) with the aversion to risk in the second-stage optimal value $\rho[\mathcal{Q}(x,\omega)]$. 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 inexact bundle method for convex first-stage feasible regions. The validity of the risk-averse inexact oracle under some assumption on the second-stage optimal value is theoretically established.

The developed inexact bundle method for our resource allocation model is studied using an extensive computational investigation. In particular, we focus on two coherent risk measures for the risk of unfulfilled demand. We show the convergence of the method for this two-stage stochastic optimization problem with a discrete first-stage feasible region numerically by comparing the computed solutions with those of the computationally prohibitive deterministic equivalent models.

The contributions of the paper can be summarized as follows.

- An oracle needed to implement the framework of inexact proximal 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 conditions (3)-(4).
- The approach with the introduced risk-averse inexact oracle is applied heuristically to a resource allocation problem arising in contingency planning. This step involves a heuristic treatment to adopt the method when the first-stage decision variables are subject to integrality constraints.

- We perform the benchmarks of the algorithm against the exact proximal 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 details of the inexact bundle method. Section 4 introduces the risk-averse oracle and proves its correctness. The modeling details for a resource allocation problem in contingency planning are explained in Section 5. Section 6 reports the results of the numerical experiments and the comparisons on the benchmark problems. We list our conclusions in Section 7.

## 2. Coherent Risk Measures and Risk-Averse Stochastic Optimization

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,a), Shapiro et al. (2014), Rockafellar (2014).

Let $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$ consist of all $\mathcal{F}$-measurable functions $Z : \omega \to \mathbb{R}$, where the set $\Omega := \{\omega_1, \cdots, \omega_N\}$ is finite with $N$ elements and $p_1, \cdots, p_N$ are probabilities of the corresponding elementary events. 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 exposition, $Z$ represents a random cost and as such smaller realizations are preferred.

**Definition 2.1** *A real-valued* coherent risk measure *is a proper function $\rho : \mathcal{Z} \to \mathbb{R}$ 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, e.g., see Theorem 6.4 in (Shapiro et al. 2014):

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

$$\rho(Z) = \max_{\mu \in \mathfrak{A}_\rho} \mathbb{E}_\mu[Z] = \max_{\mu \in \mathfrak{A}_\rho} \sum_{\omega \in \Omega} \mu_\omega z_\omega p_\omega, \quad \forall Z \in \mathcal{Z}, \tag{6}$$

*where* $\mathfrak{A}_\rho := \partial\rho(0) \subseteq \left\{\mu \in \mathbb{R}^{|\Omega|} \,\middle|\, \mu \geq 0 \text{ and } \sum_{\omega \in \Omega} \mu_\omega p_\omega = 1\right\}$.

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

$$\min_{x\in\mathcal{X}} \ \varphi(x) = \min_{x\in\mathcal{X}} \ \left\{c^\top x + \rho(\mathcal{Q}(x,\omega))\right\} = \min_{x\in\mathcal{X}} \ \left\{c^\top x + \max_{\mu\in\mathfrak{A}_\rho} \sum_{\omega\in\Omega} \mathcal{Q}(x,\omega)\mu_\omega p_\omega\right\}.$$

In the risk-averse two-stage optimization problem (1)–(2), $\mathcal{X}\subseteq\mathbb{R}^n$ and $\mathcal{Y}\subseteq\mathbb{R}^m$ are nonempty, convex, and closed. The second-stage problem is defined by $q:\mathbb{R}^m\times\Omega\to\mathbb{R}$, $T=(t_1,\cdots,t_\ell):\mathbb{R}^n\times\Omega\to\mathbb{R}^\ell$, and $R=(r_1,\cdots,r_\ell):\mathbb{R}^m\times\Omega\to\mathbb{R}^\ell$, where $t_i(x,\omega)$ and $r_i(y,\omega)$, $i=1,\cdots,\ell$ are real-valued. We assume that the real-valued functions $c(x)$ and $q(\cdot,\omega)$, and the mappings $T(\cdot,\omega)$ and $W(\cdot,\omega)$ are proper convex for every $\omega\in\Omega$. 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 stochastic problem has complete recourse, e.g., see Birge and Louveaux (1997), Shapiro et al. (2014). For any first-stage decision $x$, denote the $\ell$-vector $\chi:=T(x,\omega)$ and let $\vartheta(\chi,\omega)$ denote the optimal value of the second-stage problem, i.e., $\vartheta(\chi,\omega):=\mathcal{Q}(x,\omega)$. We assume that the regularity condition $\chi\in\mathrm{int}(\mathrm{dom}\vartheta(\cdot,\omega))$ holds, i.e., for all small perturbations of $\chi$, the second-stage problem remains feasible. We further assume that the functions $c(\cdot)$ and $T_\omega(\cdot)=T(\cdot,\omega)$ are differentiable in $x$ for every $\omega\in\Omega$, and $0\in\mathrm{int}\left\{T_\omega(x)+\nabla_x T_\omega(x)\mathbb{R}^\ell-\mathrm{dom}\,\vartheta(\cdot,\omega)\right\}$. Our intention is to describe $\partial_x\varphi(x)$, the subdifferential of $\varphi(\cdot)$ evaluated at $x$.

For $x\in\mathcal{X}$, denote $\phi(x):=\rho(\mathcal{Q}(x,\omega))$. Theorem 2.1 implies that

$$\phi(x) := \max_{\mu\in\mathfrak{A}_\rho} \sum_{\omega\in\Omega} \mu_\omega p_\omega \mathcal{Q}(x,\omega). \tag{7}$$

Let $\mathcal{Z}^*:=\mathcal{L}_\infty(\Omega,\mathcal{F},P)$ and define $\widetilde{\mathcal{Q}}:\mathcal{X}\times\mathcal{Z}^*\to\mathbb{R}$, where

$$\widetilde{\mathcal{Q}}(x,\mu) := \sum_{\omega\in\Omega} p_\omega\mu_\omega\mathcal{Q}(x,\omega). \tag{8}$$

Using this notation in equation (7) implies that $\phi(x)=\max_{\mu\in\mathfrak{A}_\rho}\widetilde{\mathcal{Q}}(x,\mu)$. It thus follows from Theorem 6.11 in Shapiro et al. (2014) (see also Theorem 2.87 in Ruszczynski (2006)) that for real-valued coherent risk measures $\phi(x)$ is subdifferentiable and

$$\partial\phi(x) = \mathrm{conv}\left[\bigcup_{\mu\in\widehat{\mathfrak{A}}_x} \partial_x\widetilde{\mathcal{Q}}(x,\mu)\right], \tag{9}$$

where $\widehat{\mathfrak{A}}_x := \{\mu^*\in\mathfrak{A}_\rho \mid \widetilde{\mathcal{Q}}(x,\mu^*)=\phi(x)\}$. In particular, for any $\mu^*\in\widehat{\mathfrak{A}}_x$, we have

$$\partial_x\widetilde{\mathcal{Q}}(x,\mu^*)\subseteq\partial\phi(x). \tag{10}$$

The rest of this section aims to specify the subdifferential $\partial_x\widetilde{\mathcal{Q}}(x,\mu)$. It follows from the Moureau-Rockafeller Theorem (see Theorem 6 in Ruszczyński and Shapiro (2003)) that

$$\partial_x\widetilde{\mathcal{Q}}(x,\mu) = \sum_{\omega\in\Omega}\mu_\omega p_\omega\partial_x\mathcal{Q}(x,\omega). \tag{11}$$

Propositions 2.21 and 2.22 in Shapiro et al. (2014) imply that the function $\mathcal{Q}(\cdot,\omega)$ is convex and $\partial_x \mathcal{Q}(x,\omega) = \nabla_x T(x,\omega)^\top \mathfrak{D}(\chi,\omega)$, where $\mathfrak{D}(\chi,\omega)$ is the set of optimal solutions of the dual problem of the second-stage problem. In particular, the function $\mathcal{Q}(\cdot,\omega)$ is differentiable at every $x$ at which $\mathfrak{D}(\chi,\omega)$ is a singleton, e.g., see Corollary 2.23 in Shapiro et al. (2014). This result along with equation (11) imply that

$$\partial_x \widetilde{\mathcal{Q}}(x,\mu) = \sum_{\omega\in\Omega} \mu_\omega p_\omega \nabla_x T(x,\omega)^\top \mathfrak{D}(\chi,\omega). \tag{12}$$

Then equation (12) at $\mu^*$ together with equation (10) yields

$$\zeta_x := \sum_{\omega\in\Omega} \mu^*_\omega p_\omega \nabla_x T(x,\omega)^\top \pi_\omega \in \partial\phi(x), \tag{13}$$

for $\pi_\omega \in \mathfrak{D}(\chi,\omega)$. Since the function $\mathcal{Q}(\cdot,\mu)$ is convex for any given $\mu$, the function $\phi(x)$ is also convex. Therefore, the subgradient calculus for convex functions implies that $g_x := c + \zeta_x \in \partial\varphi(x)$.

Evaluating the objective function $\varphi(x)$ at a given $x \in \mathcal{X}$ requires computing $\phi(x)$ in equation (7) and thus the optimal objective value of the second-stage problem, $\mathcal{Q}(x,\omega)$, for all $\omega \in \Omega$. Each $\mathcal{Q}(x,\omega)$ can be computed by solving the primal form of the problem (2). However, the regularity condition and the complete recourse assumptions of the second-stage problem imply that the strong primal-dual optimality holds for the problem (2), see Proposition 25 in Ruszczyński and Shapiro (2003). Therefore,

$$\mathcal{Q}(x,\omega) = \max_{\pi\in\mathbb{R}^\ell_+} \left\{ \pi^\top \chi + \inf_{y\in\mathcal{Y}} \mathcal{L}(y,\pi,\omega) \right\}, \tag{14}$$

where $\mathcal{L}(y,\pi,\omega) := q(y,\omega) + \pi^\top R(y,\omega)$. This suggests that one can use the dual form of the second-stage problem in (14) to derive $\mathcal{Q}(x,\omega)$ and consequently evaluate $\varphi(x)$. Our risk-averse inexact oracle builds on this dual representation.

## 3. Inexact Proximal 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. The class of cutting plane methods, in particular bundle methods, proved to be a useful approach to solve risk-averse optimization problems. For details on the exact bundle method and its convergence analysis for minimizing a nonsmooth convex function, see (Ruszczynski 2006, ch. 7.4). For the case of two-stage stochastic optimization, see (Birge and Louveaux 1997, ch. 5.2).

The essence of the (exact) proximal bundle method includes the application of the Moreau-Yosida regularization of a lower approximation of the objective function and solving a sequence of quadratic optimization problems. Localizing the iterations through regularization makes the bundle

method more practical for problems of higher dimensions, where methods such as the cutting plane method need a growing number of cuts to be stored in the master problem (Ruszczynski 2006).

We apply the bundle method to the first-stage minimization problem (1), $\min_{x\in\mathcal{X}} \varphi(x)$. At iteration $k$ of the proximal bundle method, having points $\widehat{x}^1, \cdots, \widehat{x}^k$, a piecewise-linear approximation $\varphi^k(\cdot)$ of the objective function $\varphi(\cdot)$ is constructed and used in the following *master problem*:

$$\min_{x\in\mathcal{X}} \ \varphi^k(x) + \frac{\gamma_k}{2}\left\|x-\beta^k\right\|^2. \tag{15}$$

Here, in the proximal term $\frac{\gamma_k}{2}\left\|x-\beta^k\right\|^2$, $\gamma_k > 0$ is a regularization parameter and the prox center $\beta^k \in \mathcal{X}$ is updated by a conditional rule over iterations. An optimal solution of the master problem (15), denoted by $\widehat{x}_{k+1}$, is added to the set of points and used to construct improved linear approximations over next iterations.

In the exact version of the method, values of the objective function $\varphi(\widehat{x}^1), \cdots, \varphi(\widehat{x}^k)$, and subgradients $g^1 \in \partial\varphi(\widehat{x}^1), \cdots, g^k \in \partial\varphi(\widehat{x}^k)$ are used to construct the approximate model $\varphi^k(\cdot)$ as follows:

$$\varphi^k(x) := \max_{j\in J_k} \left\{\varphi(\widehat{x}^j) + \left\langle g^j,\, x-\widehat{x}^j\right\rangle\right\}, \tag{16}$$

for some subset $J_k \subseteq \{1,\cdots,k\}$. For the risk-averse two-stage stochastic optimization problem (1)–(2), evaluating $\varphi(\widehat{x})$ and computing subgradients $g \in \partial\varphi(\widehat{x})$ are carried out by means of equations (7) and (13), which involve computing $\mathcal{Q}(x,\omega)$ by solving the second-stage problem for all $\omega \in \Omega$. This is a computationally expensive task, especially as the number of elementary elements in the probability space $N$ grows.

To mitigate this computational challenge, one can resort to inexact bundle methods, which rely on approximations of the objective and subgradient values, and try to supply admissible approximations by solving some form of the second-stage problem for only a subset of scenarios. We adopt the inexact bundle method in Kiwiel (2006). In this inexact proximal bundle method, approximate objective value $\widetilde{\varphi}_{\widehat{x}^j}$ and approximate subgradient $\widetilde{g}_{\widehat{x}^j}$ satisfying the conditions in equations (3) and (4) are derived to construct the piecewise-linear approximation $\varphi^k(x)$ as follows:

$$\varphi^k(x) = \max_{j\in J_k} \left\{\widetilde{\varphi}_{\widehat{x}^j} + \left\langle \widetilde{g}_{\widehat{x}^j},\, x-\widehat{x}^j\right\rangle\right\}. \tag{17}$$

Then, at iteration $k$, the model approximation $\varphi^k(x)$ is used in the master problem (15). With the regularization parameter $\gamma_k = \frac{1}{t_k}$, this master problem can be equivalently written as a linearly constrained quadratic optimization problem:

$$\begin{aligned} \min_{x\in\mathcal{X},\ \upsilon\in\mathbb{R}} \ & \upsilon + \frac{1}{2t_k}\left\|x-\beta^k\right\|^2 \\ \text{s.t. } & \widetilde{\varphi}_{\widehat{x}^j} + \left\langle \widetilde{g}_{\widehat{x}^j},\, x-\widehat{x}^j\right\rangle \le \upsilon, \qquad \forall\ j \in J_k. \end{aligned} \tag{18}$$

The parameter $t_k > 0$, referred to as the *stepsize*, as well as the prox center $\beta^k$ are updated during iterations (Kiwiel 1990, 2006). Algorithm 1 presents the details of this method (see Algorithm 2.1 in Kiwiel (2006)). This algorithm assumes the availability of an oracle which returns admissible approximations $\widetilde{\varphi}_{\widehat{x}^j}$ and $\widetilde{g}_{\widehat{x}^j}$ satisfying the conditions (3) and (4).

**Algorithm 1** Inexact Proximal Bundle Method (Kiwiel 2006)

Inputs: descent parameter $\varrho \in (0,1)$, stepsize bound $T_1 > 0$, stepsize $t_1 \in (0, T_1]$, stopping tolerance $\delta > 0$, an inexact oracle satisfying conditions (3) and (4).

Step 0: *(Initialization)*

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

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

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

Step 1: *(Trial point finding)* Let $(\widehat{x}^{k+1}, v^{k+1})$ be an optimal solution of problem (18), and $\{\lambda_j^k\}_{j \in J_k}$ be the Lagrangian multipliers. Compute the *aggregate subgradient* $p_k := \dfrac{1}{t_k}(\beta^k - \widehat{x}^{k+1})$, the *predicted descent* $\nu_k := \widetilde{\varphi}_{\beta^k} - v^{k+1}$, and the *aggregate linearization error* $\alpha_k := \nu_k - t_k \|p_k\|^2$.

Step 2: *(Stopping criterion)* Compute the *optimality measure* $\varpi_k := \max\{\|p_k\|,\ \alpha_k\}$. If $\varpi_k \le \delta$, stop.

Step 3: *(Stepsize correction)* If $\nu_k < -\alpha_k$, set $t_k := 10\, t_k$, $T_k := \max\{T_k, t_k\}$, $i_k := k$ and go to Step 1; else set $T_{k+1} := T_k$.

Step 4: *(Inexact oracle call)* Obtain $\widetilde{\varphi}_{\widehat{x}^{k+1}}$ and $\widetilde{g}_{\widehat{x}^{k+1}}$ from the inexact oracle satisfying (3) and (4).

Step 5: *(Descent test)*

If $\widetilde{\varphi}_{\widehat{x}^{k+1}} \le \widetilde{\varphi}_{\beta^k} - \varrho\nu_k$, declare a *descent step*: set $\beta^{k+1} := \widehat{x}^{k+1}$, $\ell := \ell + 1$, $i_{k+1} := 0$, $k(\ell) := k + 1$.

If $\widetilde{\varphi}_{\widehat{x}^{k+1}} > \widetilde{\varphi}_{\beta^k} - \varrho\nu_k$, declare a *null step*: set $\beta^{k+1} := \beta^k$, $i_{k+1} = i_k$.

Step 6: *(Bundle management)* Choose $J_{k+1} \supseteq \{j \in J_k \mid \lambda_j^k \neq 0\} \cup \{k+1\}$.

Step 7: *(Stepsize updating)* If $k(\ell) = k+1$ (i.e., after a descent step), select $t_{k+1} \in [t_k, T_{k+1}]$. Otherwise, if the step was declared null, set $t_{k+1} := t_k$. If $i_{k+1} = 0$ and $\varpi_k \le \widetilde{\varphi}_{\beta^k} - (\widetilde{\varphi}_{\widehat{x}^{k+1}} + \langle \widetilde{g}_{\widehat{x}^{k+1}},\, \beta^k - \widehat{x}^{k+1} \rangle)$, choose $t_{k+1} \in [0.1\, t_k, t_k]$. Set $k := k+1$ and go to Step 1.

Step 1 sets $\widehat{x}^{k+1}$ and $v^{k+1} = \varphi^k(\widehat{x}^{k+1})$ to be an optimal solution to the master problem (18) using the approximate piecewise-linear model $\varphi^k$ constructed by $\{\widehat{x}^j,\ \widetilde{\varphi}_{\widehat{x}^j},\ \widetilde{g}_{\widehat{x}^j}\}_{j \in J_k}$. Due to errors in the oracle approximations, $\varphi^k$ might not approximate $\widetilde{\varphi}$ from below at the prox center $\beta^k$, in which case the predicted descent $\nu_k$ may be nonpositive. In such a case, the method increases the stepsize $t_k$ and loops over a stepsize correction phase in Step 3 until $\nu_k > 0$. Note that after the stepsize correction step, $\nu_k := \widetilde{\varphi}_{\beta^k} - v^{k+1} = \widetilde{\varphi}_{\beta^k} - \varphi^k(\widehat{x}^{k+1}) > 0$. In the descent test at Step 5, a null step improves the model approximation (17) through the addition of an extra constraint to problem (18). In Step 6, the Lagrangian multipliers $\lambda_j^k$ are used to update the index set with the

aim of reducing the number of cuts when constructing 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 from equation (17). Lagrangian multipliers $\lambda_j^k$ are only necessary for bundle management, Step 6, and can be forgone at the cost of solving a larger master problem at each iteration.

With the tolerance level $\delta = 0$, the inexact bundle method has two possible outcomes (see Theorem 9 in Oliveira et al. (2011)): Either (i) the method loops forever at Step 3, in which case the last generated $\beta^k$ is $2(\epsilon_1 + \epsilon_2)$-optimal, or (ii) the method generates an infinite sequence of either descent or null steps. In this case the method generates a sequence $\{\beta^k\}_{k=1}^{\infty}$ for which each cluster point $\beta^*$ satisfies $\beta^* \in \mathcal{X}$ and $\beta^*$ is $2(\epsilon_1 + \epsilon_2)$-optimal.

This inexact proximal bundle method can find a $\epsilon_0$-optimal solution for the convex minimization problem over a nonempty closed convex set. We refer the reader to Kiwiel (2006) and de Oliveira et al. (2014) for a proof and an in-depth discussion on its convergence analysis.

## 4. Risk-Averse Inexact Oracle

A key component in Algorithm 1 is the definition of an inexact oracle capable of providing approximate value $\widetilde{\varphi}_x$ and approximate subgradient $\widetilde{g}_x$ satisfying (3)–(4). In this section, we introduce an inexact oracle specialized to work on our two-stage risk-averse stochastic optimization problem. We refer to this oracle as the *risk-averse inexact oracle*, which together with Algorithm 1 comprises the *risk-averse inexact bundle method*. The construction of our risk-averse inexact oracle extends 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 admits the form

$$\mathcal{Q}(x,\omega) = \max_{\pi \in \mathbb{R}^\ell} \left\{ \pi^\top \eta(x,\omega) \mid \pi \in \Pi(\omega) \right\}, \tag{19}$$

for some function $\eta : \mathcal{X} \times \Omega \to \mathbb{R}^\ell$ convex and differentiable in $x$, and a nonempty convex set $\Pi(\omega) \subseteq \mathbb{R}^\ell$. The feasible region $\Pi(\omega)$ can be expressed in the general form $\{\pi \in \mathbb{R}^\ell \mid c_i(\pi,\omega) \leq 0,\ i = 1, \cdots, l\}$, where $c_i(\pi,\omega)$s are convex in $\pi$.

Denote an optimal solution of the maximization problem (19) by $\pi_\omega$. Given the optimality of $\pi_\omega$ and expression (19), we have $\mathcal{Q}(x,\omega) = \pi_\omega^\top \eta(x,\omega)$. The essence of the inexact oracle is to compute approximate values for $\{\mathcal{Q}(x,\omega)\}_{\omega\in\Omega}$, without computing $\pi_\omega$ for every single $\omega \in \Omega$, and then use these values to construct approximations for $\varphi(x)$ and $\partial\varphi(x)$. This process is carried out by first selecting a subset $\mathcal{I} \subseteq \Omega$ and cluster the scenarios. The approximate oracle then computes the exact optimal values $\mathcal{Q}(x,\omega)$ for every $\omega \in \mathcal{I}$ and derives approximate optimal values, without solving the second-stage problem, for the remaining scenarios $\omega \in \Omega \setminus \mathcal{I}$.

Suppose $\mathcal{I} := \{\widehat{w}_1, \cdots, \widehat{w}_K\}$ is a subset of scenarios $\mathcal{I} \subseteq \Omega$ obtained through a clustering procedure for the scenario set $\Omega$. Therefore, $\Omega = \cup_{\widehat{\omega}_k \in \mathcal{I}} \mathcal{J}_k$, where $\mathcal{J}_k$ denotes the cluster of scenarios around the scenario $\widehat{\omega}_k \in \mathcal{I}$. Given $\mathcal{I}$ and for any $k = 1, \cdots, K$, define

$$\widehat{\Pi}_k := \left\{ \pi \in \mathbb{R}^\ell \mid \mathbb{E}_\omega \left[ c_i(\pi, \omega) | \omega \in \mathcal{J}_k \right] \leq 0,\ i = 1, \cdots, l \right\}. \tag{20}$$

Here, the expectation is given by $\mathbb{E}_\omega \left[ c_i(\pi, \omega) | \omega \in \mathcal{J}_k \right] = \sum_{\omega \in \mathcal{J}_k} p_\omega c_i(\pi, \omega) / \sum_{\omega \in \mathcal{J}_k} p_\omega$. For any $\omega \in \mathcal{I}$, consider the problem of maximizing $\pi^\top \eta(x, \omega)$ over the feasible set $\widehat{\Pi}_k$, defined in (20), and denote its optimal solution by $\widehat{\pi}_k$. Hence, for $k = 1, \cdots, K$,

$$\widehat{\pi}_k \in \arg\max_{\pi \in \mathbb{R}^\ell} \left\{ \pi^\top \eta(x, \widehat{\omega}_k) \mid \pi \in \widehat{\Pi}_k \right\}. \tag{21}$$

Note that $\widehat{\pi}_k$ depends on $x$. For a scenario $\omega \in \Omega$, let $\widehat{k}(w) \in \{1, \cdots, K\}$ denote the index of the corresponding cluster $\mathcal{J}_k$ with $\omega \in \mathcal{J}_k$. In particular, when $\omega = \widehat{\omega}_k \in \mathcal{I}$, we have $\widehat{k}(w) = k$. Then, for any $\omega \in \Omega$, the algorithm adopts an approximation of $\mathcal{Q}(x, \omega)$ given by

$$\mathcal{Q}^{\text{approx}}(x, \omega) := \widehat{\pi}_{\widehat{k}(\omega)}^\top \eta(x, \omega). \tag{22}$$

These computed values are then aggregated to form the objective function approximation $\widetilde{\varphi}_x$,

$$\widetilde{\varphi}_x := c^\top x + \max_{\mu \in \mathfrak{A}_\rho} \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}^{\text{approx}}(x, \omega) \right], \tag{23}$$

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

$$\widetilde{g}_x := c + \left[ \sum_{\omega \in \Omega} \mu^*_\omega p_\omega \widehat{\pi}_{\widehat{k}(\omega)}^\top \nabla_x \eta(x, \omega) \right], \tag{24}$$

where, $\mu^*$ is an optimal solution to the maximization problem in (23). Algorithm 2 formally states the risk-averse inexact oracle. Later in Theorem 4.1, we prove that under some conditions on problem (19) and $\mathcal{I}$, the computed $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ in equations (23) and (59) satisfy the conditions (3) and (4); hence the oracle returns admissible inputs for the Algorithm 1.

**Algorithm 2** Risk-Averse Inexact Oracle

*Inputs:* $x \in \mathcal{X}$, $\Omega = \{\omega_1, \cdots, \omega_N\}$, *a scenario clustering method* $\mathcal{J}(x)$

Step 1: (Scenario Clustering) *For the given* $x$, *call the scenario clustering method* $\mathcal{J}(x)$ *to generate clusters* $\{\mathcal{J}_1, \cdots, \mathcal{J}_K\}$ *with the set of cluster centers* $\mathcal{I}$.

Step 2: (Cluster Solutions) *For each cluster* $\mathcal{J}_k$, *construct* $\widehat{\Pi}_k$ *in* (20). *Compute a solution* $\widehat{\pi}_k$ *as in* (21).

Step 3: (Approximate $\mathcal{Q}^{\text{approx}}$) *For each* $\omega \in \Omega$, *find the cluster* $\widehat{k}(\omega)$ *containing* $\omega$. *Using solutions* $\widehat{\pi}_k$ *obtained in* Step 2, *compute* $\mathcal{Q}^{approx}(x, \omega)$.

Step 4: (Risk-averse oracle estimates) *Use* $\mathcal{Q}^{approx}$ *obtained from* Step 3 *in equation* (23) *to compute* $\widetilde{\varphi}_x$. *Compute an optimal solution* $\mu^*$ *to the maximization in equation* (23). *Apply* $\mu^*$ *in equation* (59) *to derive the subgradient approximation* $\widetilde{g}_x$.

---

Next, we show that the risk-averse inexact oracle, presented in Algorithm 2, provides approximations $\widetilde{\varphi}_x$ and $\widetilde{g}_x$ satisfying conditions (3) and (4). This property is referred to as the correctness of the risk-averse inexact oracle.

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

$$s_{\mathcal{S}}(d) := \max\left\{\pi^\top d \;\middle|\; \pi \in \mathcal{S}\right\}. \tag{25}$$

Denote $d_\omega^x := \eta(x,\omega)$. Let $\|d_\omega^x\| \le \vartheta_k^x$ for all $\omega \in \mathcal{J}_k$, for some $\vartheta_k^x < +\infty$. The following assumptions are made on the structure in (19):

[A] For any $x \in \mathcal{X}$ and any cluster $k \in \{1, \cdots, K\}$, there exists a constant $\Gamma_k^x > 0$ such that

$$|s_{\Pi(\omega)}(d_\omega^x) - s_{\widehat{\Pi}_k}(d_\omega^x)| \le \vartheta_k^x \Gamma_k^x, \quad \forall\, \omega \in \mathcal{J}_k.$$

[B] For any $x \in \mathcal{X}$ and any cluster $k \in \{1, \cdots, K\}$, there exists a constant $\kappa_k^x > 0$ such that

$$|s_{\widehat{\Pi}_k}(d_\omega^x) - \widehat{\pi}_k^\top d_\omega^x| \le \vartheta_k^x \kappa_k^x, \quad \forall\, \omega \in \mathcal{J}_k.$$

Since $\widehat{\pi}_{\widehat{k}(\omega)} \in \widehat{\Pi}_{\widehat{k}(\omega)}$, it is a feasible point when maximizing $\pi^\top d_w^x$. Hence, $s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d_\omega^x) \ge \widehat{\pi}_{\widehat{k}(\omega)}^\top d_\omega^x$, and consequently $|s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d_\omega^x) - \widehat{\pi}_{\widehat{k}(\omega)}^\top d_\omega^x| = |s_{\widehat{\Pi}_k}(d_\omega^x) - \widehat{\pi}_k^\top d_\omega^x| = s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d_\omega^x) - \widehat{\pi}_{\widehat{k}(\omega)}^\top d_\omega^x$. In addition, note that the definition of $\widehat{\pi}_k$ in equation (21), for the cluster center $w = \widehat{w}_k$, we have $s_{\widehat{\Pi}_k}(d_\omega^x) = \widehat{\pi}_k^\top d_\omega^x$, and consequently $|s_{\widehat{\Pi}_k}(d_\omega^x) - \widehat{\pi}_k^\top d_\omega^x| = 0$.

Notice that the parameters $\Gamma_\omega^x$ and $\kappa_\omega^x$, and consequently $\epsilon^*$ defined below, do depend on the set of clusters $\mathcal{I}$ and consequently the clustering mechanism through the collinearity measure (26).

**Theorem 4.1** *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* (19) *for some function* $\eta$. *Then, for every* $x \in \mathcal{X}$ *and* $\epsilon_{cos} \in (0,1)$, Algorithm 2 *along with the scenario clustering method in* Algorithm 3 *provides outputs* $\widetilde{\varphi}_x$ *and* $\widetilde{g}_x$ *satisfying equations* (3) *and* (4) *with* $\epsilon_1 = \epsilon_2 = \epsilon^* > 0$, *where* $\epsilon^* := \max_{x\in\mathcal{X}, k=1,\cdots,K}\{(\Gamma_k^x + \kappa_k^x)\vartheta_k^x\}$.

The proof of Theorem 4.1 is presented in subsection 4.3.

### 4.1. Scenario Clustering

An approach for scenario clustering in Step 1 of Algorithm 2 is to select the subset $\mathcal{I}$ such that the corresponding set of vectors $\{\eta(x,\xi)\,|\,\xi\in\mathcal{I}\}$ sufficiently deviates from collinearity. The collinearity of two scenarios $\omega$ and $\xi$ is measured by the cosine of the angle $\theta_{\omega,\xi}$ between the two vectors $\eta(x,\omega)$ and $\eta(x,\xi)$, namely

$$\theta_{\omega,\xi} := \cos^{-1}\left(\frac{\eta(x,\omega)^\top \eta(x,\xi)}{\|\eta(x,\omega)\|\,\|\eta(x,\xi)\|}\right). \tag{26}$$

Hence, two scenarios $\omega$ and $\xi$ are collinear if $\cos\theta_{\omega,\xi}=1$. 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 $\cup_{\omega\in\mathcal{I}}\mathcal{J}_\omega=\Omega$ and for every $\omega,\xi\in\mathcal{I}$ we have $\cos\theta_{\omega,\xi}\le 1-\epsilon_{\cos}$. Each cluster then includes all $\omega$-almost collinear scenarios, for some $\omega\in\mathcal{I}$. This can be carried out by a combinatorial method given in Oliveira et al. (2011). This method is presented in Algorithm 3. The generated subset $\mathcal{I}\subseteq\Omega$ from Algorithm 3 depends on the permutation of elements of $\Omega$ fixed as an input. In particular, the initial elements of the permutation of $\Omega$ are favored to be part of $\mathcal{I}$. We consider a random permutation of elements of $\Omega$ on each call to Algorithm 3.

**Algorithm 3** Selection of $\mathcal{I}\subseteq\Omega$ and clusters $\{\mathcal{J}_k\}_k$

*Inputs:* $x\in\mathcal{X}$, $\Omega=\{\omega_1,\ldots,\omega_N\}$, *collinearity parameter* $\epsilon_{cos}\in(0,1)$

Step 0: (Initialization) *Set* $k=0$, $\mathcal{I}=\emptyset$, *and* $\mathcal{I}_0:=\Omega$

Step 1: (Main procedure) *For every* $i=1,\cdots,N$ *such that* $\omega_i\in\mathcal{I}_0$ *do:*

$k:=k+1$, $\mathcal{I}=\mathcal{I}\cup\{\omega_i\}$, and define $\mathcal{J}_k:=\{\omega_i\}$

**for** every $j=i+1,\cdots,N$ such that $\omega_j\in\mathcal{I}_0$ **do**

compute $\cos\theta_{\omega_i,\omega_j}$ using vectors $\eta(x,\omega_i)$ and $\eta(x,\omega_j)$ and eq. (26)

If $\cos\theta_{\omega_i,\omega_j}>1-\epsilon_{\cos}$ then $\mathcal{J}_k:=\mathcal{J}_k\cup\{\omega_j\}$ and $\mathcal{I}_0:=\mathcal{I}_0\setminus\{\omega_j\}$

**end for**

### 4.2. The Case of Linear Second-Stage Problem

The second-stage optimal value $\mathcal{Q}(x,\omega)$ takes the form in equation (19), for example, when the second-stage problems in a linear optimization problem for any $\omega\in\Omega$, i.e., $q(y,\omega)=q_\omega^\top y$ for some $m$-vector $q_\omega$, $R(y,\omega)=Ry-h_\omega$ for a fixed recourse matrix $R$, and $\mathcal{Y}=\mathbb{R}^m_+$:

$$\begin{aligned}\mathcal{Q}(x,\omega)=\min_{y\in\mathbb{R}^m}\quad & q_\omega^\top y\\ \text{s.t.}\quad & Ry+T(x,\omega)\le h_\omega,\qquad y\ge 0,\end{aligned} \tag{27}$$

The full recourse assumption on the second-stage problem implies that the strong duality for problem (27) holds and we have

$$\begin{aligned}\mathcal{Q}(x,\omega)=\max_{\pi\in\mathbb{R}^\ell}\quad & \pi^\top\left(T(x,\omega)-h_\omega\right)\\ \text{s.t.}\quad & R^\top\pi\ge -q_\omega,\qquad \pi\ge 0.\end{aligned} \tag{28}$$

The expression of problem (28) if of the form in (19) with $\eta(x,\omega) = T(x,\omega) - h_\omega$ is convex and differentiable in $x$, and the set $\Pi(\omega) := \{\pi \in \mathbb{R}^\ell \big| -R^\top \pi - q_\omega \le 0,\ -\pi \le 0\}$ is convex.

For the second-stage problem (27), the feasible set $\widehat{\Pi}_k$ defined in equation (20) is given by $\widehat{\Pi}_k = \{\pi \in \mathbb{R}^\ell \mid \pi \ge 0,\ R^\top \pi \ge -\mathbb{E}_\omega[q_\omega | \omega \in \mathcal{J}_k]\}$, with the conditional expectation $\widehat{q}_\omega := \mathbb{E}_\omega[q_\omega|\omega \in \mathcal{J}_k] = \sum_{\xi\in\mathcal{J}_\omega} p_\xi q_\xi / \sum_{\xi\in\mathcal{J}_\omega} p_\xi$.

For the linear second-stage optimization problem (27) both assumptions [A] and [B] hold. We use the Lipschitz continuity of solutions of linear programs with respect to the right-hand-side perturbation; for the linear optimization problem (28) the optimal solution is Lipschitz with respect to perturbations in the right-hand-side vector $q_\omega$ (e.g., see Theorem 2.4 of Mangasarian and Shiau (1987)). Consider the $k$th cluster $\mathcal{J}_k$. Suppose $\max_{\pi\in\widehat{\Pi}_{\widehat{k}(\omega)}} \pi^\top d^x_\omega = \pi_k^\top d^x_\omega$. For all $\omega \in \mathcal{J}_k$, $\widehat{k}(\omega) = k$, we have

$$\begin{aligned}\left| s_{\Pi(\omega)}(d^x_\omega) - s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^x_\omega)\right| &= |\max_{\pi\in\Pi(\omega)} \pi^\top d^x_\omega - \max_{\pi\in\widehat{\Pi}_{\widehat{k}(\omega)}} \pi^\top d^x_\omega| = |\pi_\omega^\top d^x_\omega - \pi_k^\top d^x_\omega| \\ &\le \|\pi_\omega - \pi_k\| \|d^x_\omega\| \le \varsigma^x_\omega \|q_\omega - \mathbb{E}_\xi[q_\xi|\xi\in\mathcal{J}_k]\| \, \|d^x_\omega\| \le \Gamma^x_k \vartheta^x_k, \end{aligned} \tag{29}$$

for some Lipschitz constant $\varsigma^x_\omega > 0$ and $\Gamma^x_k := \max_{\omega\in\mathcal{J}_k}\{\varsigma^x_\omega \|q_\omega - \mathbb{E}_\xi[q_\xi|\xi\in\mathcal{J}_k]\|\}$. This establishes the validity of assumption [A] for linear second-stage models. In the special case, that $q_\omega = q$ for all $\omega \in \Omega$, then $\Gamma^x_k = 0$. In this case, $\widehat{\pi}_k = \pi_{\widehat{\omega}_k}$, the exact optimal solution of the second-stage problem for the scenario $\widehat{\omega}_k$.

Assumption [B] also holds when the second-stage problem is linear, as in (28). Consider cluster $\mathcal{J}_k$ and let $\omega \in \mathcal{J}_k$. Using the property $\max_{x\in X} f(x) - \max_{x\in X} g(x) \le \max_x (f(x) - g(x))$, we have

$$|s_{\widehat{\Pi}_k}(d^x_\omega) - \widehat{\pi}_k^\top d^x_\omega| = s_{\widehat{\Pi}_k}(d^x_\omega) - \widehat{\pi}_k^\top d^x_\omega = \max_{\pi\in\widehat{\Pi}_k} \pi^\top d^x_\omega - \max_{\pi\in\widehat{\Pi}_k} \pi^\top d^x_{\widehat{\omega}_k} \le \max_{\pi\in\widehat{\Pi}_k} \pi^\top (d^x_\omega - d^x_{\widehat{\omega}_k}) = \bar{\pi}^\top (d^x_\omega - d^x_{\widehat{\omega}_k}),$$

where $\check{\pi}_k$ is an optimal solution of the maximization problem $\max_{\pi\in\widehat{\Pi}_k} \pi^\top (d^x_\omega - d^x_{\widehat{\omega}_k})$. Since the feasible region of the problem (28) and consequently $\widehat{\Pi}_k$ is a polyhedron, which is nonempty due to the complete recourse assumption. Thus, we can let $\bar{\pi}_k$ be a basic optimal solution. Since any polyhedron has finitely many extreme points, let $\bar{\vartheta}_k$ be the largest $\|\bar{\pi}_k\|$. On the other hand we have $\|d^x_\omega - d^x_{\widehat{\omega}_k}\|^2 = (d^x_\omega - d^x_{\widehat{\omega}_k})^\top (d^x_\omega - d^x_{\widehat{\omega}_k}) = \|d^x_\omega\|^2 + \|d^x_{\widehat{\omega}_k}\|^2 - 2\|d^x_\omega\|\|d^x_{\widehat{\omega}_k}\| \cos(\theta_{\omega,\widehat{\omega}_k}) \le (\vartheta^x_k)^2 + (\vartheta^x_k)^2 - 2(\vartheta^x_k)^2 \epsilon_{\cos} = (\vartheta^x_k)^2 2(1-\epsilon_{\cos})$. These two bounds together imply that

$$|s_{\widehat{\Pi}_k}(d^x_\omega) - \widehat{\pi}_k^\top d^x_\omega| \le \|\check{\pi}_k\| \|d^x_\omega - d^x_{\widehat{\omega}_k}\| \le \bar{\vartheta}_k \vartheta^x_k \sqrt{2(1-\epsilon_{\cos})} \tag{30}$$

Thus, assumption [B] holds for $\kappa^x_\omega := \bar{\vartheta}_k \sqrt{2(1-\epsilon_{\cos})}$.

### 4.3. Proof of Theorem 4.1

In this section, we present the proof of Theorem 4.1.

**Proof of Theorem 4.1** Fix $x \in \mathcal{X}$. According to equation (19) and using the notation in (25), we have $\mathcal{Q}(x,\omega) = s_{\Pi(\omega)}(d^x_\omega)$. Equation (22) implies that $\mathcal{Q}^{\text{approx}}(x,\omega) := \widehat{\pi}^\top_{\widehat{k}(\omega)} d^x_\omega$, where $\widehat{\pi}_{\widehat{k}(\omega)}$ is as in (21) for $k = \widehat{k}(\omega)$. In particular, when $\omega = \widehat{\omega}_k$ for some $\widehat{\omega}_k \in \mathcal{I}$, we have $\mathcal{Q}^{\text{approx}}(x,\omega) := s_{\widehat{\Pi}_k}(d^x_\omega)$. Denote the approximation error in the second-stage value function evaluation by

$$\epsilon_\omega := \mathcal{Q}(x,\omega) - \mathcal{Q}^{\text{approx}}(x,\omega) = \begin{cases} s_{\Pi(\omega)}(d^x_\omega) - s_{\widehat{\Pi}_k}(d^x_\omega), & \text{if } \omega = \widehat{w}_k \in \mathcal{I} \\ s_{\Pi(\omega)}(d^x_\omega) - \widehat{\pi}^\top_{\widehat{k}(\omega)} d^x_\omega, & \text{if } \omega \in \Omega \setminus \mathcal{I}. \end{cases} \tag{31}$$

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

$$|\epsilon_\omega| = |s_{\Pi(\omega)}(d^x_\omega) - s_{\widehat{\Pi}_k}(d^x_\omega)| \leq \Gamma^x_k \vartheta^x_k. \tag{32}$$

For every $\omega \in \Omega \setminus \mathcal{I}$, from equation (31) we have

$$\begin{aligned} |\epsilon_\omega| = |s_{\Pi(\omega)}(d^x_\omega) - \widehat{\pi}^\top_{\widehat{k}(\omega)} d^x_\omega| &= |s_{\Pi(\omega)}(d^x_\omega) - s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^x_\omega) + s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^x_\omega) - \widehat{\pi}^\top_{\widehat{k}(\omega)} d^x_\omega| \\ &\leq |s_{\Pi(\omega)}(d^x_\omega) - s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^x_\omega)| + |s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^x_\omega) - \widehat{\pi}^\top_{\widehat{k}(\omega)} d^x_\omega| \\ &\leq \Gamma^x_\omega \vartheta^x_k + \kappa^x_\omega \vartheta^x_k, \end{aligned} \tag{33}$$

where the inequality (33) follows from assumptions [A] and [B]. Therefore, equations (32) and (33) collectively imply that for any $\omega \in \Omega$,

$$|\epsilon_\omega| \leq \max\{\Gamma^x_k \vartheta^x_k,\ \Gamma^x_k \vartheta^x_k + \kappa^x_\omega \vartheta^x_k\} = (\Gamma^x_k + \kappa^x_k)\, \vartheta^x_k \leq \epsilon^*, \quad \forall \omega \in \mathcal{J}_k. \tag{34}$$

Note that given the compactness of $\mathcal{X}$ and finiteness of $\Omega$, $\epsilon^*$ is well-defined and $\epsilon^* < \infty$. We use the bound (34) on $\mathcal{Q}$ estimation errors to establish that the risk-averse inexact oracle satisfies requirements (3) and (4). We complete this step in two parts focusing on $\widetilde{\varphi}_x$ and $\widetilde{g}_x$, respectively.
**Part 1:** *Correctness of $\widetilde{\varphi}_x$:* The expression of $\varphi(x)$ in (1) and Theorem 2.1 implies that

$$\begin{aligned} \varphi(x) &= c^\top x + \max_{\mu \in \mathfrak{A}_\rho} \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}(x,\omega) \right] = c^\top x + \max_{\mu \in \mathfrak{A}_\rho} \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \left( \mathcal{Q}^{\text{approx}}(x,\omega) + \epsilon_\omega \right) \right] \\ &= c^\top x + \max_{\mu \in \mathfrak{A}_\rho} \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}^{\text{approx}}(x,\omega) + \sum_{\omega \in \Omega} \mu_\omega p_\omega \epsilon_\omega \right] \\ &\leq c^\top x + \max_{\mu \in \mathfrak{A}_\rho} \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}^{\text{approx}}(x,\omega) + \sum_{\omega \in \Omega} \mu_\omega p_\omega \epsilon^* \right] \\ &= \widetilde{\varphi}_x + \epsilon^*. \end{aligned} \tag{35}$$

The last equality comes from the equality $\sum_{\omega\in\Omega}\mu_\omega p_\omega = 1$ from Theorem 2.1 for every $\mu\in\mathfrak{A}_\rho$, and then the definition of $\widetilde{\varphi}_x$ in equation (23). Similarly, we get

$$\begin{aligned}\varphi(x) &= c^\top x + \max_{\mu\in\mathfrak{A}_\rho}\left[\sum_{\omega\in\Omega}\mu_\omega p_\omega\left(\mathcal{Q}^{\text{approx}}(x,\omega)+\epsilon_\omega\right)\right]\\ &\geq c^\top x + \max_{\mu\in\mathfrak{A}_\rho}\left[\sum_{\omega\in\Omega}\mu_\omega p_\omega\left(\mathcal{Q}^{\text{approx}}(x,\omega)-\epsilon^*\right)\right]\\ &= \widetilde{\varphi}_x - \epsilon^*. \end{aligned} \tag{36}$$

Inequalities (35) and (36) collectively imply that $\widetilde{\varphi}_x - \epsilon^* \leq \varphi(x) \leq \widetilde{\varphi}_x + \epsilon^*$, thus satisfying the requirement (3), $\widetilde{\varphi}_x \in [\varphi(x)-\epsilon^*,\ \varphi(x)+\epsilon^*]$, with $\epsilon_1 := \epsilon_2 := \epsilon^*$.

**Part 2:** *Correctness of* $\widetilde{g}_x$. Let $z\in\mathcal{X}$ such that $z\neq x$. Let $\pi^z_\omega$ be an optimal solution for problem (19), 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 $d^z_\omega := \eta(z,\omega)$. Using assumption [A] at $d^z_\omega$, we get

$$\epsilon^* \geq \Gamma^z_{\widehat{k}(\omega)}\vartheta^z_{\widehat{k}(\omega)} \geq |s_{\Pi(\omega)}(d^z_\omega) - s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^z_\omega)| \geq s_{\widehat{\Pi}_{\widehat{k}(\omega)}}(d^z_\omega) - s_{\Pi(\omega)}(d^z_\omega) = \max_{\pi\in\widehat{\Pi}_{\widehat{k}(\omega)}}\pi^\top d^z_\omega - s_{\Pi(\omega)}(d^z_\omega). \tag{38}$$

From equation (22), $\mathcal{Q}^{\text{approx}}(x,\omega) = \eta(x,\omega)^\top\pi_x$, where $\pi_x$ is $\widehat{\pi}_{\widehat{k}(\omega)}$ corresponding to $x$. Since $\pi_x\in\widehat{\Pi}_{\widehat{k}(\omega)}$, we have $\max_{\pi\in\widehat{\Pi}_{\widehat{k}(\omega)}}\pi^\top d^z_\omega \geq \pi_x^\top d^z_\omega$. Applying this inequality in (38) yields

$$s_{\Pi(\omega)}(d^z_\omega) \geq \pi_x^\top d^z_\omega - \epsilon^*. \tag{39}$$

Using (37) and (39) along with the equality $\mathcal{Q}^{\text{approx}}(x,\omega) = \eta(x,\omega)^\top\pi_x$ implies that

$$\begin{aligned}\mathcal{Q}(z,\omega) &\geq \left(\mathcal{Q}^{\text{approx}}(x,\omega) - \eta(x,\omega)^\top\pi_x\right) + \left(\pi_x^\top d^z_\omega - \epsilon^*\right)\\ &= \mathcal{Q}^{\text{approx}}(x,\omega) - \epsilon^* + \pi_x^\top\left(\eta(z,\omega)-\eta(x,\omega)\right).\end{aligned} \tag{40}$$

The convexity and differentiability of $\eta(\cdot,\omega)$ at $x$ implies that $\eta(z,\omega)-\eta(x,\omega)\geq\nabla_x\eta(x,\omega)(z-x)$. Here, $\nabla_x\eta(x,\omega)$ is the $s\times n$ Jacobian matrix of $\eta$. Using this inequality in (40) along with $\pi_x\geq 0$ (problem (14) includes $\pi\geq 0$ as the constraints), we arrive at

$$\mathcal{Q}(z,\omega) \geq \mathcal{Q}^{\text{approx}}(x,\omega) - \epsilon^* + \pi_x^\top\nabla_x\eta(x,\omega)(z-x). \tag{41}$$

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

$$\varphi(z) = c^\top z + \max_{\mu\in\mathfrak{A}_\rho}\left[\sum_{\omega\in\Omega}\mu_\omega p_\omega\mathcal{Q}(z,\omega)\right] \geq c^\top x + \sum_{\omega\in\Omega}\mu^*_\omega p_\omega\mathcal{Q}(z,\omega) + c^\top(z-x),$$

where $\mu^*_\omega$ is an optimal solution to the optimization problem in equation (23) for $x$. Using inequality (41) in the above equality yields

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

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 (4) with $\epsilon_1 := \epsilon_2 := \epsilon^*$. This completes the proof. □

## 5. 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 Chowdhury et al. (2004), Jirutitijaroen and Singh (2006, 2008). We formulate the optimal allocation problem as a two-stage risk-averse stochastic optimization problem, where the second-stage problem is modeled as a network flow problem (Bertsekas 1998). We start by describing the network model next.

### 5.1. The Model

Consider a multi-area network with the set of nodes $\{1,\ldots,n\}$ and the set of edges $E \subseteq \{\{i,j\} \,|\, i,j \in \{1,\cdots,n\}\}$. The network is directed and we set $\tilde{E} := \{(i,j),(j,i) \,|\, \{i,j\} \in E\}$. For any arc $(i,j) \in \tilde{E}$, $i \neq j$. 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 characterized by

$t_{i,j}(\omega)$: tie-line capacity between areas $i$ and $j$ under scenario $\omega$

$c^l_i(\omega)$: cost of demand unfulfillment in area $i$ under scenario $\omega$

$g_i(\omega)$: production capacity of area $i$ under scenario $\omega$

$l_i(\omega)$: demand in area $i$ under scenario $\omega$

Above-mentioned functions of scenario $\omega$ are known and given. In addition, the following parameters are given:

$c_i^b$: 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

We do not assume that the events of network disruptions are independent. Our computational studies let the tie-line distributions be dependent. We refer the reader to Garver (1966), Lago-conzalez and Singh (1989), Mitra and Singh (1999), Lawton et al. (2003), Jirutitijaroen and Singh (2006, 2008) 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).

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 uncertain inputs $t_{i,j}$, $c_i$, $g_i$, 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 $\{x_i\}_{i=1}^n$, flows in the network for each scenario constitute the second-stage decision variables. Denote the flow from arc $i$ to $j$ for the scenario $\omega$ by $y_{ij}(w)$. The precise formulation of this two-stage problem is as follows:

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{n} c_i^b x_i + \rho\left[\mathcal{Q}(x,\omega)\right] \tag{42}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i \le B,\ x \ge 0,\ x \text{ is integral}, \tag{43}$$

where $\rho(\cdot)$ is a coherent risk measure, and for any $\omega \in \Omega$,

$$\mathcal{Q}(x,\omega) = \min_{\substack{y\in\mathbb{R}^{|\tilde{E}|}\\ y_G, y_L\in\mathbb{R}^n}} \quad \sum_{i=1}^{n} c_i^l(\omega)\left(l_i(\omega) - y_{L,i}\right) \tag{44}$$

$$\text{s.t.} \quad y_{G,i} \le g_i(\omega) + G_i^b x_i, \quad i = 1,2,\cdots,n, \tag{45}$$

$$y_{L,i} \le l_i(\omega), \quad i = 1,2,\cdots,n, \tag{46}$$

$$|y_{ji} - y_{ij}| \le t_{i,j}(\omega), \quad (i,j) \in \tilde{E}, \tag{47}$$

$$\sum_{(j,i)\in\tilde{E}} y_{ji} - \sum_{(i,j)\in\tilde{E}} y_{ij} + y_{G,i} - y_{L,i} = 0, \quad i = 1,2,\cdots,n, \tag{48}$$

$$y_{ij}, y_{G,i}, y_{L,i} \ge 0, \quad i,j = 1,2,\cdots,n. \tag{49}$$

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 (45)–(47) correspond to the maximum capacity flow in the network. These constraints consider generation, demand, and tie-line capacity, respectively. Equation (48) 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 (45), $c_i^l(\omega)(l_i(\omega) - y_{L,i}(\omega))$, 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 $\Omega$. Equation (43) 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 (42) 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 $\rho$ 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). 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.

### 5.2. Coherent Risk Measures for Network Reliability Assessment

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

For $\alpha \geq 0$, the *mean-upper-semideviation*, denoted by $\mathrm{MUSD}_\alpha$, measures the risk of losses exceeding the expectation (Ogryczak and Ruszczyński (1999, 2001)) and is defined by

$$\mathrm{MUSD}_\alpha(Z) := \mathbb{E}[Z] + \alpha\mathbb{E}\left[Z - \mathbb{E}[Z]\right]_+, \qquad \forall\ Z \in \mathcal{Z}, \tag{50}$$

The *conditional value-at-risk* at level $\alpha \in [0,1]$ (Rockafellar and Uryasev (2000, 2002), Ruszczyński and Shapiro (2006b)), denoted by $\mathrm{CVaR}_\alpha$, is the expectation of $Z$ in the conditional distribution of its upper $\alpha$-tail, and is given by

$$\mathrm{CVaR}_\alpha(Z) := \inf_{t\in\mathbb{R}} \left\{ t + \alpha^{-1}\mathbb{E}[Z-t]_+ \right\}. \tag{51}$$

Given a finite probability space with $\Omega = \{\omega_1, \cdots, \omega_N\}$, the corresponding risk envelopes are

$$\mathfrak{A}_{\mathrm{MUSD}_\alpha} = \{\mu \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) \mid \mu = 1 + \tau - \mathbb{E}[\tau],\ \ \|\tau\|_\infty \leq \alpha\}, \tag{52}$$

$$\mathfrak{A}_{\mathrm{CVaR}_\alpha} = \left\{\mu \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) \,\middle|\, \mu(\omega) \in [0, \alpha^{-1}] \ \text{ a.e. } \omega \in \Omega,\ \ \mathbb{E}[\mu] = 1\right\}. \tag{53}$$

For other examples of coherent risk measures and their representations see Section 6.3.2 in Shapiro et al. (2014) and Rockafellar (2014).

For a discussion on deploying coherent risk measures for handling the risk of system failure and reliability analysis, see Minguez et al. (2011), Rockafellar and Royset (2015), Gardoni (2017). The two-stage model (42)-(49) for a finite probability space can be represented as a one-stage deterministic linear optimization problems. The deterministic equivalent optimization formulations for these risk functionals are presented in Appendix A. Solving the deterministic equivalent formulations becomes quickly computationally expensive as the number of scenarios $N$ grows.

### 5.3. Application of Risk-Averse Inexact Proximal Bundle Method

This section presents the details of applying the inexact proximal bundle method outlined in Section 3 with the risk-averse inexact oracle described in Section 4 for the two-stage optimization model (42)-(49).

For each cluster $\mathcal{J}_k$, the solutions $\widehat{\pi}_k$ is computed as in equations (21). Let $\widehat{\pi}_k^g, \widehat{\pi}_k^l, \widehat{\pi}_k^{t+}, \widehat{\pi}_k^{t-}$, and $\widehat{\pi}_k^b$ refer to part of $\widehat{\pi}_k$ corresponding to the set of constraints (45)-(48) in the second-stage optimization problem. Hence, equations (20) and (21) are expressed as

$$\widehat{\pi}_k \in \arg\max_{\pi \in \widehat{\Pi}_k} \left\{ \sum_{i=1}^{n} (g_i(\widehat{\omega}_k) + G_i^b x_i)\pi_i^g + \sum_{i=1}^{n} l_i(\widehat{\omega}_k)\pi_i^l + \sum_{(i,j)\in E} t_{i,j}(\widehat{\omega}_k)(\pi_{ij}^{t+} + \pi_{ij}^{t-}) \right\}, \tag{54}$$

where

$$\widehat{\Pi}_k = \left\{ \pi = (\pi^g, \pi^l, \pi^{t+}, \pi^{t-}, \pi^d) \in \mathbb{R}^{3n+2|E|} \;\middle|\; \begin{array}{l} \pi_i^g + \pi_i^b \le 0, \quad i = 1, \cdots, n \\ \pi_i^L - \pi_i^b \le \mathbb{E}[c_i^l(\omega) | \omega \in \mathcal{J}_k], \quad i = 1, \cdots, n \\ \pi_{ij}^{t+} - \pi_{ij}^{t-} + \pi_j^b - \pi_i^b \le 0, \quad (i,j) \in E \\ \pi_i^G, \pi_i^L, \pi_{ij}^{t+}, \pi_{ij}^{t-} \ge 0 \end{array} \right\}. \tag{55}$$

Note that here $\mathbb{E}[c_i^l(\omega)|\omega \in \mathcal{J}_k] = \sum_{\omega \in \mathcal{J}_k} p_\omega c_i^l(\omega) / \sum_{\omega \in \mathcal{J}_k} p_\omega$. These solutions are used to construct $\mathcal{Q}^{\text{approx}}(x, \omega)$. For some $\omega \in \Omega$, suppose that $\widehat{k}(\omega) = k$. Thus,

$$\mathcal{Q}^{\text{approx}}(x,\omega) = \sum_{i=1}^{n} (g_i(\omega) + G_i^b x_i)\widehat{\pi}_{k,i}^g + \sum_{i=1}^{n} l_i(\omega)\widehat{\pi}_{k,i}^l + \sum_{(i,j)\in E} t_{i,j}(\omega)(\widehat{\pi}_{k,ij}^{t+} + \widehat{\pi}_{k,ij}^{t-}). \tag{56}$$

These approximate optimal values for the second-stage problem are used to derive $\widetilde{\varphi}_x$ as stated in equation (23). In particular, using (53) for $\text{CVaR}_\alpha$ we set

$$\begin{aligned} \widetilde{\varphi}_x := c^\top x + \max_{\mu \in \mathbb{R}^N} \quad & \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}^{\text{approx}}(x,\omega) \right] \\ \text{s.t.} \quad & \sum_{s=1}^{N} p_s \mu_s = 1, \;\; 0 \le \mu_s \le \frac{1}{\alpha} \quad s = 1, \cdots, N. \end{aligned} \tag{57}$$

Using the expression in (52) for $\text{MUSD}_\alpha$ we compute

$$\begin{aligned} \widetilde{\varphi}_x := c^\top x + \max_{\mu \in \mathbb{R}^N} \quad & \left[ \sum_{\omega \in \Omega} \mu_\omega p_\omega \mathcal{Q}^{\text{approx}}(x,\omega) \right] \\ \text{s.t.} \quad & \mu_s = 1 + \tau - \sum_{s=1}^{N} p_s \tau_s + \tau, \quad 0 \le \tau_s \le \alpha, \quad s = 1, \cdots, N. \end{aligned} \tag{58}$$

Suppose that $\mu^* \in \mathbb{R}^N$ solves the maximization in the description of $\widetilde{\varphi}_x$ above. Then, the approximate subgradient follows from $\nabla_x \mathcal{Q}^{\text{approx}}(x,\omega) = [\widehat{\pi}^g_{k,1} G^b_1 \cdots \widehat{\pi}^g_{k,n} G^b_n]^\top$, that is

$$\widetilde{g}_x := c + \sum_{\omega \in \Omega} \mu^*_\omega p_\omega \begin{bmatrix} G^b_1 \widehat{\pi}^g_{k,1} \\ \vdots \\ G^b_n \widehat{\pi}^g_{k,n} \end{bmatrix}. \tag{59}$$

### 5.4. Convergence with First-Stage Integrality Constraints

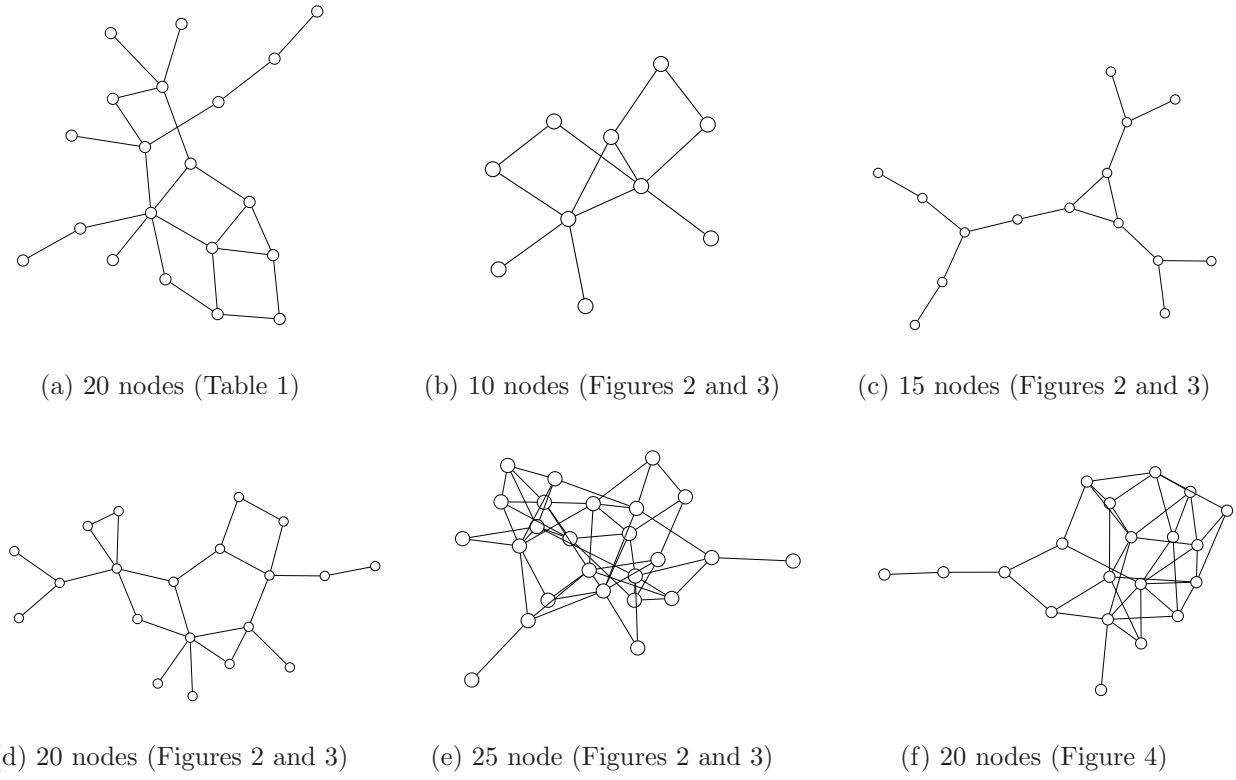
The convergence analysis of the inexact proximal bundle method established in Kiwiel (2006) relies on the optimality condition for convex optimization problems. Hence, this analysis cannot be directly employed for the resource allocation problem introduced in (42)–(49) due to its first-stage integrality constraints. We heuristically apply Algorithm 1 to the resource allocation problem by letting the regularized master problem (18) at each iteration be an integer optimization problem. In addition, in our implementation, the bundle management in Step 6 is forgone.

We numerically investigate the convergence of the algorithm for the resource allocation problem of our interest with first-stage integer decision variables. For this purpose, we develop the deterministic equivalent program (see e.g. Birge and Louveaux (1997)) for our two-stage stochastic optimization problem when $\rho$ is replaced by expectation, mean-upper semideviation, and CVaR. These deterministic equivalent formulations are presented in Appendix A. For a given scenario set, deterministic equivalent programs provide an optimal solution of the two-stage stochastic optimization problem. However, these formulations are generally computationally prohibitive as the number of scenarios grows. We solve these mixed-integer problems using the Gurobi's branch and bound algorithm (Gurobi Optimization 2016). We use the solutions computed from the deterministic equivalent formulations to assess the optimality of the solutions obtained from the application of Algorithm 1 to our resource allocation problem. For this study, we consider a randomly generated network of 20 nodes and 78 arcs and vary the number of scenarios from 100 to 2000 scenarios. The rest of the simulation parameters are as in section 6. We observe that for all three risk functionals and for all scenario sets, the relative error is consistently less than 0.01%. For expectation, Algorithm 1 with an integer master problem is often capable to offer the exact optimal solution obtained from the deterministic equivalent program.

## 6. Computational Results

In our numerical experiments we randomly generate sparse connected networks of different sizes using the Networkx Python library (Hagberg et al. 2008). Figure 1 presents some of the networks used in our subsequent analyses.

Sets of scenarios $\Omega$ specifying $t_{i,j}(\omega)$, $g_i(\omega)$, $l_i(\omega)$, $c^l_i(\omega)$ are generated by simulation. We build on the existing literature to determine our simulation setting and parameter values. The details of our



(a) 20 nodes (Table 1) (b) 10 nodes (Figures 2 and 3) (c) 15 nodes (Figures 2 and 3)

(d) 20 nodes (Figures 2 and 3) (e) 25 node (Figures 2 and 3) (f) 20 nodes (Figure 4)

**Figure 1 Examples of randomly generated networks**

scenario simulation are presented in Appendix B. In addition, we set $c_i^b = \$205,100$ and $G_i^b = 0.1\text{MW}$ for every node $i$. We set the maximum number of extra reserve resources $B$ to be 3% of the total maximum generation capacity on the network. That is, $B = 0.03\sum_{i=1}^{n} \bar{G}_i$, where $\bar{G}_i$ is the upper bound of the support of the probability distribution for $g_i(\omega)$, see Table 2a. The software containing these problems is available at `https://archive.org/details/contingency-data.-7z`.

We consider three risk functionals $\rho$: expectation, mean-upper semideviation, and CVaR. Throughout, the risk parameters in CVaR and mean-upper semideviation are set to $\alpha = 0.40$. For the exact proximal bundle method, the regularization parameter is set to $\gamma = 0.31$. In the inexact proximal bundle method, presented in Section 3, the descent parameter is $\varrho = 0.3$, the initial stepsize bound is $T_1 = 0.05$, and the initial stepsize is $t_1 = 0.1$. In both the exact and inexact bundle methods, the stopping tolerance $\delta = 10^{-6}$ is used.

The master problem in the proximal bundle methods and deterministic formulations (60), (61), and (62) are solved using Gurobi quadratic, mixed-integer, and dual simplex solvers (Gurobi Optimization 2016), implemented in the Python programming language. In the exact method, the objective functions and subgradients are evaluated using equations (7) and (13). The inexact method employs equations (23) and (59). 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 and 128GB of RAM is used in these numerical experiments.

### 6.1. Accuracy: Inexact Bundle Method versus Exact Bundle Method

Table 1 reports the run-times of the inexact bundle method for different risk measures and its corresponding suboptimality levels against the exact bundle method. In this section, all parameters are fixed except $\epsilon_{\cos}$. The suboptimility is computed as the absolute value of the difference of optimal values of exact and inexact methods, divided by the optimal value of the exact method. In this analysis, a network with 20 nodes and 48 arcs is considered. Figure 1a illustrates the undirected version of this network, with $48/2 = 24$ arcs. The size of the scenario set is $|\Omega| = 100$. For each risk measure, the exact running time is also provided.

Results in Table 1 show that the running time can be reduced by more than 2.85 times for CVaR, 7.98 times for Mean-Upper Semideviation, and 8 times for expectation at about 10% suboptimality. In addition, Table 1 suggests that $\epsilon_{\cos} = 0.1$ offers an acceptable tradeoff between approximation error and run-time. This observation informs the numerical experiments in the subsequent sections to focus on the inexact proximal bundle method with two "extreme" conditions of $\epsilon_{\cos} = 0.2$ and $\epsilon_{\cos} = 0.05$, and a "reasonable" condition with $\epsilon_{\cos} = 0.1$.

**CVaR**

| $\epsilon_{\cos}$ | Time | Suboptimality |
|---|---|---|
| 0.9 | 2.6 | 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 | 28.1 | 10.51% |
| 0.08 | 21.77 | 5.28% |
| 0.06 | 41.88 | 2.95% |
| 0.05 | 94.53 | 1.76% |
| 0.04 | 61.97 | 0.23% |
| 0.02 | 71.03 | 0.00% |
| **Exact running time: 80.43** | | |

**Mean-Upper 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 | 5.5 | 32.65% |
| 0.2 | 7.23 | 14.81% |
| 0.1 | 15.22 | 0.31% |
| 0.08 | 23.73 | 2.31% |
| 0.06 | 42.43 | 2.05% |
| 0.05 | 69.67 | 1.67% |
| 0.04 | 79.26 | 0.12% |
| 0.02 | 61.04 | 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% |
| 0.08 | 22.52 | 0.84% |
| 0.06 | 49.93 | 1.16% |
| 0.05 | 103.80 | 1.27% |
| 0.04 | 148.38 | 0.14% |
| 0.02 | 124.28 | 0.00% |
| **Exact running time: 124.87** | | |

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

Figure 2 exhibits the percentage of approximation error (suboptimality) of the solution computed from the inexact bundle 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 3 to 25 nodes. Some of these graphs are presented in Figures 1b-1e. In the

plots in Figure 2, each point represents the average over 5 simulations on a fixed network with 100 randomly generated scenarios. Other settings are similar to those explained previously.

The error in approximation improves with smaller values of $\epsilon_{\cos}$. This is expected since for a fixed scenario indexing, a smaller $\epsilon_{\cos}$ tends to create finer partitions $\mathcal{I}$, thus leading to solving more second-stage scenarios in Algorithm 2. The suboptimality level is relatively consistent among different risk measures and different scenario sizes. Right plots in Figure 2 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 make 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.

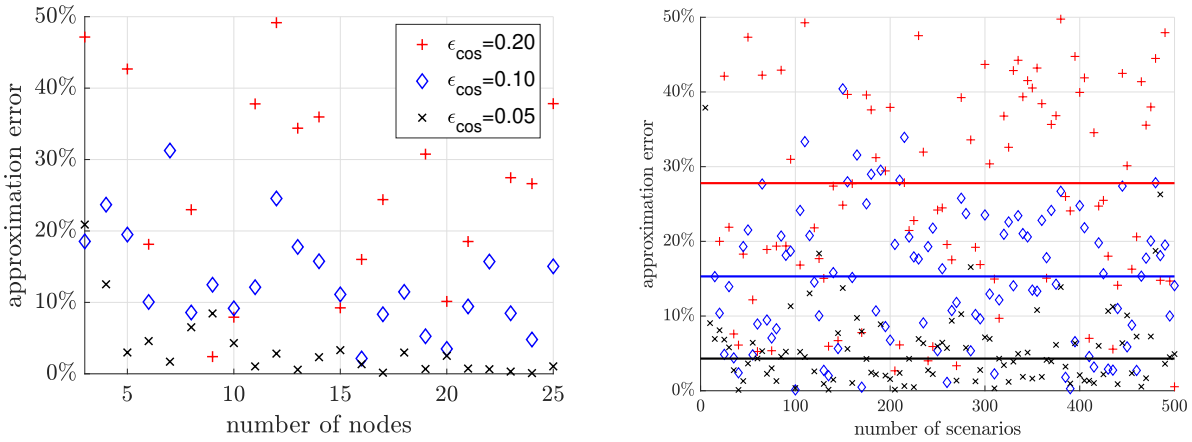
### 6.2. Computational Time: Inexact Bundle Method versus Exact Bundle Method

This section reports the run-time of the inexact proximal bundle method and compares it with the run-time of the exact proximal 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 $\Omega$, and the parameter $\epsilon_{\cos}$ in the inexact risk-averse oracle.

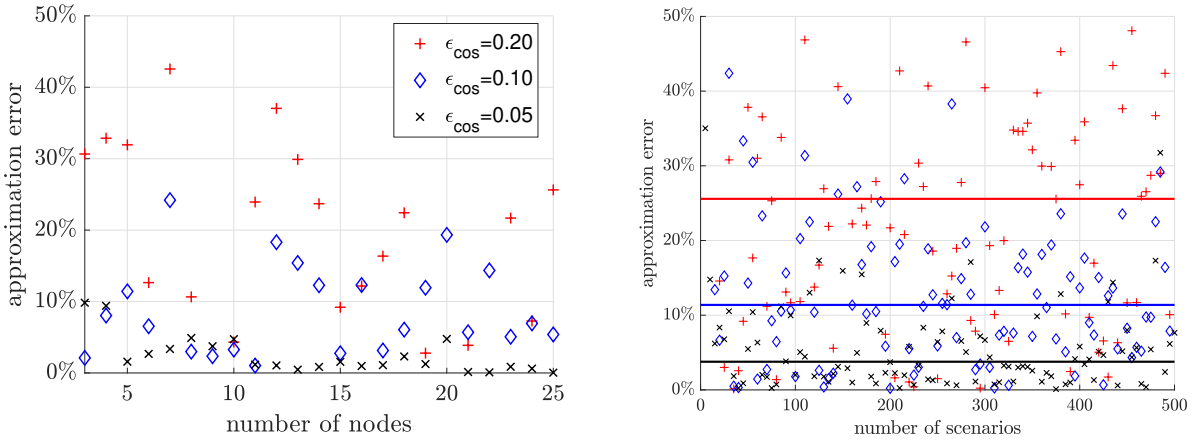
Figure 3 illustrates the run-times (in seconds) of the exact method and the inexact bundle methods for $\epsilon_{\cos} = 0.2, 0.1, 0.05$. The left plots in Figure 3 illustrate the run-times to randomly generated networks of 5 to 25 nodes, some of which are presented in Figures 1b-1e, and a scenario set with $|\Omega| = 100$ scenarios. Every point in these plots represent the average over 5 simulations. The left plots in Figure 3 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.

The right plots in Figure 3 depict the run-times of the exact and inexact methods as the number of scenarios $|\Omega|$ increases. Here, we consider a network with 10 nodes 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.

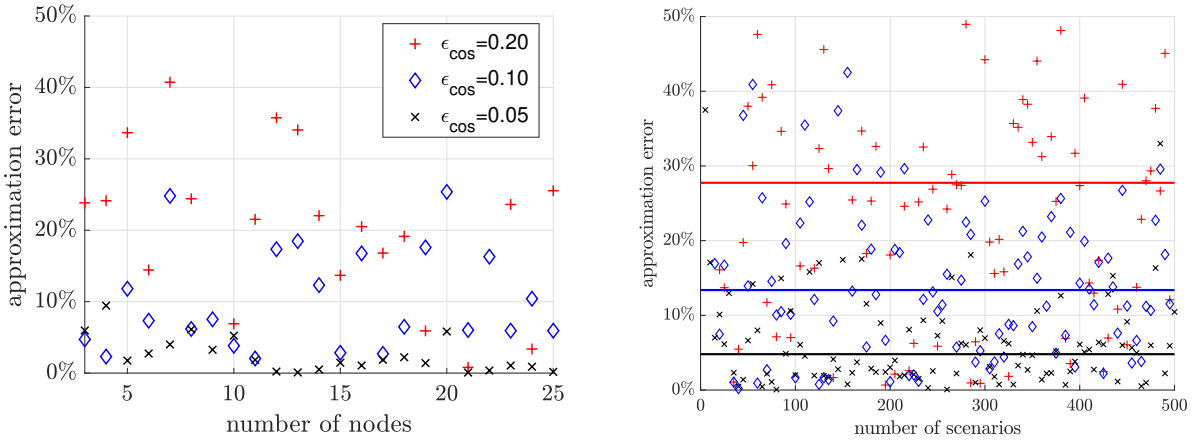
Figure 4 illustrates the run-time of the exact and inexact methods as the number of scenarios increases to 5000. This study considers a network with 20 nodes as in Figure 1f. In this experiment, a single-threaded code environment is used. All parameters are set as before. The plots in Figure 4




(a) Conditional Value-at-Risk

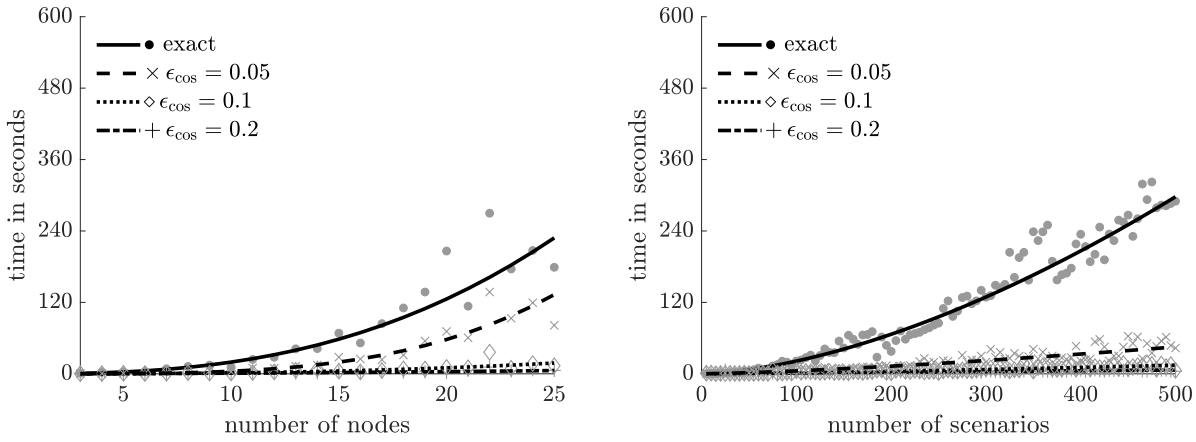



(b) Mean-Upper Semideviation

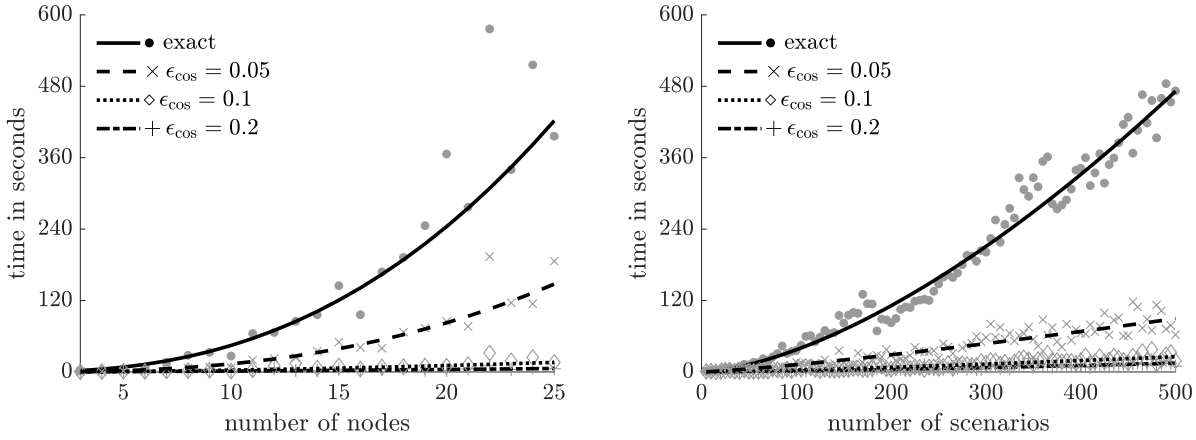



(c) Expectation

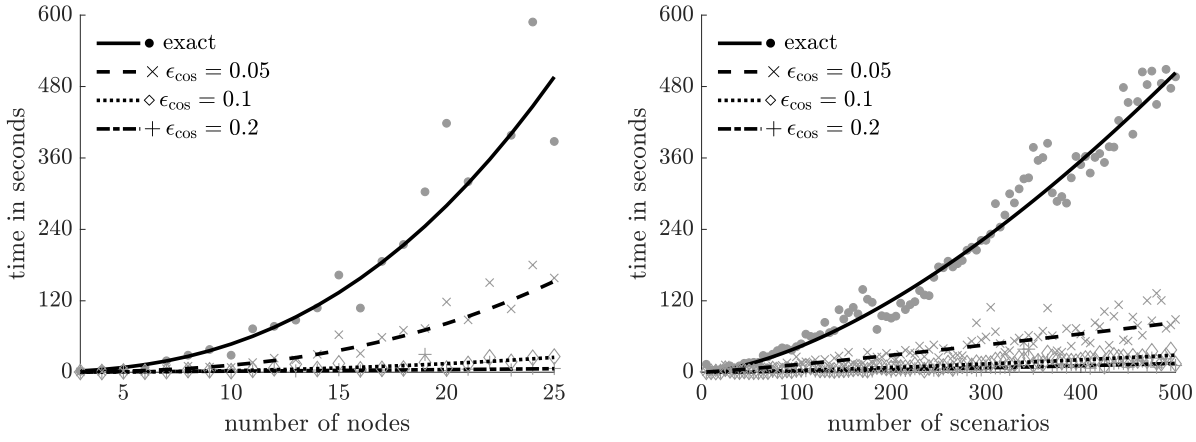
**Figure 2** **Approximation error of inexact proximal bundle methods with $\epsilon_{\cos} = 0.2$ (red), $0.1$ (blue), $0.05$ (black). Left plot: approximation error vs. number of nodes. Right plot: approximation error vs. number of scenarios. The lines show the medians.**




(a) Conditional Value-at-Risk



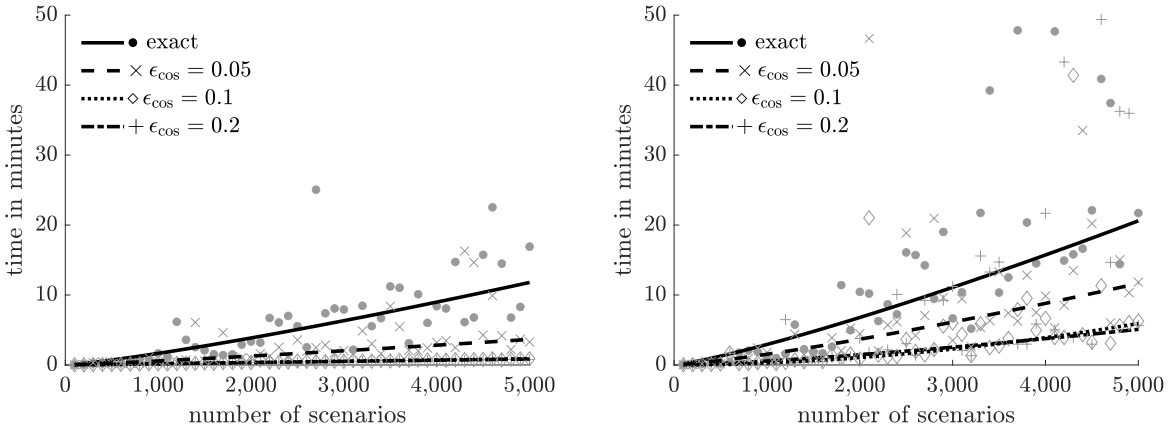

(b) Mean-Upper Semideviation




(c) Expectation

**Figure 3** **Run-time of exact and inexact proximal bundle methods with $\epsilon_{\cos} = 0.05, 0.1, 0.2$. Left plots: run-time vs. number of nodes. Right plots: run-time vs. number of scenarios. The curves show the fitted robust power models.**

also show a fitted power model $t = cN^{\gamma}$, where $t$ denotes the running time, $N$ is the number of scenarios, and $c, \gamma$ are model parameters. The plots indicate that generally significant improvements in the run-times can be achieved by the inexact method for both risk measures. The run time improvement relative to that of the exact method becomes more prominent as the number of scenarios increases. Consider that there is a run-time budget of five minutes. We observe that allowing the method with the CVaR objective run for up to five minutes, the exact method solves problems up to $2,472$ scenarios, whereas the inexact methods can handle problems up to $6,565$ for $\epsilon_{\cos} = 0.05$, up to $28,682$ for $\epsilon_{\cos} = 0.1$, and up to $30,829$ for $\epsilon_{\cos} = 0.2$. Similarly, for the mean-upper semideviation objective, letting the methods run for up to five minutes, the exact method solves problems up to $1,553$ scenarios, whereas the inexact methods can handle problems up to $2,552$ for $\epsilon_{\cos} = 0.05$, up to $4,565$ for $\epsilon_{\cos} = 0.1$, and up to $4,931$ for $\epsilon_{\cos} = 0.2$.




(a) Conditional Value-at-Risk (b) Mean-Upper Semideviation

**Figure 4 Run-time of exact and inexact proximal bundle methods as the number of scenarios grows.**

## 7. Conclusion

This paper studies the resource allocation problem as a two-stage stochastic optimization problem with risk-averse recourse. Solving this problem using the deterministic equivalents of the problem or exact proximal bundle method becomes computationally expensive for large number of scenarios. An inexact proximal bundle method with a risk-averse inexact oracle to compute approximate objective function values and subgradients is developed, for coherent risk measures and convex second-stage problems. Sufficient conditions are established for the correctness of the risk-averse oracle, when the second-stage optimal value for each scenario admits a linear representation. The performance of the methodology is investigated for the resource allocation problem for reserve

resources arising in contingency planning, modeled as a risk-averse two-stage stochastic optimization. Computational experiments are conducted for various networks and scenario sets, collected as a library of test problems. A sensitivity analysis on the scenario clustering parameter $\epsilon_{\cos}$ in the risk-averse inexact oracle for this two-stage risk averse stochastic problem is carried out to guide on the selection of an appropriate value for this parameter. Our numerical results exhibit that the inexact proximal bundle method can provide significant improvement in the run-time to achieve an approximate solution for this two-stage problem, comparing to the exact bundle method. Such runtime improvements depend on the decision maker's approximation preference controlled by the choice of the clustering method. Future research includes investigating the performance of the risk-averse inexact oracle when applied to problem classes beyond linear second-stage problems and to other resource allocation applications. Studying the effect of alternative scenario clustering mechanisms on the solution approximation and the granularity of the results constitutes another direction for future work.

## Acknowledgments

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

### Appendix A: Deterministic Equivalent Formulations

In this section, $\mathbb{R}^n_+$ denotes the set of all real vectors of dimension $n$ with nonnegative elements. Let $\mathcal{X} = \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, x \text{ is integer}, \sum_{i=1}^n x_i \leq B\}$.

**Expectation:**

$$\begin{aligned}
\min_{x\in\mathcal{X},\ y\in\mathbb{R}_+^{|\mathbf{E}||\Omega|}} \quad & \sum_{i=1}^{n} c_i^b x_i + \sum_{\omega\in\Omega} p_\omega \sum_{i=1}^{n} c_i^l(\omega)\left(l_i(\omega) - y_{L,i,\omega}\right) \\
\text{s.t.} \quad & y_{G,i,\omega} \leq g_i(\omega) + G_i^b x_i, \quad i \in \{1,2,\cdots,n\},\ \omega\in\Omega, \\
& y_{L,i,\omega} \leq l_i(\omega), \quad i \in \{1,2,\cdots,n\},\ \omega\in\Omega, \\
& |y_{ji,\omega} - y_{ij,\omega}| \leq t_{i,j}(\omega), \quad i,j \in \{1,2,\cdots,n\},\ i\neq j,\ \omega\in\Omega, \\
& \sum_{\substack{j\in I\\ j\neq i}} y_{ji,\omega} - \sum_{\substack{j\in I\\ j\neq i}} y_{ij,\omega} + y_{G,i,\omega} - y_{L,i,\omega} = 0, \quad i \in \{1,2,\cdots,n\},\ \omega\in\Omega.
\end{aligned} \tag{60}$$

**Mean-Upper Semideviation:**

$$\begin{aligned}
\min_{x\in\mathcal{X},\ y\in\mathbb{R}_+^{|\mathbf{E}||\Omega|},\ T\in\mathbb{R}^{|\Omega|},\ R\in\mathbb{R}_+^{|\Omega|}} \quad & \sum_{i=1}^{n} c_i^b x_i + \sum_{\omega\in\Omega} p_\omega T_\omega + \alpha \sum_{\omega\in\Omega} p_\omega R_\omega \\
\text{s.t.} \quad & y_{G,i,\omega} \le g_i(\omega) + G_i^b x_i, \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& y_{L,i,\omega} \le l_i(\omega), \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& |y_{ji,\omega} - y_{ij,\omega}| \le t_{i,j}(\omega), \quad i,j\in\{1,2,\cdots,n\},\ i\ne j,\ \omega\in\Omega, \\
& \sum_{\substack{j\in I\\ j\ne i}} y_{ji,\omega} - \sum_{\substack{j\in I\\ j\ne i}} y_{ij,\omega} + y_{G,i,\omega} - y_{L,i,\omega} = 0, \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& T_\omega = \sum_{i=1}^{n} c_i^l(\omega)\left(l_i(\omega) - y_{L,i,\omega}\right),\ \omega\in\Omega, \\
& R_\omega \ge T_\omega - \sum_{\omega\in\Omega} p_\omega T_\omega,\ \omega\in\Omega.
\end{aligned} \tag{61}$$

**CVaR:**

$$\begin{aligned}
\min_{x\in\mathcal{X},\, t\in\mathbb{R}, y\in\mathbb{R}_+^{|\mathbf{E}||\Omega|}, T\in\mathbb{R}^{|\Omega|}, R\in\mathbb{R}_+^{|\Omega|}} \quad & \sum_{i=1}^{n} c_i^b x_i + \sum_{\omega\in\Omega} p_\omega T_\omega + \alpha \sum_{\omega\in\Omega} p_\omega R_\omega \\
\text{s.t.} \quad & y_{G,i,\omega} \le g_i(\omega) + G_i^b x_i, \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& y_{L,i,\omega} \le l_i(\omega), \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& |y_{ji,\omega} - y_{ij,\omega}| \le t_{i,j}(\omega), \quad i,j\in\{1,2,\cdots,n\},\ i\ne j,\ \omega\in\Omega, \\
& \sum_{\substack{j\in I\\ j\ne i}} y_{ji,\omega} - \sum_{\substack{j\in I\\ j\ne i}} y_{ij,\omega} + y_{G,i,\omega} - y_{L,i,\omega} = 0, \quad i\in\{1,2,\cdots,n\},\ \omega\in\Omega, \\
& T_\omega = \sum_{i=1}^{n} c_i^l(\omega)\left(l_i(\omega) - y_{L,i,\omega}\right),\ \omega\in\Omega, \\
& R_\omega \ge T_\omega - t,\ \omega\in\Omega.
\end{aligned} \tag{62}$$

## Appendix B: Scenario Simulation

In order to capture the local structure of the network we generate distance-based covariance matrices $\Sigma_I$ and $\Sigma_E$. To do this, we first obtain matrices $D_I$ and $D_E$ of distance in the underlying undirected network between nodes in $I$ and between arcs in $E$, respectively. For every $a,b\in\{1,\cdots,n\}$ and $e,d\in E$ we define

$$\Sigma_I(a,b) := \frac{1}{2}\left[\exp\left(-\left[\frac{D_I(a,b)}{\rho_I n}\right]^2\right) + \delta_{a,b}\right] \quad \text{and} \quad \Sigma_E(e,d) := \frac{1}{2}\left[\exp\left(-\left[\frac{D_E(e,d)}{\rho_E |E|}\right]^2\right) + \delta_{e,d}\right],$$

where $\delta_{i,j}$ is the Kronecker delta and $\rho_I, \rho_E$ are parameters controlling the strength of linear correlation. In our experiments we set $\rho_I = 0.3$ and $\rho_E = 0.4$. The distance $D_I(a,b)$ is the number of arcs on the shortest path from node $a$ to $b$. The distance $D_E(e,d)$ is the number of arcs on the shortest path from node $e$ to $d$ in the line graph of the network. We avoid numerical instabilities in the calculation of $\Sigma_I$ and $\Sigma_E$ by running them through the algorithm in Higham (1988) to find a nearest symmetric positive semidefinite matrix. We use the matrices $\Sigma_I(a,b)$ and $\Sigma_E(e,d)$ to generate correlated simulated values, described next.

$t_{i,j}(\omega)$: To each tie-line $(i,j)\in\tilde{E}$, we attach an independent discrete *total failure* distribution taking values 100 and 0.0 with probability 0.7 and 0.3, respectively. The high probability of total failure

| **max gen. 500 MW** | | **max gen. 600 MW** | |
|---|---|---|---|
| **Cap. (MW)** | **Prob.** | **Cap. (MW)** | **Prob.** |
| – | – | 600 | 0.564474 |
| 500 | 0.620921 | 500 | 0.338684 |
| 400 | 0.310461 | 400 | 0.084671 |
| 300 | 0.062092 | 300 | 0.011289 |
| 200 | 0.006209 | 200 | 0.000847 |
| 100 | 0.000310 | 100 | 0.000034 |
| 0 | 0.000006 | 0 | 0.000001 |

(a) Probability models for Area Generation

| **max load 500** | | **max load 600** | |
|---|---|---|---|
| **Cap. (MW)** | **Prob.** | **Cap. (MW)** | **Prob.** |
| – | – | 600 | 0.028257 |
| 500 | 0.028257 | 500 | 0.275288 |
| 400 | 0.275288 | 400 | 0.436651 |
| 300 | 0.436651 | 300 | 0.259803 |
| 200 | 0.259803 | – | – |

(b) Regimes for Area Load

**Table 2 Discrete probability distributions to generate two regimes for area generation $g_i$ and load $l_i$**

simulates times of great stress in the network such as storms and other natural disasters. In order to capture the structure of the network, we let $t_{i,j}$ be obtained from a multivariate probability with a Gaussian copula and total failure distribution marginals with values correlated by the tie-line distance covariance matrix $\Sigma_E$.

$g_i(\omega)$: Each area is randomly assigned one of two *generation distributions* defining regimes constraining maximum generation to 500MW or 600MW, see Table 2a. We let $g_i$ be obtained from a multivariate probability with a Gaussian copula and generation distribution marginals with values correlated by the covariance matrix $\Sigma_I$. In this way, generation values on closer areas are more correlated than on far apart ones, thus simulating local constraints on the generation areas.

$l_i(\omega)$: Each area is randomly assigned one of two *load distributions* defining maximum-load regimes, see Table 2b. We let $l_i := \Xi_i + \Lambda_i$, where $\Xi_i$ is obtained from a multivariate probability with a Gaussian copula and load distribution marginals with values correlated by the covariance matrix $\Sigma_I$ and $\Lambda_i$ is an independent Poisson distribution. $\Lambda_i$ models the load spikes integral to load values in electricity networks, see Carmona and Coulon (2014) for more details on modeling electricity markets.

$c_i^l(\omega)$: The cost of demand loss function is expressed through a *customer damage function* (CDF) that relates different types of load and interruption duration to cost per MW. We use the CDF appeared in Lawton et al. (2003), which is defined as follows $c^l = \exp(6.48005 + 0.38489 D_\omega - 0.02248 D_\omega^2)$, where $D_\omega$ is the mean duration of each state, see e.g. Samaan (2004), Jirutitijaroen and Singh (2006, 2008). This quantity is defined

$$D_\omega = 24 \left[ \sum_{i \in I} \lambda_{g_i}^{\omega+} + \sum_{i \in I} \lambda_{g_i}^{\omega-} + \sum_{\substack{i,j \in I \\ i \neq j}} \lambda_{t_{ij}}^{\omega+} + \sum_{\substack{i,j \in I \\ i \neq j}} \lambda_{t_{ij}}^{\omega-} + \sum_{k=1}^{m_k} \lambda_{l_k}^{\omega} \right]^{-1}.$$

Here, $m_t$ is the total number of area load states, $\lambda_{l_k}^{\omega}$ is the equivalent transition rate of area load from state $\omega$ to other load states, and $\lambda_{g_i}^{\omega+}$ ($\lambda_{g_i}^{\omega-}$) expresses the equivalent transition rate of generation in area $i$ from a capacity of state $\omega$ to higher (lower) capacity. Similarly, $\lambda_{t_{ij}}^{\omega+}$ ($\lambda_{t_{ij}}^{\omega-}$) is the equivalent transition rate of transmission line from area $i$ to area $j$ from a capacity of state $\omega$ to higher (lower) capacity. The values of these parameters are given in Table 3, see also Mitra and Singh (1999).

| Cap. (MW) | max gen. 500 | | max gen. 600 | |
|---|---|---|---|---|
| | $\lambda^{\omega+}_{g_i}$ | $\lambda^{\omega-}_{g_i}$ | $\lambda^{\omega+}_{g_i}$ | $\lambda^{\omega-}_{g_i}$ |
| 600 | – | – | 0.6 | 0 |
| 500 | 0.5 | 0 | 0.5 | 1 |
| 400 | 0.4 | 1 | 0.4 | 2 |
| 300 | 0.3 | 2 | 0.3 | 3 |
| 200 | 0.2 | 3 | 0.2 | 4 |
| 100 | 0.1 | 4 | 0.1 | 5 |
| 0 | 0 | 5 | 0 | 6 |

(a) Transition rate regimes for generation

| index | $\lambda^{\omega}_{l_k}$ |
|---|---|
| 0 | 1.3429 |
| 1 | 0.0206 |
| 2 | 0.3394 |
| 3 | 1.9753 |
| 4 | 0.0278 |
| 5 | 0.0085 |
| 6 | 1.3399 |
| 7 | 2.1036 |
| 8 | 0.0452 |
| 9 | 2.237 |

(b) Transition rate for loads

| Cap. (MW) | $\lambda^{\omega+}_{t_{ij}}$ | $\lambda^{\omega-}_{t_{ij}}$ |
|---|---|---|
| 100 | 0.274 | 0 |
| 0 | 0 | 3 |

(c) Transition rates for tie-lines

**Table 3 Generation, load, and tie-line transition rates**

## 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.

Asamov, Tsvetan, Warren B. Powell. 2018. Regularized decomposition of high-dimensional multistage stochastic programs with markov uncertainty. *SIAM on Optimization* **28**(1) 575–595.

Avlov, Alexander, Dmitry Ivanov, Dmitry Pavlov, Alexey Slinko. 2019. Optimization of network redundancy and contingency planning in sustainable and resilient supply chain resource management under conditions of structural dynamics. *Annals of Operations Research* doi:10.1007/s10479-019-03182-6.

Behzadi, Golnar, Michael Justin O'Sullivan, Tava Lennon Olsen. 2020. On metrics for supply chain resilience. *European Journal of Operational Research* **287**(1) 145–158.

Bertsekas, Dimitri. 1998. *Network Optimization: Continuous and Discrete Models*. Athena Scientific.

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

Carmona, René, Michael Coulon. 2014. A survey of commodity markets and structural models for electricity prices. *Quantitative Energy Finance*. Springer, 41–83.

Chen, Richard Li-Yang, Neng Fan, Ali Pinar, Jean-Paul Watson. 2017. Contingency-constrained unit commitment with post-contingency corrective recourse. *Annals of Operations Research* **249**(1-2) 381–407.

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, Ricardo A, Dávid Papp, Andrzej 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.

de Oliveira, W., C. Sagastizábal, C. Lemaréchal. 2014. Convex proximal bundle methods in depth: a unified analysis for inexact oracles. *Mathematical Programming* **148**(1) 241–277.

de Oliveira, Welington, Mikhail Solodov. 2020. *Bundle Methods for Inexact Data*. Springer International Publishing, Cham, 417–459.

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

Gardoni, Paolo. 2017. *Risk and Reliability Analysis: Theory and Applications*. Springer.

Garver, L. L. 1966. Effective load carrying capability of generating units. *IEEE Transactions on Power Apparatus and Systems* **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.

Guo, Zhaomiao, Richard Li-Yang Chen, Neng Fan, Jean-Paul Watson. 2016. Contingency-constrained unit commitment with intervening time for system adjustments. *IEEE Transactions on Power Systems* **32**(4) 3049–3059.

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`.

Hagberg, Aric A., Daniel A. Schult, Pieter J. Swart. 2008. Exploring network structure, dynamics, and function using networkx. 11–15.

Higham, Nicholas J. 1988. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications* **103** 103–118.

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.

Kleindorfer, Paul R, Germaine H Saad. 2005. Managing disruption risks in supply chains. *Production and operations management* **14**(1) 53–68.

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.

Le, Quoc V, Alex J Smola, Svn Vishwanathan. 2008. Bundle methods for machine learning. *Advances in neural information processing systems*. 1377–1384.

Liu, Junyi, Suvrajeet Sen. 2020. Asymptotic results of stochastic decomposition for two-stage stochastic quadratic programming. *SIAM Journal on Optimization* **30**(1) 823–852.

Mäkelä, Marko. 2002. Survey of bundle methods for nonsmooth optimization. *Optimization Methods and Software* **17**(1) 1–29. doi:10.1080/10556780290027828.

Mäkelä, Marko M, Napsu Karmitsa, Adil Bagirov. 2013. Subgradient and bundle methods for nonsmooth optimization. *Numerical Methods for Differential Equations, Optimization, and Technological Problems*. Springer, 275–304.

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

Minguez, R., A.J. Conejo, R. Garcia-Bertrand. 2011. Reliability and decomposition techniques to solve certain class of stochastic programming problems. *Reliability Engineering and System Safety* **96**(2) 314 – 323. doi:https://doi.org/10.1016/j.ress.2010.09.011.

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.

Moreno, Alfredo, Douglas Alem, Deisemara Ferreira, Alistair Clark. 2018. An effective two-stage stochastic multi-trip location-transportation model with social concerns in relief supply chains. *European Journal of Operational Research* **269**(3) 1050–1071. doi:https://doi.org/10.1016/j.ejor.2018.02.022.

Nesterov, Yurii. 2018. Nonsmooth convex optimization. *Lectures on Convex Optimization*. Springer, 139–240.

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

Ogryczak, Włodzimierz, Andrzej Ruszczyński. 1999. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European journal of operational research* **116**(1) 33–50.

Ogryczak, Włodzimierz, Andrzej Ruszczyński. 2001. On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming* **89**(2) 217–232.

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

Parajuli, Anubhuti, Onur Kuzgunkaya, Navneet Vidyarthi. 2017. Responsive contingency planning of capacitated supply networks under disruption risks. *Transportation Research Part E: Logistics and Transportation Review* **102** 13–37. doi:https://doi.org/10.1016/j.tre.2017.03.010.

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

Rockafellar, R. T., S. Uryasev. 2000. Optimization of conditional value-at-risk. *The Journal of Risk* **2**(3) 21–41.

Rockafellar, R. Tyrrell. 2014. *Coherent Approaches to Risk in Optimization Under Uncertainty*, chap. Chapter 3. 38–61. doi:10.1287/educ.1073.0032.

Rockafellar, R. Tyrrell, Johannes O. Royset. 2015. Engineering decisions under risk averseness. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **1**(2) 04015003. doi:10.1061/AJRUA6.0000816.

Rockafellar, R Tyrrell, Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of banking & finance* **26**(7) 1443–1471.

Ruszczyanski, Andrzej, Artur Swietanowski. 1997. Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research* **101**(2) 328–342.

Ruszczynski, A. 1986. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming* **35** 309–333.

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. 2010. Risk-averse dynamic programming for markov decision processes. *Mathematical Programming* **125**(2) 235–261.

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. *Mathematics of Operations Research* **31**(3) 544–561.

Ruszczyński, Andrzej, Alexander Shapiro. 2006b. Optimization of convex risk functions. *Mathematics of Operations Research* **31**(3) 433–452.

Samaan, N. 2004. Reliability assessment of electric power systems using genetic algorithms. Ph.D. thesis, Texas A&M Univerisity, Texas A&M Univerisity, College Station, Texas.

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.

van Ackooij, Wim, W. de Oliveira, Yongjia Song. 2019. On level regularization with normal solutions in decomposition methods for multistage stochastic programming problems. *Computational Optimization and Application* **74** 1–42.