Distributionally Robust Chance Constrained p-Hub Center Problem

Yue Zhao[1], Zhi Chen[2], and Zhenzhen Zhang[3]

[1]Institute of Operations Research and Analytics, National University of Singapore, Singapore
[2]Department of Management Sciences, City University of Hong Kong, Hong Kong SAR, China
[3]School of Economics and Management, Tongji University, Shanghai, China
yuezhao@u.nus.edu, zhi.chen@cityu.edu.hk, zhenzhenzhang222@gmail.com

The p-hub center problem is a fundamental model for the strategic design of hub location. It aims at constructing p fully interconnected hubs and links from nodes to hubs so that the longest path between any two nodes is minimized. Existing literature on the p-hub center problem under uncertainty often assumes a joint distribution of travel times, which is difficult (if not impossible) to elicit precisely. In this paper, we bridge the gap by investigating two distributionally robust chance constrained models, which cover the existing stochastic one with independent normal distribution and the sample average approximation approach as a special case, respectively. We derive deterministic reformulations as a mixed-integer program wherein a large number of constraints can be dynamically added via a constraint generation approach to accelerate computation. Extensive numerical experiments demonstrate the encouraging out-of-sample performance of our proposed models as well as the effectiveness of the constraint generation approach.

Key words: p-hub center problem; ambiguous chance constraints; Wasserstein distance; elliptical distribution; empirical distribution.

History: April 16, 2022

---

1. Introduction

The hub-and-spoke network is a centralized delivery system that identifies a set of fully interconnected hubs and allocates non-hub nodes to hubs. In this network, commodities between any pair of origin and destination (O-D) are shipped through one or multiple hubs that might be sorting centers or local warehouses. Its configuration balances the building cost, operation cost, and service level, making it a fundamental logistic model in real-world transportation systems.

In this paper, we focus on the stochastic p-hub center problem originated from Campbell (1996), where the choice of p hubs and the allocation of other nodes to the hubs are simultaneously determined in order to minimize the maximal travel time among all O-D pairs. The stochastic version of this problem is usually addressed by the chance constrained program (Hult et al. 2014, Sim et al. 2009), where the maximal travel time among all O-D pairs is guaranteed by a chance constraint that could also be interpreted as the *service-level commitment*. For example, if the risk threshold of the chance constraint is 0.95 and the maximal travel time is one day, then the resulting hub-and-spoke network based on the p-hub center model could promise one-day delivery with a probability of 0.95. Such a service-level commitment is particularly important in the booming e-commerce as there is a massive amount of packages through the delivery network and the delivery speed is considered as one of the core competitiveness of e-commerce companies (*e.g.*, Amazon offers the options of 2-Hour Delivery, Same-Day Delivery, and etc. to its Prime membership[1]).

The service-level commitment is largely affected by the uncertain travel times, which, however, is not yet fully addressed in the existing literature. The chance constrained p-hub center problem usually comes with (i) an assumption that the random travel times between nodes are independent and jointly follow a known multivariate normal distribution and (ii) individual chance constraints on the travel time of each O-D pair; see, *e.g.*, Hult et al. (2014), Sim et al. (2009). Such an approach to handling uncertainty may raise several issues. First, the normal distribution assumption remains questionable. Since transportation systems are typically affected by many factors such as weather conditions (Hao et al. 2020), normality assumption may not be able to help hedge against risks under real-world scenarios. Second, parameters of the normal distribution such as the mean and variance are assumed to be known or given for granted. In the era of big data, decision-makers could estimate such distributional information directly from data, or even seamlessly incorporate the data into the optimization models. The so-called data-driven methods are shown to have more advantages over conventional paradigms in terms of optimization methodology, statistical property, and numerical test (Mohajerin Esfahani and Kuhn 2018, Zhu et al. 2022). Finally, random travel times between nodes in the same network are not necessarily independent (or more precisely, they are more likely to be correlated) while it is hard to capture stochastic dependencies by individual chance constraints. Ignoring correlation (or stochastic dependency) in the hub-and-spoke network may profoundly change the hub location decision and result in a huge loss. For example, Shen et al. (2021) utilize data from the Storm Prediction Center of the National Oceanic and Atmospheric Administration to estimate the correlation of disruptions among 25 airports, and they demonstrate that it is important to consider the disruption correlation among hubs in order to minimize the cost in a hub-and-spoke network.

Motivated by the above points, we propose distributionally robust chance constrained models to mitigate the potential issues. To capture the correlations among travel times of all O-D pairs in the network, we introduce to the p-hub center problem the *joint chance constraint*. We then show that the nominal p-hub center problem from Hult et al. (2014) can be generalized to the elliptical distribution while the joint chance constrained p-hub center problem can be reformulated under

[1] `https://www.amazon.com/primeinsider/tips/prime-delivery-faq.html`

the empirical distribution. Although the generalization captures a wider variety of distributions, it still assumes a specific form of the underlying distribution, limiting its modeling power and potentially jeopardizing the performance under the out-of-sample test (Smith and Winkler 2006). To address this concern, we introduce the Wasserstein ambiguity set that captures a family of distributions within a spherical neighbourhood of a central reference distribution, thus optimizing over the worst-case distribution in such a set could better hedge against the uncertainty. Together with the Wasserstein ambiguity set, we proposed the distributionally robust p-hub center problem with both individual and joint chance constraints under the data-driven setting, and we derive reformulations as a mixed-integer program. To speed up the solution procedure, we propose a constraint generation approach to dynamically add the chance constraints as lazy cuts—a strategy that can accelerate the computation up to hundreds of times as shown in our numerical experiments. We further extend our framework to multiple allocations, where each node can be assigned to multiple hubs. The numerical experiments illustrate the merits of our proposed models over the current benchmark, in many aspects.

Related Literature

A fundamental problem in facility location concerns a hub-and-spoke network configuration that locates a set of fully interconnected hubs, to one of which each origin/destination is allocated. This problem, known as the hub location problem, has a long history that may be traced back to the 1960s (Hakimi 1964, 1965), and its widespread applications have been found in transportation, logistics, and telecommunication systems. Since the 1980s (O'Kelly 1986, 1987), many variants with their own emphases on different features and objectives have been developed in the literature; we refer interested readers to Snyder (2006), Campbell and O'Kelly (2012), and Correia and da Gama (2015) for comprehensive reviews. One main challenge of the hub location problem is the inevitable uncertainty arising from several sources, including setup costs (*e.g.*, the price for land acquisition), customer demands, and travel times (which directly impact transportation service and costs). Hence, ignoring uncertainty would lead to a hub-and-spoke network that is far from optimal. For a review on models and methodologies of hub location under uncertainty, we refer to Alumur et al. (2012) and Wang et al. (2020).

Among the rich variants of the hub location problem, the p-hub center problem is perhaps the most frequently studied one. It aims at constructing p fully interconnected hubs and allocating non-hub nodes to hubs so that the longest path between any two nodes is minimized. The deterministic formulation of the single-allocation p-hub center problem is first proposed in Campbell (1994) and is proved to be NP-hard by Kara and Tansel (2000). Ernst et al. (2009) extend from single allocation to multiple allocations and establish the NP-hardness. Sim et al. (2009) investigate the p-hub center problem under travel time uncertainty by modeling with chance constraints and by assuming an independent multivariate normal distribution, which is solved by a deterministic mixed-integer reformulation. Hult et al. (2014) further improve the reformulation and propose valid cuts as well as an efficient separation algorithm. Minimizing the maximal travel time of the hub-and-spoke network under the chance constraint, the p-hub center problem provides a natural interpretation of service-level commitment for delivery companies (Sim et al. 2009). Nevertheless, the independent normal assumption in the aforementioned papers may be strong. This calls for modeling paradigms, such as stochastic programming and distributionally robust optimization, to deal with modeling and optimization under travel time uncertainty—our focus in this paper.

Our work is also closely related to the literature of distributionally robust chance constrained programs. The true distribution in many real-world applications is hard to be obtained, if not possible. A remedy for this difficulty is to adopt the distributionally robust optimization framework, where the true distribution is ambiguous and can be taken most severely from an ambiguity set. It is early observed that for several kinds of ambiguity sets, distributionally robust individual chance constraints can be reformulated tractably if the random variable is assumed independent, symmetric, and bounded by utilizing concentration inequalities (Ben-Tal and Nemirovski 2000, Bertsimas and Sim 2004). Other works based on statistical results that either reformulate or approximate the distributionally robust individual chance constraints can be found, *e.g.*, in Bertsimas et al. (2018), Calafiore and Ghaoui (2006), Xu et al. (2012). Joint chance constraints, where multiple conditions are considered simultaneous, are much more involved than individual ones (Nemirovski and Shapiro 2007). We refer interested readers to Hanasusanto et al. (2015, 2017) for comprehensive studies on distributionally robust chance constrained programs.

Recently, distributionally robust chance constraints based on the Wasserstein distance have emerged, given the nice statistical properties and attractive modeling flexibility of the Wasserstein ambiguity set (Mohajerin Esfahani and Kuhn 2018). Chen et al. (2018) and Xie (2021) discover that distributionally robust individual and joint chance constrained programs under the Wasserstein ambiguity set can be reformulated as a mixed-integer problem, based on which Ho-Nguyen et al. (2021) strengthen the reformulation and provide valid inequalities. Our paper, built upon recent advances in these works, further deals with the bilinear terms of decision variables that appear in the chance constraint. To the best of our knowledge, we are the first to adopt a constraint generation approach to handle the large number of constraints in the reformulation as we observe that many of them are indeed redundant. This insight may be of independent interest for developing efficient algorithms to solve distributionally robust chance constrained programs.

Contributions

Our contributions may be summarized as follows.

1. We consider the stochastic p-hub center problem with joint chance constraint to capture possible correlations among O-D pairs. To better hedge against travel time uncertainty, we propose the distributionally robust p-hub center problem with both individual and joint chance constraints under the Wasserstein ambiguity set.
2. We derive tractable reformulations of the proposed models by exploring the structure of p-hub center problem and leveraging recent advances in distributionally robust chance constrained programs. We show that the distributionally robust individual chance constraint is equivalent to the nominal individual chance constraint with a larger risk threshold that could be efficiently found by the bisection search while the distributionally robust joint chance constraint is mixed-integer representable.
3. We propose a constraint generation approach to dynamically add those constraints (of a large number) introduced in the reformulation. As shown in our numerical experiments, the potential speedup could be up to hundreds of times and such an acceleration tends to be more significant as the number of nodes grows. This technique supports the possibility of applying our models to the real-world network of a relatively large size.
4. Numerical experiments showcase encouraging performances of the proposed distributionally robust chance constrained models, which outperform existing models in out-of-sample tests. In particular, the ability to fulfill the service-level commitment of a distributionally robust model is significantly better than that of existing models, leading to plenty of managerial insights for the decision-makers.

Organization and Notation

The remainder of the paper proceeds as follows. In Section 2 we introduce the chance constrained p-hub center problems under known distributions. We then propose distributionally robust individual and joint chance constrained models based on the Wasserstein distance in Section 3. To accelerate the solution procedure, we propose in Section 4 a constraint generation approach. Extensions to p-hub center problems under multiple allocations are discussed in Section 5. Numerical experiments are presented in Section 6, where we analyze the service-level commitment, compare the out-of-sample performance, and show the effectiveness as well as scalability of the proposed constraint generation approach.

Throughout this paper, a random variable is associated with a tilde sign. Boldface lowercase and uppercase letters represent vectors and matrices, respectively. For any positive integer N, $[N]$ denotes the set $\{1, \cdots, N\}$. We use $\|\cdot\|$ to denote a general norm, whose dual norm is $\|\cdot\|_*$. The indicator function is $\mathbb{1}[\mathcal{L}]$, which takes a value of 1 if the logical expression $\mathcal{L}$ is true and 0 otherwise. The vector of all ones of a proper size is denoted by $\mathbf{1}$.

2. p-Hub Center Problem under Uncertainty

In this section, we first introduce the deterministic p-hub center problem from Campbell (1994) and the nominal p-hub center problem under uncertainty from Hult et al. (2014) and Sim et al. (2009). We also consider the joint chance constrained program in order to include stochastic dependencies of travel times among all O-D pairs and present the nominal models under two widely used distributions—the elliptical distribution and empirical distribution.

2.1. Nominal p-Hub Center Problem

Consider a set of V nodes, denoted by $[V] = \{1, 2, \ldots, V\}$, the decision-makers aim to construct p fully interconnected hubs identified in $[V]$ and the links from each of the other nodes to *exactly one* hub so that the longest path between any pair of nodes is minimized. This problem is known as the *single allocation* p-hub center problem. Define binary variable x_{kk} with $x_{kk} = 1$ if and only if k is selected as a hub and binary variable x_{ik} taking $x_{ik} = 1$ if and only if k is a hub and node i is assigned to k. Then the feasible set of decisions is given by

$$\mathcal{X} = \left\{ \boldsymbol{x} \in \{0,1\}^{V \times V} \;\middle|\; \begin{array}{ll} \sum_{k \in [V]} x_{kk} = p & \\ \sum_{k \in [V]} x_{ik} = 1 & \forall i \in [V] \\ x_{ik} \le x_{kk} & \forall i, k \in [V] \end{array} \right\},$$

where the first constraint selects exactly p hubs, the second collection of constraints specifies that each node is assigned to exactly one hub, and the third collection of constraints describes that nodes can only be assigned to selected hubs. If nodes i and j are allocated to two hubs k and m respectively, then the transportation path from node i to node j is $i \to k \to m \to j$. Here, hubs k and m can be identical, meaning that one hub is involved in the path.

We now present the deterministic p-hub center problem. Let u_{ij} be the travel time between nodes i, j and $\alpha \in (0,1)$ be the discount factor for the inter-hub transportation that reflects the transportation economies of scale. The total travel time along the path $i \to k \to m \to j$ is then given by $v_{ikjm} = u_{ik} + \alpha u_{km} + u_{jm}$. The deterministic formulation of p-hub center problem is

$$\begin{aligned} \min_{\boldsymbol{x}, \beta} \quad & \beta \\ \text{s.t.} \quad & \beta \ge (u_{ik} + \alpha u_{km} + u_{jm}) x_{ik} x_{jm} \quad \forall i, j, k, m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \; \beta \in \mathbb{R}_+, \end{aligned} \tag{DETERMINISTIC}$$

where the objective function β is the maximal travel time among all possible pairs of nodes in the network. The objective is to design a high-quality network by minimizing the maximal travel time among all paths.

The p-hub center problem is important in many applications (*e.g.*, emergency medical services) that are sensitive to the worst-case scenarios. However, the DETERMINISTIC model assumes away the uncertainty of travel times, which might lead to inferior solutions. To address this issue, the stochastic formulation based on the chance constraint is considered as follows.

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \mathbb{P}\big[\beta \geq (\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm})x_{ik}x_{jm}\big] \geq \gamma \quad \forall i,j,k,m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+, \end{aligned} \tag{NOMINAL}$$

where $\gamma \in (0,1)$ is a given risk threshold, $\tilde{u}_{ij}$ is the random travel time between nodes i and j, and $\mathbb{P}$ is the underlying joint distribution of travel times between nodes. The NOMINAL model can be reformulated as a mixed-integer program under the assumption of independent normal distribution (which we will refer to as the normal NOMINAL model).

LEMMA 1 (proposition 1 in Hult et al. 2014). *Suppose that for any two nodes $i,j \in [V]$, the travel time $\tilde{u}_{ij}$ between them (i) follows a normal distribution with mean μ_{ij} and standard deviation σ_{ij}, and (ii) is independent of the travel time $\tilde{u}_{kl}$ for any $(k,l) \neq (i,j)$, then the normal* NOMINAL *model is equivalent to the following mixed-integer program*

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \beta \geq \sum_{k \in [V]} (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})(x_{ik} + x_{jm} - 1) \quad \forall i,j,m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+, \end{aligned} \tag{1}$$

where $\mu_{ikjm} = \mu_{ik} + \alpha\mu_{km} + \mu_{jm}$ and $\sigma_{ikjm} = \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{km}^2 + \sigma_{jm}^2}$ are the mean and standard deviation of travel time in the path $i \to k \to m \to j$, respectively, and $\Phi^{-1}(\gamma)$ is the γ-quantile of the standard normal distribution.

Proof of Lemma 1. With the independent normal distribution, for any $i,j,k,m \in [V]$ the random travel time $\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm}$ follows a normal distribution of mean μ_{ikjm} and standard deviation σ_{ikjm}. Hence, its γ-quantile amounts to $\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm}$. It then holds that

$$\mathbb{P}\big[\beta \geq (\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm})x_{ik}x_{jm}\big] \geq \gamma \quad \Longleftrightarrow \quad \beta \geq (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})x_{ik}x_{jm}.$$

Because $\beta \in \mathbb{R}_+$ and x_{ik}, x_{jm} are both binary decisions, we have

$$\beta \geq (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})x_{ik}x_{jm} \quad \Longleftrightarrow \quad \beta \geq (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})(x_{ik} + x_{jm} - 1).$$

Given $i,j,m \in [V]$ and $x \in \mathcal{X}$, we observe that

$$\begin{aligned} & \beta \geq \sum_{k \in [V]} (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})(x_{ik} + x_{jm} - 1) \\ \Longleftrightarrow \ & \beta \geq (\mu_{ikjm} + \Phi^{-1}(\gamma) \cdot \sigma_{ikjm})(x_{ik} + x_{jm} - 1) \quad \forall k \in [V], \end{aligned}$$

concluding the proof. □

Compared to the DETERMINISTIC model, the normal NOMINAL model under uncertainty has advantages from both modeling and application perspectives. First of all, the chance constrained formulation takes the uncertainty of travel times into consideration and is flexible in handling the worst-case scenarios by adjusting the risk threshold γ. In addition, the risk threshold γ has an intuitive interpretation as the service-level commitment: with a probability of at least γ, it guarantees that the travel time between any O-D pair does not exceed β. In real-world scenarios, the service-level commitment has a wide range of implications. For example, the ambulance needs to show up within 15 minutes or the package is promised to be delivered within 2 days with a high probability. However, the assumption of independent normal distribution limits the modeling power of the normal NOMINAL model. We will mitigate the issue by generalizing the result to elliptical distributions and by handling the stochastic dependency with joint chance constraints in the following two sections, respectively.

2.2. Extension to Elliptical Distribution

We extend the normal NOMINAL model to the elliptical distribution—a generalization of the multivariate normal distribution—that has been widely studied and adopted in risk management (Embrechts et al. 2002, Landsman and Valdez 2003, Pérignon and Smith 2010).

Definition 1 (Elliptical Distribution). The probability density function of an elliptical distribution is given by

$$\phi_g(\boldsymbol{u}) = \kappa \cdot g\left(\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{u}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu}$ is a location parameter, $\boldsymbol{\Sigma}$ is a positive definite matrix, κ is a positive normalization scalar, and g is a generator function. We denote the random vector $\tilde{\boldsymbol{u}}$ follows such an elliptical distribution by $\tilde{\boldsymbol{u}} \sim \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$. ♣

The elliptical distribution has a notable preservation property under linear transformation. That is, for a random vector $\tilde{\boldsymbol{u}}$ and matrices $\boldsymbol{A}, \boldsymbol{B}$ of proper sizes,

$$\tilde{\boldsymbol{u}} \sim \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)} \Longrightarrow \boldsymbol{A}\tilde{\boldsymbol{u}} + \boldsymbol{B} \sim \mathbb{P}_{(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{B},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top,g)}.$$

By this property, we could use the elliptical random vector $\tilde{\boldsymbol{\xi}} \sim \mathbb{P}_{(\boldsymbol{0},\boldsymbol{I},g)}$ to represent any random vector $\tilde{\boldsymbol{u}} \sim \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$ by $\tilde{\boldsymbol{u}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\xi}}$. In particular, when the random variable is a scalar, we have the *standard elliptical distribution* $\mathbb{P}_{(0,1,g)}$ generated by g. The cumulative distribution function of the standard elliptical distribution is

$$\Phi_g(a) = \int_{-\infty}^{a} \kappa \cdot g\left(\frac{z^2}{2}\right) \mathrm{d}z,$$

whose inverse (*i.e.*, the quantile function) we denote by Φ_g^{-1}. A commonly used example would be the standard normal distribution where the generator function is $g(x) = \exp(-x)$ and the normalization scalar is $\kappa = 1/\sqrt{2\pi}$.

EXAMPLE 1 (ELLIPTICAL TRAVEL TIMES BETWEEN O-D PAIRS). Suppose that the random travel times of all paths in the network jointly follow an elliptical distribution $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$. We denote the random travel time from i to j through their hubs k and m by

$$\tilde{v}_{ikjm} := \tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm},$$

which can be written as $\boldsymbol{a}_{ikjm}^\top \tilde{\boldsymbol{u}}$ for some vector $\boldsymbol{a}_{ikjm}$. Then the random scalar $\tilde{v}_{ikjm}$ follows an elliptical distribution $\mathbb{P}_{(\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu},\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm},g)}$ with the same generator function g, $\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} = \mu_{ik} + \alpha\mu_{km} + \mu_{jm}$, and $\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm} = \Sigma_{ik,ik} + \alpha^2\Sigma_{km,km} + \Sigma_{jm,jm} + 2\Sigma_{ik,jm} + 2\alpha(\Sigma_{ik,km} + \Sigma_{jm,km})$ with $\Sigma_{ik,jm}$ being the covariance between $\tilde{u}_{ik}$ and $\tilde{u}_{jm}$. Moreover, it is not hard to see that $(\tilde{v}_{ikjm} - \boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu})/\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}$ follows the standard elliptical distribution $\mathbb{P}_{(0,1,g)}$. Hence, its γ-quantile of $\tilde{v}_{ikjm}$ amounts to

$$\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} + \Phi_g^{-1}(\gamma)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}},$$

which is a desirable closed-form expression. ♣

As shown in Example 1, the travel time between any O-D pair also follows an elliptical distribution. Therefore, the mixed-integer reformulation (1) of the normal NOMINAL model can be naturally extended to the elliptical distribution (which we call the elliptical NOMINAL model).

LEMMA 2. *Suppose that the underlying distribution in the* NOMINAL *model is an elliptical distribution* $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$, *then the corresponding elliptical* NOMINAL *model is equivalent to the following mixed-integer program*

$$\begin{aligned}
\min_{\boldsymbol{x},\beta}\quad & \beta \\
\text{s.t.}\quad & \beta \geq \sum_{k\in[V]} \left(\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} + \Phi_g^{-1}(\gamma)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}\right)(x_{ik} + x_{jm} - 1) \quad \forall i,j,m\in[V] \\
& \boldsymbol{x}\in\mathcal{X},\ \beta\in\mathbb{R}_+,
\end{aligned}$$

where $\boldsymbol{a}_{ikjm}^\top\tilde{\boldsymbol{u}} = \tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm}$.

Proof of Lemma 2. As discussed in Example 1, the random travel time $\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm}$ still follows an elliptical distribution and its γ-quantile can be written as

$$\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} + \Phi_g^{-1}(\gamma)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}.$$

Thus for any $i,j,k,m\in[V]$, we have

$$\mathbb{P}\left[\beta \geq (\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm})x_{ik}x_{jm}\right] \geq \gamma \quad\Longleftrightarrow\quad \beta \geq \left(\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} + \Phi_g^{-1}(\gamma)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}\right)x_{ik}x_{jm}.$$

The remaining proof is similar to that of Lemma 1. □

2.3. Extension to Joint Chance Constraint

One fundamental issue of the NOMINAL model is the assumption of stochastic independence that ignores correlation (or stochastic dependency) among different O-D pairs. However, recent works show that taking the correlation of uncertainty into consideration is of vital importance under the hub-and-spoke structure. For example, Shen et al. (2021) study a reliable hub location problem with random node disruptions, and they find that the disruption correlation among hub nodes is critical to control the total cost. Likewise, in our model, the travel times among O-D pairs could be correlated and severely affect the network performance.

To capture the correlation, we can consider the following joint chance constrained problem:

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \mathbb{P}\big[\forall i,j,k,m \in [V] : \beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+. \end{aligned} \tag{2}$$

Although the problem could explicitly model the stochastic dependency, it is more difficult to solve than the NOMINAL model that consists of individual chance constraints. Even if we assume the travel time follows a multivariate normal distribution, it remains challenging to derive an exact deterministic reformulation. To see this, we rewrite the joint chance constraint as

$$\mathbb{P}\big[\beta \geq \max\{\forall i,j,k,m \in [V] : (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\}\big] \geq \gamma.$$

Suppressing $x_{ik}x_{jm}$, the probability still involves an order statistic of random components from the multivariate normal distribution, and we need to calculate its cumulative distribution function in order to solve the problem (*i.e.*, to optimize over $\boldsymbol{x}$). Fortunately, if we take $\mathbb{P}$ as the empirical distribution from data, problem (2) then admits a mixed-integer programming reformulation.

Definition 2 (Empirical Distribution). Given data samples $\{\hat{\boldsymbol{u}}_n\}_{n=1}^N$, the empirical distribution is defined as

$$\hat{\mathbb{P}} = \frac{1}{N} \sum_{n=1}^{N} \delta_{\hat{\boldsymbol{u}}_n},$$

where $\delta_{\hat{\boldsymbol{u}}_n}$ is the Dirac distribution that places a unit mass on the sample $\hat{\boldsymbol{u}}_n$. ♣

The empirical distribution is a discrete distribution that uniformly distributes on each sample. It is a key ingredient of the popular sample average approximation (SAA) in stochastic programming, wherein the underlying distribution is often assumed to be the empirical distribution; see, *e.g.*, Birge and Louveaux (2011), Ruszczyński and Shapiro (2003). Recently, the empirical distribution gains popularity in distributionally robust optimization, for being a decent central reference distribution in the data-driven Wasserstein ambiguity set that we will introduce and adopt subsequently (Chen et al. 2018, Mohajerin Esfahani and Kuhn 2018).

We next derive the tractable reformulation of the joint chance constrained p-hub center problem (2) under empirical distribution $\hat{\mathbb{P}}$, which we call the SAA model since the approach is also referred to as the sample average approximation in stochastic programming.

LEMMA 3. *Suppose that the underlying probability $\mathbb{P}$ is approximated by the empirical distribution $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n=1}^{N}\delta_{\hat{\boldsymbol{u}}_n}$ in problem (2), then the corresponding* SAA *model*

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \hat{\mathbb{P}}\big[\forall i,j,k,m \in [V] : \beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+. \end{aligned} \tag{SAA}$$

is equivalent to the following mixed-integer program

$$\begin{aligned} \min_{\boldsymbol{x},\beta,\boldsymbol{q}} \quad & \beta \\ \text{s.t.} \quad & \mathbf{1}^\top \boldsymbol{q} \leq (1-\gamma)N \\ & \beta + \mathrm{M} q_n \geq \sum_{k\in[V]} ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})(x_{ik} + x_{jm} - 1) \quad \forall n \in [N], \ i,j,m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+, \ \boldsymbol{q} \in \{0,1\}^N, \end{aligned} \tag{3}$$

where M *is a sufficiently large (but finite) number.*

Proof of Lemma 3. Using the standard technique to reformulate probabilistic constraints under discrete distribution (see, *e.g.*, Ruszczyński 2002), the SAA model is equivalent to

$$\begin{aligned} \min_{\boldsymbol{x},\beta,\boldsymbol{q}} \quad & \beta \\ \text{s.t.} \quad & \mathbf{1}^\top \boldsymbol{q} \leq (1-\gamma)N \\ & \beta + \mathrm{M} q_n \geq ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}) x_{ik} x_{jm} \quad \forall n \in [N], \ i,j,k,m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+, \ \boldsymbol{q} \in \{0,1\}^N, \end{aligned}$$

where $q_n = 1$ (resp., $q_n = 0$) corresponds to the situation when sample $\hat{\boldsymbol{u}}_n$ does not (resp., does) satisfy the chance constraint $\beta \geq (\tilde{u}_{ik} + \alpha\tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}$ $\forall i,j,k,m \in [V]$. Then $\mathbf{1}^\top \boldsymbol{q} \leq (1-\gamma)N$ guarantees that the probability of satisfying the chance constraint is not smaller than γ under the empirical distribution. Following similar arguments as in the proof of Lemma 1, we can first linearize the term $x_{ik}x_{jm}$ by $x_{ik} + x_{jm} - 1$ and then sum over $k \in [V]$ to obtain

$$\begin{aligned} & \beta + \mathrm{M} q_n \geq ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}) x_{ik} x_{jm} && \forall n \in [N], \ i,j,k,m \in [V] \\ \iff \ & \beta + \mathrm{M} q_n \geq \sum_{k\in[V]} ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})(x_{ik} + x_{jm} - 1) && \forall n \in [N], \ i,j,m \in [V]. \end{aligned}$$

The proof then completes. $\square$

Remark 1. Although in Lemma 3 we have established a tractable formulation for the joint chance constrained p-hub center problem (2) under empirical distribution, it is still hard to develop a tractable formulation under elliptical distribution. As we have discussed that the joint chance constraint under normal distribution is hard, it remains the same for elliptical distribution as we need to explore the distribution of an order statistics (that is, maximum) over many elliptical distributions. Furthermore, we would like to point out that developing the tractable reformulation for individual chance constrained p-hub center problem under empirical distribution may encounter unnecessary computation issues. To see this, following the same spirit in Lemma 3, for each individual chance constraint we need to introduce N binary variables. Hence, there will be in total $\mathcal{O}(NV^3)$ binary variables in the reformulation, leading to computational difficulties. Considering these points, in the rest of the paper we only study the *elliptical individual* chance constrained p-hub center problem and the *empirical joint* chance constrained p-hub center problem. ♣

3. Distributionally Robust p-Hub Center Models

A critical issue of the chance constrained models in the last section is the assumption of known distribution of random travel times, which, in practice, is hard (if not impossible) to precisely specify based on historical data. The elliptical distribution may not fit well with data, while the empirical distribution may result in disappointing out-of-sample performance (termed as the *optimizer's curse* in stochastic programming; see, *e.g.*, Smith and Winkler 2006). In p-hub center problem, this issue will cause the failure of service-level commitment to the customers, as the configuration of the network and the promised travel time (*i.e.*, the objective β) could be over-optimistic.

To address the issue, we will adopt the distributionally robust chance constrained program that broadly falls into the category of data-driven distributionally robust optimization (Erdoğan and Iyengar 2006, Goh and Sim 2010, Mohajerin Esfahani and Kuhn 2018, Chen et al. 2018). Instead of assuming a known distribution, the distributionally robust chance constrained models consider the worst-case distribution from a family of distributions. In particular, we will introduce the distributionally robust p-hub center problem with (i) individual chance constraints that consider individually the satisfactory probability of each path and (ii) the joint chance constraint that considers jointly the satisfactory probability of all paths. To begin with, we introduce the family of distributions based on the Wasserstein distance.

3.1. Wasserstein Ambiguity Set

The Wasserstein distance is widely adopted to measure the closeness between two distributions. It could be understood as the minimal transportation cost of moving from one distribution to the other. Formally, the Wasserstein distance between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\mathbb{R}^L$ is defined by

$$d_{\mathrm{W}}(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{P} \in \mathcal{P}(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\mathbb{P}}[\|\tilde{\boldsymbol{u}}_1 - \tilde{\boldsymbol{u}}_2\|],$$

where $\tilde{\boldsymbol{u}}_1 \sim \mathbb{P}_1, \tilde{\boldsymbol{u}}_2 \sim \mathbb{P}_2$, $\|\cdot\|$ is a general norm on $\mathbb{R}^L$ to measure the cost of moving a Dirac point mass from $\tilde{\boldsymbol{u}}_1$ to $\tilde{\boldsymbol{u}}_2$, and $\mathcal{P}(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all distributions on $\mathbb{R}^L \times \mathbb{R}^L$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$. The Wasserstein ambiguity set $\mathcal{F}(\theta)$ is then defined as

$$\mathcal{F}(\theta) = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^L) \mid d_{\mathrm{W}}(\mathbb{P}, \mathbb{P}_{\mathrm{R}}) \le \theta\}, \tag{4}$$

which is a ball of radius $\theta \ge 0$ in the probability space centered at a given reference distribution $\mathbb{P}_{\mathrm{R}}$. The choice of central reference distribution $\mathbb{P}_{\mathrm{R}}$ could be directly constructed from data, *e.g.*, the empirical distribution; or assumed to be in a certain family of parameterized distributions such as elliptical distribution and one can estimate the moment information from data. By adjusting the radius of the Wasserstein ball, the decision-maker can obtain different levels of robustness. In particular, when $\theta = 0$, the Wasserstein ambiguity set shrinks to a singleton consisting of only the central reference distribution $\mathbb{P}_{\mathrm{R}}$.

Because the computation of Wasserstein distance is generally hard (Villani 2009), it has long been believed that distributionally robust optimization with Wasserstein distance would inevitably lead to a hard global optimization problem. Recently, Mohajerin Esfahani and Kuhn (2018) find out that in many cases the distributionally robust model under a Wasserstein ambiguity set can be reformulated as tractable convex programs, and Chen et al. (2018) further establish, for the content of distributionally robust chance constraints, reformulations as a mixed-integer conic program—a technique we next leverage to reformulate our distributionally robust chance constrained models.

3.2. Distributionally Robust Individual Chance Constraint

The distributionally robust chance constrained model requires the satisfaction of the risk threshold for the worst-case distribution in an ambiguity set. Mathematically, the distributionally robust individual chance constrained (DR-ICC) programs for p-hub center problem with *individual* chance constraints read as follows.

$$\begin{aligned} \min_{\boldsymbol{x}, \beta} \quad & \beta \\ \text{s.t.} \quad & \forall\, \mathbb{P} \in \mathcal{F}(\theta) : \mathbb{P}\left[\beta \ge (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\right] \ge \gamma \quad \forall i, j, k, m \in [V] \\ & \boldsymbol{x} \in \mathcal{X}, \ \beta \in \mathbb{R}_+, \end{aligned} \tag{DR-ICC}$$

where the $\mathcal{F}(\theta)$ is the Wasserstein ambiguity set defined in (4).

Of note, this model could be viewed as a natural extension of the NOMINAL model. One can verify that when the radius of Wasserstein distance becomes zero and the central reference distribution is the normal distribution, the DR-ICC model then recovers normal NOMINAL from Hult et al. (2014). When the radius θ increases, the feasible region of DR-ICC shrinks, thus the model becomes more conservative. In other words, the minimum of maximal travel time over each O-D pair β is expected to be larger, but the service-level commitment is more likely to be satisfied.

We next derive an exact reformulation of DR-ICC under the Wasserstein ambiguity set with $\mathbb{P}_{\mathrm{R}}$ being the elliptical distribution (which we call elliptical DR-ICC model). Quite notably, the reformulation is equivalent to an elliptical NOMINAL model at a higher risk level.

THEOREM 1. *Suppose that $\mathbb{P}_{\mathrm{R}}$ in the Wasserstein ambiguity set (4) is an elliptical distribution $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$ with a positive definite matrix $\boldsymbol{\Sigma}$, and the Wasserstein distance is equipped with the Mahalanobis norm associated with $\boldsymbol{\Sigma}$ defined by $\|\boldsymbol{\xi}\| = \sqrt{\boldsymbol{\xi}^\top\boldsymbol{\Sigma}\boldsymbol{\xi}}$, then the corresponding elliptical* DR-ICC *model is equivalent to the following mixed-integer program*

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \beta \geq \sum_{k\in[V]} \left(\boldsymbol{a}_{ikjm}^\top \boldsymbol{\mu} + \Phi_g^{-1}(\gamma_{ikjm}^\star)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}\right)(x_{ik}+x_{jm}-1) \quad \forall i,j,m\in[V] \\ & \boldsymbol{x}\in\mathcal{X},\ \beta\in\mathbb{R}_+, \end{aligned} \tag{5}$$

where $\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu} = \mu_{ik}+\alpha\mu_{km}+\mu_{jm}$ and $\gamma_{ikjm}^\star = \Phi_g(\eta_{ikjm}^\star)\geq\gamma$ with

$$\eta_{ikjm}^\star = \inf_{\eta}\left\{\eta\geq\Phi_g^{-1}(\gamma) : \eta(\Phi_g(\eta)-\gamma) - \int_{(\Phi_g^{-1}(\gamma))^2/2}^{\eta^2/2}\kappa g(z)\mathrm{d}z \geq \frac{\theta\|\boldsymbol{a}_{ikjm}\|_*}{\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}}}\right\}.$$

Proof of Theorem 1. We first look at the following individual robust chance constraint

$$\forall\,\mathbb{P}\in\mathcal{F}(\theta) : \mathbb{P}\left[\beta\geq\boldsymbol{a}^\top\tilde{\boldsymbol{u}}\right]\geq\gamma,$$

for some generic coefficient vector $\boldsymbol{a}\in\mathbb{R}^L$. The above chance constraint is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\left[\beta<\boldsymbol{a}^\top\tilde{\boldsymbol{u}}\right]\leq 1-\gamma \iff \sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\left[\beta\leq\boldsymbol{a}^\top\tilde{\boldsymbol{u}}\right]\leq 1-\gamma.$$

Here, by using proposition 3 in Gao and Kleywegt (2016), it is indifferent if we replace the strict inequality on the left-hand side with a weak one on the right-hand side.

Now we introduce the worst-case value-at-risk (VaR) of the Wasserstein ambiguity set $\mathcal{F}(\theta)$ (Chen and Xie 2021). For a random scalar $\tilde{u}$,

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\text{-VaR}_\gamma[\tilde{u}] := \inf_{v\in\mathbb{R}}\left\{v : \sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}[\tilde{u}>v]\leq 1-\gamma\right\}.$$

Exploring the definition of the worst-case VaR, we note that

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\left[\beta\leq\boldsymbol{a}^\top\tilde{\boldsymbol{u}}\right]\leq 1-\gamma \iff \sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\text{-VaR}_\gamma[\boldsymbol{a}^\top\tilde{\boldsymbol{u}}-\beta]\leq 0.$$

By theorem 7 of Chen and Xie (2021), it holds that

$$\sup_{\mathbb{P}\in\mathcal{F}(\theta)}\mathbb{P}\text{-VaR}_\gamma[\boldsymbol{a}^\top\tilde{\boldsymbol{u}}-\beta] = \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\text{-VaR}_{\gamma^\star}[\boldsymbol{a}^\top\tilde{\boldsymbol{u}}-\beta],$$

where $\gamma^\star = \Phi_g(\eta^\star) \geq \gamma$ with $\eta^\star$ being the smallest $\eta \geq \Phi_g^{-1}(\gamma)$ that satisfies

$$\eta(\Phi_g(\eta) - \gamma) - \int_{(\Phi_g^{-1}(\gamma))^2/2}^{\eta^2/2} \kappa g(z)\mathrm{d}z \geq \frac{\theta\|\boldsymbol{a}\|_*}{\sqrt{\boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}}}.$$

In other words, the worst-case VaR around the elliptical distribution $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$ at the risk level γ is equal to the nominal elliptical VaR at a larger risk level $\gamma^\star \geq \gamma$. We thus obtain

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\text{-VaR}_\gamma[\boldsymbol{a}^\top \tilde{\boldsymbol{u}} - \beta] \leq 0 \quad &\Longleftrightarrow \quad \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\text{-VaR}_{\gamma^\star}[\boldsymbol{a}^\top \tilde{\boldsymbol{u}} - \beta] \leq 0 \\
&\Longleftrightarrow \quad \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\big[\beta \leq \boldsymbol{a}^\top \tilde{\boldsymbol{u}}\big] \leq 1 - \gamma^\star \\
&\Longleftrightarrow \quad \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\big[\beta \geq \boldsymbol{a}^\top \tilde{\boldsymbol{u}}\big] \geq \gamma^\star,
\end{aligned}$$

where the last equivalence follows from $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$ being a continuous distribution. Therefore, for any $i,j,k,m \in [V]$, if $x_{ik} = x_{jm} = 1$, we have

$$\begin{aligned}
& \forall\, \mathbb{P} \in \mathcal{F}(\theta): \mathbb{P}\big[\beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma \\
\Longleftrightarrow \quad & \forall\, \mathbb{P} \in \mathcal{F}(\theta): \mathbb{P}\big[\beta \geq \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}\big] \geq \gamma \\
\Longleftrightarrow \quad & \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\big[\beta \geq \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}\big] \geq \gamma^\star_{ikjm} \\
\Longleftrightarrow \quad & \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}\big[\beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma^\star_{ikjm} \\
\Longleftrightarrow \quad & \beta \geq \Big(\boldsymbol{a}_{ikjm}^\top \boldsymbol{\mu} + \Phi_g^{-1}(\gamma^\star_{ikjm}) \cdot \sqrt{\boldsymbol{a}_{ikjm}^\top \boldsymbol{\Sigma} \boldsymbol{a}_{ikjm}}\Big) x_{ik} x_{jm} \\
\Longleftrightarrow \quad & \beta \geq \Big(\boldsymbol{a}_{ikjm}^\top \boldsymbol{\mu} + \Phi_g^{-1}(\gamma^\star_{ikjm}) \cdot \sqrt{\boldsymbol{a}_{ikjm}^\top \boldsymbol{\Sigma} \boldsymbol{a}_{ikjm}}\Big) (x_{ik} + x_{jm} - 1).
\end{aligned}$$

Here, if $x_{ik} = 0$ or $x_{jm} = 0$, then both the distributionally robust chance constraint and the nominal chance constraint at a larger risk level are redundant since $\beta \geq 0$. Finally, similar to the observation in Lemma 1, notice that for any fixed $i,j,m \in [V]$ and $\boldsymbol{x} \in \mathcal{X}$, we have the below observation

$$\begin{aligned}
& \beta \geq \Big(\boldsymbol{a}_{ikjm}^\top \boldsymbol{\mu} + \Phi_g^{-1}(\gamma^\star_{ikjm}) \cdot \sqrt{\boldsymbol{a}_{ikjm}^\top \boldsymbol{\Sigma} \boldsymbol{a}_{ikjm}}\Big) (x_{ik} + x_{jm} - 1) \ \forall k \in [V] \\
\Longleftrightarrow \quad & \beta \geq \sum_{k\in[V]} \Big(\boldsymbol{a}_{ikjm}^\top \boldsymbol{\mu} + \Phi_g^{-1}(\gamma^\star_{ikjm}) \cdot \sqrt{\boldsymbol{a}_{ikjm}^\top \boldsymbol{\Sigma} \boldsymbol{a}_{ikjm}}\Big) (x_{ik} + x_{jm} - 1),
\end{aligned}$$

This concludes the proof. □

Theorem 1 helps equivalently transform the distributionally robust individual chance constraints within the Wasserstein ball to simple chance constraints with a higher risk threshold. The proof mainly utilizes the relationship between distributionally robust individual chance constraint and the worst-case value-at-risk. The remaining computational issue is how to find the risk threshold $\gamma^\star_{ikjm}$. Indeed, based on the fact that the function $f(\eta) = \eta(\Phi_g(\eta) - \gamma) - \int_{(\Phi_g^{-1}(\gamma))^2/2}^{\eta^2/2} \kappa g(z)\mathrm{d}z$ is monotonically increasing in η, $\gamma^\star_{ikjm}$ can be efficiently found by the bisection search algorithm; see more detail in Appendix B. Considering that the computational overhead of the bisection search is quite small and can be done offline, the time complexity of solving the elliptical DR-ICC model is approximately the same as that of the elliptical NOMINAL model.

3.3. Distributionally Robust Joint Chance Constraint

Although being a general framework to model the uncertainty and capture the nominal p-hub center problem as a special case, the DR-ICC model in the previous section still has a significant drawback: it neglects the correlation among different O-D pairs. Recently, Shen et al. (2021) study reliable hub location problems with random disruptions and they find it helpful to consider the disruption correlation in reducing the cost, especially when the underlying correlation is strong.

Now we introduce the *distributionally robust joint chance constrained* (DR-JCC) program for the p-hub center problem

$$\begin{aligned} \min_{\boldsymbol{x},\beta} \quad & \beta \\ \text{s.t.} \quad & \forall\, \mathbb{P}\in\mathcal{F}(\theta):\ \mathbb{P}\big[\forall i,j,k,m\in[V]:\beta\geq(\tilde{u}_{ik}+\alpha\tilde{u}_{km}+\tilde{u}_{jm})x_{ik}x_{jm}\big]\geq\gamma \\ & \boldsymbol{x}\in\mathcal{X},\ \beta\in\mathbb{R}_+. \end{aligned} \tag{DR-JCC}$$

Unlike the DR-ICC model, this problem requires a joint probability satisfaction among all O-D pairs. The service-level commitment in DR-JCC is *global* across the whole network, whereas in DR-ICC it only focuses on *local* O-D pairs.

Note that given the same ambiguity set $\mathcal{F}(\theta)$ and risk threshold γ, the DR-ICC model always has a more optimistic solution than that of the DR-JCC model. This is because that any feasible solution $(\boldsymbol{x},\beta)$ to DR-JCC is also feasible to DR-ICC, since for any $i',j',k',m'\in[V]$ we have

$$\mathbb{P}\big[\beta\geq(\tilde{u}_{i'k'}+\alpha\tilde{u}_{k'm'}+\tilde{u}_{m'j'})x_{i'k'}x_{j'm'}\big]\geq\mathbb{P}\big[\forall i,j,k,m\in[V]:\beta\geq(\tilde{u}_{ik}+\alpha\tilde{u}_{km}+\tilde{u}_{jm})x_{ik}x_{jm}\big].$$

In other words, the optimal objective β of DR-ICC will be equal to or smaller than that of DR-JCC.

For the DR-JCC model under the Wasserstein ambiguity set with $\mathbb{P}_{\mathrm{R}}$ being the empirical distribution (which we call empirical DR-JCC model), exploring the structure of the p-hub center problem, we can reformulate it as a mixed-integer program.

THEOREM 2. *Suppose that* $\mathbb{P}_{\mathrm{R}}$ *in the Wasserstein ambiguity set* (4) *is an empirical distribution* $\hat{\mathbb{P}}=\frac{1}{N}\sum_{n=1}^{N}\delta_{\hat{\boldsymbol{u}}_n}$*, then the corresponding empirical* DR-JCC *model is equivalent to the following mixed-integer program*

$$\begin{aligned} \min_{\boldsymbol{x},\beta,\boldsymbol{r},\boldsymbol{s},t,\boldsymbol{q}} \quad & \beta \\ \text{s.t.} \quad & (1-\gamma)Nt-\mathbf{1}^\top\boldsymbol{s}\geq\theta N\alpha_* \\ & \mathrm{M}(1-q_n)\geq t-s_n && \forall n\in[N] \\ & r_n+\mathrm{M}q_n\geq t-s_n && \forall n\in[N] \\ & \beta-r_n\geq\sum_{k\in[V]}([\hat{\boldsymbol{u}}_n]_{ik}+\alpha[\hat{\boldsymbol{u}}_n]_{km}+[\hat{\boldsymbol{u}}_n]_{jm})(x_{ik}+x_{jm}-1) && \forall n\in[N],\ i,j,m\in[V] \\ & \boldsymbol{x}\in\mathcal{X},\ \beta\in\mathbb{R}_+,\ \boldsymbol{r}\in\mathbb{R}^N,\ \boldsymbol{s}\in\mathbb{R}_+^N,\ t\in\mathbb{R},\ \boldsymbol{q}\in\{0,1\}^N, \end{aligned} \tag{6}$$

where $\alpha_*=\|(1,1,\alpha)\|_*$ *and* M *is a sufficiently large (but finite) number.*

Proof of Theorem 2. We only need to reformulate in DR-JCC, the distributionally robust joint chance constraint for a given $\boldsymbol{x} \in \mathcal{X}$:

$$\forall\, \mathbb{P} \in \mathcal{F}(\theta): \mathbb{P}\big[\forall i,j,k,m \in [V]: \beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma. \tag{7}$$

We first rewrite (7) as

$$\inf_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\big[\forall i,j,k,m \in [V]: \beta \geq (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) x_{ik} x_{jm}\big] \geq \gamma,$$

which involves $\mathcal{O}(V^4)$ safety conditions with indices $i,j,k,m \in [V]$. If $x_{ik} = 0$ or $x_{jm} = 0$, then the corresponding safety condition becomes $\beta \geq 0$, which is independent of the uncertainty and is naturally satisfied. With this observation, we can re-express (7) as the following one with a much less number of safety conditions:

$$\begin{aligned}
&\inf_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\big[\forall i,j,k,m \in [V] \text{ s.t. } x_{ik} = x_{jm} = 1: \beta \geq \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}\big] \geq \gamma \\
\iff &\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\big[\exists i,j,k,m \in [V] \text{ s.t. } x_{ik} = x_{jm} = 1: \beta < \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}\big] \leq 1 - \gamma.
\end{aligned}$$

Indeed, by proposition 3 of Gao and Kleywegt (2016), we can replace the strict inequalities with weak ones and obtain the following re-expression:

$$\begin{aligned}
&\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\big[\exists i,j,k,m \in [V] \text{ s.t. } x_{ik} = x_{jm} = 1: \beta \leq \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}\big] \leq 1 - \gamma \\
\iff &\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\Bigg[\tilde{\boldsymbol{u}} \in \bigcup_{i,j,k,m \in [V]: x_{ik} = x_{jm} = 1} \big\{\boldsymbol{u} \mid \beta \leq u_{ik} + \alpha u_{km} + u_{jm}\big\}\Bigg] \leq 1 - \gamma \\
\iff &\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}\Bigg[\tilde{\boldsymbol{u}} \notin \bigcap_{i,j,k,m \in [V]: x_{ik} = x_{jm} = 1} \big\{\boldsymbol{u} \mid \beta > u_{ik} + \alpha u_{km} + u_{jm}\big\}\Bigg] \leq 1 - \gamma.
\end{aligned} \tag{8}$$

By theorem 3 and lemma 2 of Chen et al. (2018), constraint (8) is satisfiable if and only if there are $\boldsymbol{s} \in \mathbb{R}_+^N$, and $t \in \mathbb{R}$ satisfying the following constraint system:

$$\begin{cases}
(1-\gamma) N t - \mathbf{1}^\top \boldsymbol{s} \geq \theta N \\
\dfrac{(\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha [\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}))^+}{\|(1,1,\alpha)\|_*} \geq t - s_n \quad \forall n \in [N],\ i,j,k,m \in [V]: x_{ik} = x_{jm} = 1.
\end{cases}$$

To eliminate the maximum operator, we could introduce binary variable $\boldsymbol{q} \in \{0,1\}^N$ and continuous variable $\boldsymbol{r} \in \mathbb{R}^N$ to re-write the constraint system equivalently to:

$$\begin{cases}
\mathrm{M}(1 - q_n) \geq t - s_n & \forall n \in [N] \\
r_n + \mathrm{M} q_n \geq t - s_n & \forall n \in [N] \\
\dfrac{\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha [\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})}{\|(1,1,\alpha)\|_*} \geq r_n & \forall n \in [N],\ i,j,k,m \in [V]: x_{ik} = x_{jm} = 1,
\end{cases}$$

for some suitably large (but finite) positive constant M. At optimality, we have $q_n = 0$ if $\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha [\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}) > 0$ $\forall i,j,k,m : x_{ik} = x_{jm} = 1$, and $q_n = 1$ otherwise.

To see that M is finite, first notice that we can set $\boldsymbol{r} \geq \boldsymbol{0}$ without changing the optimality. At optimality: (i) If $q_n = 0$, we have $\mathrm{M} \geq t - s_n$ and

$$\frac{\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})}{\|(1,1,\alpha)\|_*} \geq r_n \geq t - s_n \;\; \forall i,j,k,m \in [V] : x_{ik} = x_{jm} = 1$$

where the optimality must be obtained when the equalities hold, since otherwise we could always reduce the value of β to get a better objective value if either one of the inequalities is not tight. Therefore, we have $r_n = (\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}))/\|(1,1,\alpha)\|_* > 0$. (ii) If $q_n = 1$, we have $0 \geq r_n$, $0 \geq t - s_n$, and $r_n + \mathrm{M} \geq t - s_n$. The value of r_n does not influence the optimality, so we can set $r_n = 0$ in this case. Now, we can see that we just need a sufficiently large M such that $\mathrm{M} \geq t - s_n$ whenever $q_n = 0$ or $q_n = 1$. Since $(\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}))^+/\|(1,1,\alpha)\|_* \geq t - s_n$, setting

$$\mathrm{M} = \max_{\substack{\beta \in \mathbb{R}_+, n \in [N] \\ i,j,k,m \in [V]}} \frac{|\beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})|}{\|(1,1,\alpha)\|_*}$$

will be enough. Given a Wasserstein ambiguity set $\mathcal{F}(\theta)$ with finite $\theta > 0$, it is easy to see that β is upper bounded, thus M in the above equation is finite.

Applying the variable substitutions $(\boldsymbol{r}, \boldsymbol{s}, t) \leftarrow (\boldsymbol{r}, \boldsymbol{s}, t)/\|(1,1,\alpha)\|_*$, we see that constraint (8) is satisfiable if and only if there exist $\boldsymbol{q} \in \{0,1\}^N$, $\boldsymbol{r} \in \mathbb{R}^N_+$, $\boldsymbol{s} \in \mathbb{R}^N_+$, and $t \in \mathbb{R}$ satisfying

$$\begin{cases} (1-\gamma)Nt - \boldsymbol{1}^\top \boldsymbol{s} \geq \theta N \|(1,1,\alpha)\|_* & \\ \mathrm{M}(1 - q_n) \geq t - s_n & \forall n \in [N] \\ r_n + \mathrm{M} q_n \geq t - s_n & \forall n \in [N] \\ \beta - ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}) \geq r_n & \forall n \in [N],\; i,j,k,m \in [V] : x_{ik} = x_{jm} = 1, \end{cases}$$

Further notice that given $n \in [N]$ and $\boldsymbol{x} \in \mathcal{X}$,

$$\beta - r_n \geq [\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm} \;\; \forall i,j,k,m \in [V] : x_{ik} = x_{jm} = 1$$

is equivalent to

$$\beta - r_n \geq ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})(x_{ik} + x_{jm} - 1) \;\; \forall i,j,k,m \in [V], \tag{9}$$

which obviously holds when $x_{ik} = x_{jm} = 1$. If either $x_{ik} = 0$ or $x_{jm} = 0$, then both $\beta - r_n \geq 0$ and $\beta - r_n \geq -([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})$ are redundant, because there must exist i', j', k', m' such that $x_{i'k'} = x_{j'm'} = 1$ and $\beta - r_n \geq ([\hat{\boldsymbol{u}}_n]_{i'k'} + \alpha[\hat{\boldsymbol{u}}_n]_{k'm'} + [\hat{\boldsymbol{u}}_n]_{m'j'})$. Finally, by a similar argument as in the proof of Lemma 1, constraint (9) is equivalent to

$$\beta - r_n \geq \sum_{k \in [V]} ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})(x_{ik} + x_{jm} - 1) \;\; \forall i,j,m \in [V].$$

This concludes the proof. □

In reformulation (6), there are $\mathcal{O}(NV^3)$ constraints and $\mathcal{O}(N+V^2)$ binary decision variables. With similar treatments, we can also derive a tractable reformulation of DR-ICC with $\mathbb{P}_{\mathrm{R}}$ being the empirical distribution. However, there will be $\mathcal{O}(NV^3)$ constraints and $\mathcal{O}(NV^3)$ decision variables including the binary ones in this case. This makes the problem much computationally harder to solve, especially when the number of nodes V grows. On the contrary, in reformulation (5) of elliptical DR-ICC, there are only $\mathcal{O}(V^3)$ constraints and $\mathcal{O}(V^2)$ decision variables. We therefore omit the reformulation of DR-ICC when the central reference distribution is the empirical distribution.

4. Constraint Generation Approach

In the exact reformulation (6) of empirical DR-JCC, the number of $\mathcal{O}(NV^3)$ constraints may make the problem computationally prohibited. In this section, we propose a constraint generation approach to add these constraints dynamically as lazy constraints in the solution procedure. As a result, the model size can be greatly reduced and the computational time can be largely shortened.

For notation simplicity, we define a set $\mathcal{R} := \{(n,i,j,m) : \forall n \in [N],\ i,j,m \in [V]\}$. Then the fourth collection of constraints in problem (6) is equivalent to

$$\beta - r_n \geq \sum_{k\in[V]} ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm})(x_{ik} + x_{jm} - 1) \quad \forall (n,i,j,m) \in \mathcal{R}.$$

The constraint generation approach iteratively solves a relaxation of problem (6) that replaces $\mathcal{R}$ with its subset $\mathcal{R}'$. Whenever an optimal solution of the relaxation is obtained, we check whether it violates any constraint in the set $\mathcal{R}\setminus\mathcal{R}'$. If it does, we add the triple (i',j',m') that corresponds to the most violated constraint into $\mathcal{R}'$ to cut off the current incumbent infeasible solution and then repeat the process; otherwise the current solution is feasible and optimal to the original problem, and we terminate the solution procedure. The details are formalized as follows.

Constraint Generation

Input: $\mathcal{R}' \leftarrow \emptyset$.

1. **Solve Relaxed Problem**

    2.1. Relax the problem (6) by replacing $\mathcal{R}$ with $\mathcal{R}'$.

    2.2. Solve the relaxation to obtain a solution $(\boldsymbol{x}', \beta', \boldsymbol{r}', \cdots)$.

2. **Check Feasibility and Select Constraint**

    3.1. Given the solution $\boldsymbol{x}'$, let k_i and m_j be hubs to which nodes i and j are allocated, respectively.

    3.2. Find $\Delta_{\max} = (n',i',j') = \underset{n\in[N],i,j\in[V]}{\arg\max}\ [\hat{\boldsymbol{u}}_n]_{ik_i} + \alpha[\hat{\boldsymbol{u}}_n]_{k_i m_j} + [\hat{\boldsymbol{u}}_n]_{jm_j} - \beta' + r'_n$.

    3.3. If $\Delta_{\max} > 0$, add $(n',i',j',m_{j'})$ into $\mathcal{R}'$ and return to step 2; otherwise terminate the procedure.

Output: $(\boldsymbol{x}', \cdots)$ as the optimal solution.

The constraint generation approach can also be applied to reformulation (3) of SAA and reformulation (5) of elliptical DR-ICC, respectively. In its implementation, we use the callback function provided by the off-the-shelf solvers like CPLEX to achieve better efficiency. We will experimentally show that the constraint generation approach could save the introduction of over 99% of the total constraints in $\mathcal{R}$ and speed up the solution procedure up to hundreds of times in Section 6.

5. Extension to Multiple Allocations

So far we have been focusing on single-allocation models where each non-hub node is allocated to a unique hub. We next discuss how our method can be extended to *multiple-allocation* models where non-hub nodes can be allocated to more than one hub.

Because now one node can be allocated to multiple hubs, the path between any pair of nodes must be explicitly specified and thus a four-index formulation is introduced for the multiple-allocation p-hub center problem. Let $\boldsymbol{y} \in \{0,1\}^V$ be the vector of binary variables defined by $y_k = 1$ if and only if node k is a hub. For $i, j, k, m \in [V]$, let z_{ikjm} be a binary variable such that $z_{ikjm} = 1$ if and only if that for the O-D pair (i, j), i is allocated to hub k and j is to hub m. Here, although a non-hub node is allowed to allocate to different hubs, only one path is explicitly specified for one O-D pair. The feasible set that describes all potential designs of the hub-and-spoke network is given by

$$\mathcal{Y} = \left\{ \begin{array}{l|ll} & \sum_{k\in[V]} y_k = p & \\ \boldsymbol{y} \in \{0,1\}^V & \sum_{k\in[V]}\sum_{m\in[V]} z_{ikjm} = 1 & \forall i,j \in [V] \\ \boldsymbol{z} \in \{0,1\}^{V\times V\times V\times V} & \sum_{m\in[V]} z_{ikjm} \le y_k & \forall i,j,k \in [V] \\ & \sum_{k\in[V]} z_{ikjm} \le y_m & \forall i,j,m \in [V] \end{array} \right\}.$$

Here, the first constraint restricts to locate exactly p hubs, the second collection of constraints ensures that there is one and only one path between any O-D pair, and the third and fourth group of constraints guarantee that any O-D pair can only be connected through hubs. The deterministic p-hub center problem under multiple allocations (Ernst et al. 2009) is then captured by

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{z},\beta} \quad & \beta \\ \text{s.t.} \quad & \beta \ge \sum_{k\in[V]}\sum_{m\in[V]} (u_{ik} + \alpha u_{km} + u_{jm}) z_{ikjm} \quad \forall i,j \in [V] \\ & (\boldsymbol{y},\boldsymbol{z}) \in \mathcal{Y}, \ \beta \in \mathbb{R}_+, \end{aligned}$$

where the binary decision $\boldsymbol{z}$ could be relaxed to $\boldsymbol{z} \in [0,1]^{V\times V\times V\times V}$ without changing the optimality (proposition 3.1 in Ernst et al. 2009). The reasoning is as follows. At optimality, a positive $z_{ikjm} > 0$

represents a path between an O-D pair (i,j) and is associated with a travel cost $u_{ik}+\alpha u_{km}+u_{jm}$. Given an optimal solution with $z_{ikjm}>0$, suppose there exists (k',m') such that $k'\neq k$, $m'\neq m$, and $z_{ik'jm'}>0$, then we must have $u_{ik}+\alpha u_{km}+u_{jm}=u_{ik'}+\alpha u_{k'm'}+u_{m'j}$, because otherwise the optimal objective value can be lower if giving more weights to the path with less travel cost. Therefore, we could reassign the value of z_{ikjm} to be 1 and others to be 0 without changing the optimal objective value.

Similar to the previous sections, we can derive the mixed-integer reformulations for the SAA, elliptical DR-ICC, and empirical DR-JCC models under multiple allocations.

PROPOSITION 1. *Suppose that the underlying distribution $\mathbb{P}$ is approximated by the empirical distribution $\hat{\mathbb{P}}=\frac{1}{N}\sum_{n=1}^{N}\delta_{\hat{\boldsymbol{u}}_n}$, then the SAA model of the joint chance constrained p-hub center problem under multiple allocations*

$$\begin{aligned}\min_{\boldsymbol{y},\boldsymbol{z},\beta}\quad & \beta\\ \text{s.t.}\quad & \hat{\mathbb{P}}\Big[\forall i,j\in[V]:\beta\geq\sum_{k\in[V]}\sum_{m\in[V]}(\tilde{u}_{ik}+\alpha\tilde{u}_{km}+\tilde{u}_{jm})z_{ikjm}\Big]\geq\gamma\\ & (\boldsymbol{y},\boldsymbol{z})\in\mathcal{Y},\ \beta\in\mathbb{R}_+,\end{aligned}$$

is equivalent to the following mixed-integer program

$$\begin{aligned}\min_{\boldsymbol{y},\boldsymbol{z},\beta,\boldsymbol{q}}\quad & \beta\\ \text{s.t.}\quad & \mathbf{1}^\top\boldsymbol{q}\leq(1-\gamma)N\\ & \beta+\mathrm{M}q_n\geq\sum_{k\in[V]}\sum_{m\in[V]}([\hat{\boldsymbol{u}}_n]_{ik}+\alpha[\hat{\boldsymbol{u}}_n]_{km}+[\hat{\boldsymbol{u}}_n]_{jm})z_{ikjm}\quad\forall n\in[N],\ i,j\in[V]\\ & (\boldsymbol{y},\boldsymbol{z})\in\mathcal{Y},\ \beta\in\mathbb{R}_+,\ \boldsymbol{q}\in\{0,1\}^N,\end{aligned}$$

where M *is a sufficiently large (but finite) number.*

PROPOSITION 2. *Suppose that $\mathbb{P}_{\mathrm{R}}$ in the Wasserstein ambiguity set (4) is an elliptical distribution $\mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma},g)}$ with a positive definite matrix $\boldsymbol{\Sigma}$, and the Wasserstein distance is equipped with the Mahalanobis norm associated with $\boldsymbol{\Sigma}$ defined by $\|\boldsymbol{\xi}\|=\sqrt{\boldsymbol{\xi}^\top\boldsymbol{\Sigma}\boldsymbol{\xi}}$, then the corresponding DR-ICC p-hub center problem under multiple allocations*

$$\begin{aligned}\min_{\boldsymbol{y},\boldsymbol{z},\beta}\quad & \beta\\ \text{s.t.}\quad & \forall\,\mathbb{P}\in\mathcal{F}(\theta):\mathbb{P}\Big[\beta\geq\sum_{k\in[V]}\sum_{m\in[V]}(\tilde{u}_{ik}+\alpha\tilde{u}_{km}+\tilde{u}_{jm})z_{ikjm}\Big]\geq\gamma\quad\forall i,j\in[V]\\ & (\boldsymbol{y},\boldsymbol{z})\in\mathcal{Y},\ \beta\in\mathbb{R}_+,\end{aligned}$$

is equivalent to the following mixed-integer program

$$\begin{aligned}\min_{\boldsymbol{y},\boldsymbol{z},\beta}\quad & \beta\\ \text{s.t.}\quad & \beta\geq\sum_{k\in[V]}\sum_{m\in[V]}(\boldsymbol{a}_{ikjm}^\top\boldsymbol{\mu}+\Phi_g^{-1}(\gamma_{ikjm}^\star)\cdot\sqrt{\boldsymbol{a}_{ikjm}^\top\boldsymbol{\Sigma}\boldsymbol{a}_{ikjm}})z_{ikjm}\quad\forall i,j\in[V]\\ & (\boldsymbol{y},\boldsymbol{z})\in\mathcal{Y},\ \beta\in\mathbb{R}_+,\end{aligned}$$

where $\boldsymbol{a}_{ikjm}^\top \tilde{\boldsymbol{u}} = \tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}$ *and* $\gamma_{ikjm}^\star = \Phi_g(\eta_{ikjm}^\star) \geq \gamma$ *with*

$$\eta_{ikjm}^\star = \inf_\eta \left\{ \eta \geq \Phi_g^{-1}(\gamma) : \eta(\Phi_g(\eta) - \gamma) - \int_{(\Phi_g^{-1}(\gamma))^2/2}^{\eta^2/2} \kappa g(z) \mathrm{d}z \geq \frac{\theta \|\boldsymbol{a}_{ikjm}\|_*}{\sqrt{\boldsymbol{a}_{ikjm}^\top \boldsymbol{\Sigma} \boldsymbol{a}_{ikjm}}} \right\}.$$

Proposition 3. *Suppose that* $\mathbb{P}_{\mathrm{R}}$ *in the Wasserstein ambiguity set* (4) *is an empirical* $\hat{\mathbb{P}} = \frac{1}{N}\sum_{n=1}^N \delta_{\hat{\boldsymbol{u}}_n}$, *then the corresponding DR-JCC p-hub center problem under multiple allocations*

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{z},\beta} \quad & \beta \\ \text{s.t.} \quad & \forall\, \mathbb{P} \in \mathcal{F}(\theta) : \mathbb{P}\left[\forall i, j \in [V] : \beta \geq \sum_{k\in[V]} \sum_{m\in[V]} (\tilde{u}_{ik} + \alpha \tilde{u}_{km} + \tilde{u}_{jm}) z_{ikjm}\right] \geq \gamma \\ & (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y}, \ \beta \in \mathbb{R}_+, \end{aligned}$$

is equivalent to the following mixed-integer linear program

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{z},\beta,\boldsymbol{r},\boldsymbol{s},t,\boldsymbol{q}} \quad & \beta \\ \text{s.t.} \quad & (1-\gamma)Nt - \mathbf{1}^\top \boldsymbol{s} \geq \theta N \alpha_* \\ & \mathrm{M}(1 - q_n) \geq t - s_n && \forall n \in [N] \\ & r_n + \mathrm{M} q_n \geq t - s_n && \forall n \in [N] \\ & \beta - r_n \geq \sum_{k\in[V]} \sum_{m\in[V]} ([\hat{\boldsymbol{u}}_n]_{ik} + \alpha [\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}) z_{ikjm} && \forall n \in [N],\ i, j \in [V] \\ & (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y}, \ \beta \in \mathbb{R}_+, \ \boldsymbol{r} \in \mathbb{R}^N, \ \boldsymbol{s} \in \mathbb{R}_+^N, \ t \in \mathbb{R}, \ \boldsymbol{q} \in \{0,1\}^N, \end{aligned}$$

where $\alpha_* = \|(1,1,\alpha)\|_*$ *and* M *is a sufficiently large (but finite) number.*

Proofs of Proposition 1 to Proposition 3 are similar to those in the single-allocation counterparts and are thus omitted. The proposed constraint generation approach in Section 4 can be readily applied to all reformulations herein, and the binary decision variables $\boldsymbol{z}$ could be relaxed to be continuous by a similar argument as that in the deterministic case. Compared with the single-allocation models, there are fewer constraints but more decision variables in the multiple-allocation models. We summarize the comparison in Table 1.

6. Numerical Results

We demonstrate the numerical results of the distributionally robust chance constrained models on the widely used Civil Aeronautic Board (CAB) data set[2] for hub location and travel time. The travel distance from CAB is set to be the mean travel time for the underlying true distribution. All numerical results were produced on an Intel Xeon 3.00GHz processor with 32GB memory using Java Openjdk 15.0.2.

[2] Available at `https://www.researchgate.net/publication/269396247_cab100_mok`.

	Single Allocation			Multiple Allocations		
	SAA	DR-ICC	DR-JCC	SAA	DR-ICC	DR-JCC
constraint	$\mathcal{O}(NV^3)$	$\mathcal{O}(V^3)$	$\mathcal{O}(NV^3)$	$\mathcal{O}(NV^2)$	$\mathcal{O}(V^2)$	$\mathcal{O}(NV^2)$
binary decision	$\mathcal{O}(N+V^2)$	$\mathcal{O}(V^2)$	$\mathcal{O}(N+V^2)$	$\mathcal{O}(N+V)$	$\mathcal{O}(V)$	$\mathcal{O}(N+V)$
continuous decision	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(V^4)$	$\mathcal{O}(V^4)$	$\mathcal{O}(V^4)$

Table 1 Numbers of constraints, binary decision variables, and continuous decision variables.

We evaluate the following four p-hub center models: the normal NOMINAL model from Hult et al. (2014) with reformulation (1), the SAA model with reformulation (3), the elliptical DR-ICC model with reformulation (5), and the empirical DR-JCC model with reformulation (6). The constraint generation approach in Section 4 is applied to solve all the models. For ease of expression, we denote these four models as Hult, SAA, ICC, and JCC, respectively. Note that Hult is a special case of ICC when the radius of Wasserstein ball $\theta = 0$, and SAA is also a special case of JCC when $\theta = 0$. In the ICC model, the elliptical distribution is set to be the multivariate normal distribution with parameters estimated from the training data by a standard estimation process. We use the same risk threshold γ for both ICC and JCC models, but their meanings are different: the former is the individual probability and the latter is the joint probability, therefore we should expect more optimistic solutions of ICC (*i.e.,* smaller value of β) compared to those of JCC.

The experiment follows the data-driven scheme. Assuming an underlying true distribution, we first generate $N = 30$ training samples to solve each model, then we evaluate its optimal solution based on $N' = 10000$ testing samples for the out-of-sample performance. The procedure is repeated in 50 trials with different random seeds in order to eliminate the random bias. Let $\mathcal{E}$ denote the set of total edges, we first set the mean value of travel time $\boldsymbol{\mu} \in \mathbb{R}^{|\mathcal{E}|}$ as the distance in the CAB dataset, and the standard deviation $\boldsymbol{\sigma} = \rho\boldsymbol{\mu}$ where $\rho \in \{0.25, 0.5\}$ is a parameter to control the magnitude of deviation. We then construct a multivariate normal distribution $\tilde{\boldsymbol{u}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the underlying true distribution, where $\boldsymbol{\Sigma} \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$ with components $\Sigma_{ee} = \sigma_e^2 \ \forall e \in \mathcal{E}$ and $\Sigma_{e_1 e_2} = 0.5\sigma_{e_1}\sigma_{e_2} \ \forall e_1 \neq e_2 \in \mathcal{E}$. The travel time is truncated at 0, *i.e.*, a negative value is set to be 0.

6.1. Trade-off between Service-level Commitment Rate and Objective Value

A natural way to evaluate the performance is to compare the optimal value β obtained in each model. However, as the *optimizer's curse* (Smith and Winkler 2006) suggests, the performance of a solution in the out-of-sample tests is usually inferior to that in the optimization stage, resulting in post-decision disappointment. Indeed, this phenomenon is also observed in our numerical experiment that the optimal objective value of SAA is usually too optimistic. We thus propose the *service-level commitment rate* below to measure the out-of-sample performance.

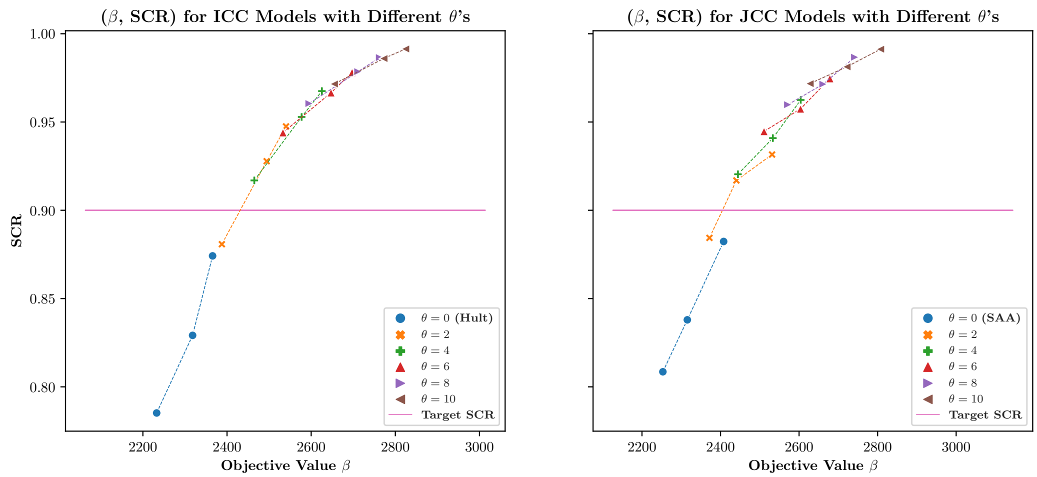

Figure 1 Objective value β and SCR ($\rho = 0.25$ and $\gamma = 0.9$).

Suppose the model solves an optimal solution $\boldsymbol{x}$, let

$$\mathcal{E}(\boldsymbol{x}) := \{(i,j,k,m) : x_{ik} = x_{jm} = 1\ \forall i,j,k,m \in [V]\}$$

be the set of paths among all O-D pairs. We use $e := (i,j,k,m) \in \mathcal{E}(\boldsymbol{x})$ to represent a path, the travel time in e for sample n is then denoted by $[\hat{\boldsymbol{u}}_n]_e = [\hat{\boldsymbol{u}}_n]_{ik} + \alpha[\hat{\boldsymbol{u}}_n]_{km} + [\hat{\boldsymbol{u}}_n]_{jm}$. The service-level commitment rate (SCR) is the probability that the delivery is within the promised time. Suppose the model solves to an optimal solution of $\boldsymbol{x}$ and β, then the corresponding SCR evaluated in the testing samples is defined by

$$\frac{1}{N'} \sum_{n \in [N']} \mathbb{1}[\beta \geq [\hat{\boldsymbol{u}}_n]_e,\ \forall e \in \mathcal{E}(\boldsymbol{x})],$$

where N' is the number of testing samples. For the final comparison of SCR, we solve the model in 50 trials with different random seeds and compute the average.

We depict the SCR and objective value β for ICC and JCC models with different θ's in Figures 1 and 2. Each line in one color represents the performance of the solution with a different value of θ (when $\theta = 0$, ICC recovers Hult and JCC reduces to SAA). For example, a point in the left plot of Figure 1, say the blue dot, represents a solution of ICC with $\theta = 0$, whose horizontal value is the corresponding objective value β and vertical value is the corresponding SCR; thus its coordinate is $(\beta, \mathrm{SCR}(\boldsymbol{x}))$. Similarly, the coordinate of a point in the right plot will be $(\beta, \mathrm{SCR}(\boldsymbol{x}))$ for a solution of the JCC model. For better illustration, we solve each model in 50 trials with different random seeds and choose $0.25, 0.5, 0.75$-quantiles of both β and SCR to present each model.

It can be observed that the larger θ is, the more conservative the model will be. In other words, if a large θ is chosen, the maximal travel time promised by the decision-maker will be large but SCR

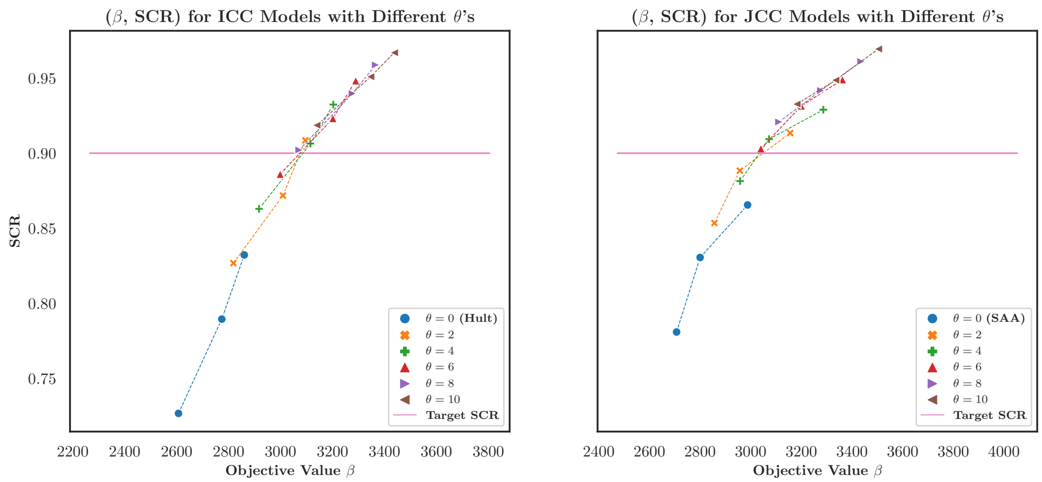

Figure 2 Objective value β and SCR ($\rho = 0.5$ and $\gamma = 0.9$).

will be high, meaning that the promised delivery time will be satisfied at a higher level. Comparing Figures 1 and 2 (the former std/mean ratio is $\rho = 0.25$ and the latter is $\rho = 0.5$), we find that if the data has a large variance, the model should be chosen to be more robust in order to meet the service-level commitment. We have also conducted experiments for larger risk thresholds. The main message conveyed from the experiment remains the same, thus we omit the result.

From the above results, we have the following managerial insights for decision-makers. (i) There is a trade-off between SCR and objective value β (*e.g.*, this could be the promised maximal delivery time): when the model becomes more robust (*i.e.*, θ becomes larger), the SCR will be large, which could mean the promised delivery time to the customer is more likely to be met, but that promised time will also be larger. (ii) In terms of SCR, both Hult and SAA models are over-optimistic as they are both under the line of the target SCR. This could result in failures of meeting the promised delivery time, whereas a slightly robust ICC or JCC model could make a decent balance between SCR and promised delivery time. (iii) Although the ICC model is expected to be less robust than the JCC model, it still performs quite well in terms of SCR when θ is selected properly, *e.g.*, choose $\theta = 6$ in the left plot of Figure 2. (iv) From the comparison between Figure 1 where $\rho = 0.25$ and Figure 2 where $\rho = 0.5$, the larger the variance is, the larger the θ should be, in order to meet the promised delivery time.

6.2. Comparison of Out-of-Sample Performance

After showing the trade-off between SCR and objective value, we can see that Hult or SAA is always over-optimistic and fails to meet the service-level commitment, whereas JCC could balance this by choosing an appropriate θ. It is demonstrated that JCC is the most robust in terms of meeting SCR. In this section, we will further show the comparison of out-of-sample travel times of different models and illustrate that JCC indeed outperforms other models in this aspect.

Out-of-sample travel time statistics

We calculate the following three quantities

$$\begin{aligned}
\mathcal{A}_n(\boldsymbol{x}) &:= \frac{1}{|\mathcal{E}(\boldsymbol{x})|}\sum_{e\in\mathcal{E}(\boldsymbol{x})}[\hat{\boldsymbol{u}}_n]_e && \forall n\in[N'] \\
\mathcal{M}_n(\boldsymbol{x}) &:= \max_{e\in\mathcal{E}(\boldsymbol{x})}\,[\hat{\boldsymbol{u}}_n]_e && \forall n\in[N'] \\
\mathcal{Q}_n(\boldsymbol{x}) &:= \gamma\text{-quantile of } \{[\hat{\boldsymbol{u}}_n]_e : e\in\mathcal{E}(\boldsymbol{x})\} && \forall n\in[N'],
\end{aligned} \tag{10}$$

which are the average, maximum, and γ-quantile of the out-of-sample travel times over all paths in $\mathcal{E}(\boldsymbol{x})$. We then compute the frequency of one model is better or equal to another over all of the test instances. Let $\boldsymbol{x}^{\text{Hult}}, \boldsymbol{x}^{\text{SAA}}$, and $\boldsymbol{x}^{\text{ICC}}, \boldsymbol{x}^{\text{JCC}}$ denote the optimal solution of each model. Take the comparison between ICC model and JCC model for example, the frequencies are

$$\begin{aligned}
\mathcal{A}(\text{JCC/ICC}) &:= \frac{1}{N'}\sum_{n\in[N']}\mathbb{1}[\mathcal{A}_n(\boldsymbol{x}^{\text{JCC}}) \le \mathcal{A}_n(\boldsymbol{x}^{\text{ICC}})] \\
\mathcal{M}(\text{JCC/ICC}) &:= \frac{1}{N'}\sum_{n\in[N']}\mathbb{1}[\mathcal{M}_n(\boldsymbol{x}^{\text{JCC}}) \le \mathcal{M}_n(\boldsymbol{x}^{\text{ICC}})] \\
\mathcal{Q}(\text{JCC/ICC}) &:= \frac{1}{N'}\sum_{n\in[N']}\mathbb{1}[\mathcal{Q}_n(\boldsymbol{x}^{\text{JCC}}) \le \mathcal{Q}_n(\boldsymbol{x}^{\text{ICC}})].
\end{aligned} \tag{11}$$

For the final comparison of these metrics, we also consider 50 trials and compute the average of them. If the value of $\mathcal{A}(\text{JCC/ICC})$ is larger than 0.5, we then say JCC is better than ICC for out-of-sample tests in terms of the average travel time, otherwise ICC is better than JCC.

In Table 2, we show the comparison of out-of-sample travel time statistics defined in (11). To get a fair comparison, we fix in both ICC and JCC models $\theta = 6$ across different parameter settings. The first column specifies different settings of $\gamma \in \{0.9, 0.95\}$ and $\rho \in \{0.25, 0.5\}$. It can be observed that: (*i*) JCC model consistently performs better than others over different combinations of (ρ, γ) and comparison metrics; (*ii*) ICC is slightly better than SAA in most cases and significantly better than Hult, especially when the std/mean ratio ρ increases; (*iii*) SAA also performs better than Hult; (*iv*) Among comparison metrics, the distributionally robust JCC and ICC models have an outstanding performance under the maximum travel time metric, suggesting the advantage of distributionally robust models in improving the maximum travel time of the network.[3]

We present the computational time of solving each model over different parameter settings in Table 3. ICC and Hult are the most computationally efficient models which are solved in a few seconds. Although JCC is more time-consuming than other models, its computational time is still within an acceptable level that is less than half a minute.

[3] We conduct additional experiments with other commonly used distributions (lognormal, uniform, and two-point distributions). We relegate the results to Appendix A as we do not observe any significant qualitative difference.

(ρ, γ)	metric	JCC/ICC	JCC/SAA	JCC/Hult	ICC/SAA	ICC/Hult	SAA/Hult
(0.25, 0.9)	$\mathcal{A}$	0.646	0.677	0.672	0.555	0.621	0.625
(0.25, 0.9)	$\mathcal{M}$	0.725	0.740	0.729	0.653	0.814	0.642
(0.25, 0.9)	$\mathcal{Q}$	0.636	0.657	0.665	0.549	0.638	0.639
(0.25, 0.95)	$\mathcal{A}$	0.584	0.633	0.588	0.508	0.666	0.587
(0.25, 0.95)	$\mathcal{M}$	0.658	0.687	0.671	0.625	0.852	0.642
(0.25, 0.95)	$\mathcal{Q}$	0.583	0.623	0.603	0.524	0.708	0.606
(0.5, 0.9)	$\mathcal{A}$	0.665	0.720	0.676	0.573	0.794	0.602
(0.5, 0.9)	$\mathcal{M}$	0.644	0.736	0.648	0.616	0.875	0.596
(0.5, 0.9)	$\mathcal{Q}$	0.642	0.708	0.650	0.588	0.797	0.589
(0.5, 0.95)	$\mathcal{A}$	0.570	0.647	0.553	0.509	0.777	0.582
(0.5, 0.95)	$\mathcal{M}$	0.605	0.709	0.597	0.591	0.858	0.582
(0.5, 0.95)	$\mathcal{Q}$	0.570	0.674	0.559	0.540	0.798	0.573

Table 2 Out-of-sample performance of all models (25 nodes and 3 hubs).

(ρ, γ)	(0.25, 0.9)	(0.25, 0.95)	(0.5, 0.9)	(0.5, 0.95)
JCC	13.58	3.57	17.84	4.36
ICC	0.71	0.70	0.71	0.70
SAA	4.10	2.46	5.04	2.76
Hult	0.72	0.70	0.68	0.70

Table 3 Average CPU time (seconds) of solving each model over 50 random instances with the multivariate normal distribution (25 nodes and 3 hubs).

We also evaluate the performance when the total number of nodes varies from $V = 25$. We set the number of nodes in $\{10, 15, 20, 25, 30, 35\}$, fix $(\rho, \gamma) = (0.5, 0.9)$ and consider 3 hubs. The results are demonstrated in Table 4. We do not observe any significant qualitative change except that when there are more nodes, SAA might slightly beat ICC.

Out-of-sample travel time density plots

In Figures 3 and 4, we present the density plot (by kernel smooth estimation) of the out-of-sample travel time over all testing samples in all 50 trials. To compare the performance based on the density plot, we note that if the density plot of model A is to the left of that of model B, then with the same probability model A has a smaller travel time than model B. We could see that the

V	JCC/ICC	JCC/SAA	JCC/Hult	ICC/SAA	ICC/Hult	SAA/Hult
10	0.649	0.834	0.645	0.647	0.850	0.609
15	0.627	0.754	0.634	0.586	0.861	0.613
20	0.574	0.722	0.599	0.504	0.791	0.631
25	0.642	0.708	0.650	0.588	0.797	0.589
30	0.661	0.659	0.679	0.450	0.632	0.606
35	0.626	0.627	0.643	0.418	0.661	0.613

Table 4 Comparison metrics $\mathcal{Q}$ for different number of nodes.

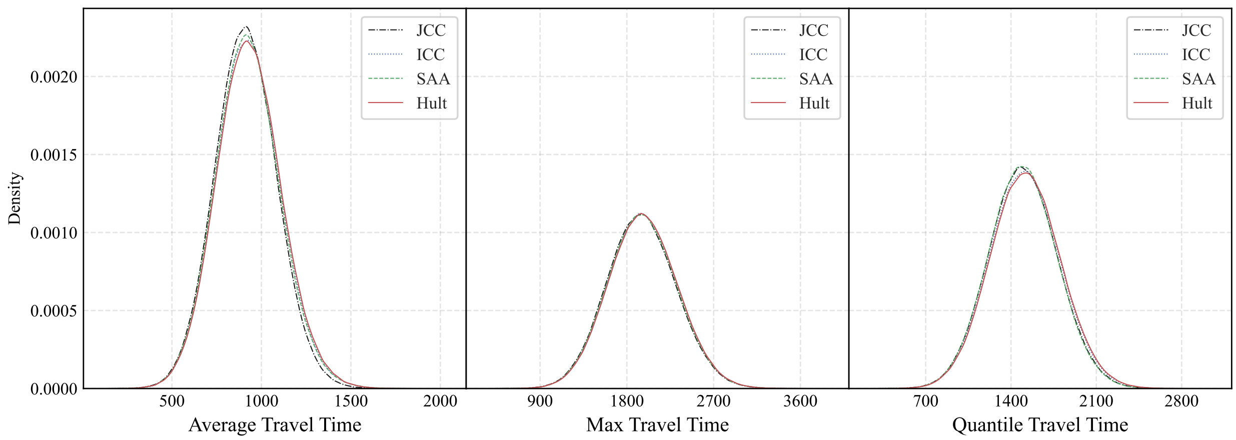

Figure 3 Density plot under the multivariate normal distribution ($\rho = 0.25$ and $\gamma = 0.9$).

density plot of JCC (resp., Hult) is leftmost (resp., rightmost), while those of ICC and SAA are quite close. Therefore, the conclusion from the density plots is similar to that from the comparison metrics, although the difference among these models is not quite large.

6.3. Effectiveness and Scalability of Constraint Generation

In this section, we test the effectiveness and scalability of our constraint generation approach. We further denote the model solved by the constraint generation approach as ICC-CG, SAA-CG, and JCC-CG, and compare them with the basic ICC, SAA, and JCC models without using the constraint generation approach. The number of constraints is $V^3/2 + V^2/2 - V$ in the basic ICC model, and is $N(V^3/2 + V^2/2 - V)$ in SAA and JCC models. Since the computational time of the Hult model is similar to that of ICC (as shown in Table 3), we only evaluate ICC in this section.

The effectiveness of the constraint generation approach is demonstrated in Table 5. We set the maximal CPU time to 1800 seconds, (ρ, γ) to $(0.5, 0.9)$, and the number of hubs to $p = 3$. The result suggests the notable speedup of our constraint generation approach. For example, it solves the

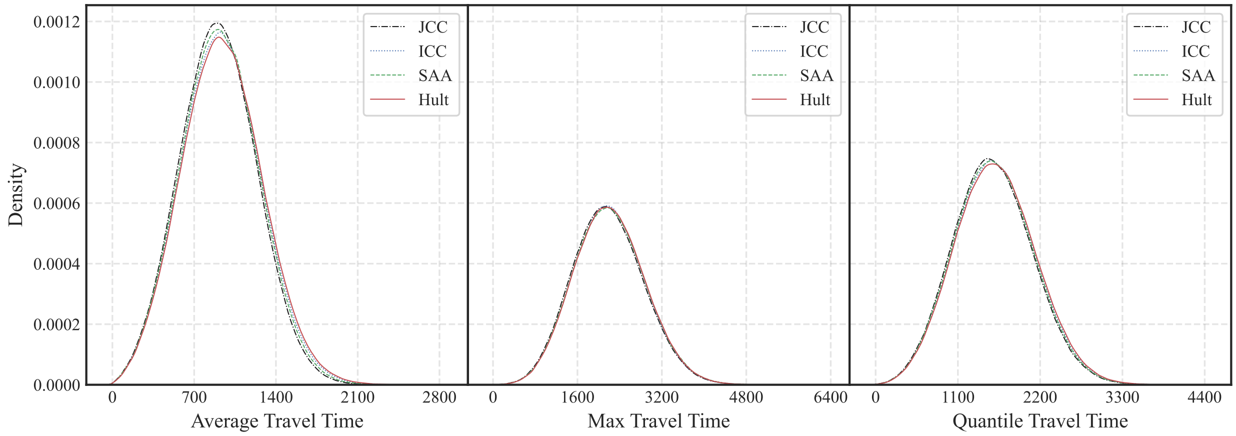

Figure 4 Density plot under the multivariate normal distribution ($\rho = 0.5$ and $\gamma = 0.9$).

	ICC	ICC-CG		SAA	SAA-CG		JCC	JCC-CG	
V	CPU time	CPU time	ratio	CPU time	CPU time	ratio	CPU time	CPU time	ratio
10	0.13	0.12	16.49%	5.57	0.16	1.60%	7.05	0.19	1.44%
15	0.57	0.43	11.03%	31.79	0.45	0.91%	82.81	1.04	0.84%
20	1.85	1.28	9.57%	265.67	1.93	0.76%	1353.54	5.91	0.67%
25	7.47	2.50	6.95%	1516.41	5.04	0.56%	1800	17.84	0.49%
30	13.68	5.75	4.00%	1800	6.73	0.35%	1800	11.46	0.29%
35	31.82	10.83	3.43%	1800	16.21	0.31%	1800	28.00	0.28%

Table 5 Average CPU time (seconds) and the ratio between the number of constraints evaluated with/without constraint generation approach over 50 trials.

SAA model **34** (resp., **300**) times faster than the basic formulation when $V = 10$ (resp., $V = 25$). The main reason for such acceleration is that through the constraint generation approach, much fewer constraints are introduced. In fact, more than 99% of the total constraints are redundant for solving the optimum when $V \geq 15$ in both SAA and JCC models. The acceleration is more significant as the number of nodes grows. Similar acceleration for the ICC model is also observed, although not that significant because ICC has fewer constraints and variables.

We also demonstrate the scalability of our constrained generation approach for each model in Table 6. We set the total number of nodes from $\{40, 50, \cdots, 100\}$ and the number of hubs from $\{4, 5, \cdots, 10\}$. We set the maximal CPU time to 1800 seconds and (ρ, γ) to $(0.5, 0.9)$. As expected, the ICC model has the best scalability and it could be solved to near optimality (less than 1% relative gap) when the total numbers of nodes and hubs reach 80 and 8, respectively. Even if we utilize the full-size CAB dataset (*i.e.*, $V = 100$ nodes), ICC can still be solved to solutions with a

	ICC-CG		SAA-CG		JCC-CG	
$(V,\ p)$	CPU time	relative gap	CPU time	relative gap	CPU time	relative gap
(40, 4)	10.13	0.00%	285.96	0.00%	725.59	0.05%
(50, 5)	27.32	0.00%	1404.42	3.59%	1736.07	9.00%
(60, 6)	201.07	0.00%	1800	47.97%	1800	43.70%
(70, 7)	558.79	0.02%	1800	48.90%	1800	49.79%
(80, 8)	1422.41	0.71%	1800	65.13%	1800	58.53%
(90, 9)	1800	21.96%	1800	100.00%	1800	100.00%
(100, 10)	1800	17.82%	1800	100.00%	1800	100.00%

Table 6 Average CPU time (seconds) and relative gap over 50 trials.

relative gap of around 20% within 30 minutes. On the other hand, both SAA and JCC models do not scale well. When the number of nodes reaches 60, the relative gap is already over 40%, when there are 90 nodes or more, it cannot even narrow the relative gap to 99.99% within the time limit.

7. Conclusion

In the paper, we address the p-hub center problem under uncertainty by distributionally robust chance constrained approaches. We first study individual chance constraints under the Wasserstein ambiguity set around an elliptical distribution, which includes the benchmark model in Hult et al. (2014) as a special case. We further investigate the joint chance constraint under the Wasserstein ambiguity set around the data-driven empirical distribution. Numerical results of the proposed models are constantly better than the existing methods under various settings. In terms of computation, we further propose a constraint generation approach that could help speed up the solution procedure up to hundreds of times.

We believe there are many aspects to explore in the future. From the modeling perspective, one can consider the transition waiting time or scheduling problem in the hub. From the computation perspective, it is interesting to establish some theoretical guarantees for the constraint generation approach, especially in the data-driven setting.

References

Alumur, Sibel, Stefan Nickel, Francisco Saldanha-da Gama. 2012. Hub location under uncertainty. *Transportation Research Part B: Methodological* **46**(4) 529–543.

Ben-Tal, Aharon, Arkadi Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming* **88**(3) 411–424.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2018. Data-driven robust optimization. *Mathematical Programming* **167**(2) 235–292.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Birge, John, Francois Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Calafiore, Giuseppe Carlo, L El Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* **130**(1) 1–22.

Campbell, James. 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* **72**(2) 387–405.

Campbell, James. 1996. Hub location and the p-hub median problem. *Operations Research* **44**(6) 923–935.

Campbell, James, Morton O'Kelly. 2012. Twenty-five years of hub location research. *Transportation Science* **46**(2) 153–169.

Chen, Zhi, Daniel Kuhn, Wolfram Wiesemann. 2018. Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210*.

Chen, Zhi, Weijun Xie. 2021. Sharing the value-at-risk under distributional ambiguity. *Mathematical Finance* **31**(1) 531–559.

Correia, Isabel, Francisco Saldanha da Gama. 2015. Facility location under uncertainty. *Location Science*. Springer, 177–203.

Embrechts, Paul, Alexander McNeil, Daniel Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* **1** 176–223.

Erdoğan, Emre, Garud Iyengar. 2006. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming* **107**(1) 37–61.

Ernst, Andreas, Horst Hamacher, Houyuan Jiang, Mohan Krishnamoorthy, Gerhard Woeginger. 2009. Uncapacitated single and multiple allocation p-hub center problems. *Computers & Operations Research* **36**(7) 2230–2241.

Gao, Rui, Anton Kleywegt. 2016. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Goh, Joel, Melvyn Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations Research* **58**(4-part-1) 902–917.

Hakimi, S Louis. 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research* **12**(3) 450–459.

Hakimi, S Louis. 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research* **13**(3) 462–475.

Hanasusanto, Grani, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151**(1) 35–62.

Hanasusanto, Grani, Vladimir Roitch, Daniel Kuhn, Wolfram Wiesemann. 2017. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research* **65**(3) 751–767.

Hao, Zhaowei, Long He, Zhenyu Hu, Jun Jiang. 2020. Robust vehicle pre-allocation with uncertain covariates. *Production and Operations Management* **29**(4) 955–972.

Ho-Nguyen, Nam, Fatma Kılınç-Karzan, Simge Küçükyavuz, Dabeen Lee. 2021. Distributionally robust chance-constrained programs with right-hand side uncertainty under wasserstein ambiguity. *Mathematical Programming* 1–32.

Hult, Edward, Houyuan Jiang, Daniel Ralph. 2014. Exact computational approaches to a stochastic uncapacitated single allocation p-hub center problem. *Computational Optimization and Applications* **59**(1-2) 185–200.

Kara, Bahar, Barbaros Tansel. 2000. On the single-assignment p-hub center problem. *European Journal of Operational Research* **125**(3) 648–655.

Landsman, Zinoviy, Emiliano Valdez. 2003. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal* **7**(4) 55–71.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1) 115–166.

Nemirovski, Arkadi, Alexander Shapiro. 2007. Convex approximations of chance constrained programs. *SIAM Journal on Optimization* **17**(4) 969–996.

O'Kelly, Morton. 1986. The location of interacting hub facilities. *Transportation Science* **20**(2) 92–106.

O'Kelly, Morton. 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* **32**(3) 393–404.

Pérignon, Christophe, Daniel Smith. 2010. Diversification and value-at-risk. *Journal of Banking & Finance* **34**(1) 55–66.

Ruszczyński, Andrzej. 2002. Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming* **93**(2) 195–215.

Ruszczyński, Andrzej, Alexander Shapiro. 2003. Stochastic programming models. *Handbooks in operations research and management science* **10** 1–64.

Shen, Hao, Yong Liang, Zuo-Jun Max Shen. 2021. Reliable hub location model for air transportation networks under random disruptions. *Manufacturing & Service Operations Management* **23**(2) 388–406.

Sim, Thaddeus, Timothy Lowe, Barrett Thomas. 2009. The stochastic p-hub center problem with service-level constraints. *Computers & Operations Research* **36**(12) 3166–3177.

Smith, James, Robert Winkler. 2006. The optimizer's curse: skepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.

Snyder, Lawrence. 2006. Facility location under uncertainty: a review. *IIE transactions* **38**(7) 547–564.

Villani, Cédric. 2009. *Optimal transport: old and new*, vol. 338. Springer.

Wang, Shuming, Zhi Chen, Tianqi Liu. 2020. Distributionally robust hub location. *Transportation Science* **54**(5) 1189–1210.

Xie, Weijun. 2021. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming* **186**(1) 115–155.

Xu, Huan, Constantine Caramanis, Shie Mannor. 2012. Optimization under probabilistic envelope constraints. *Operations Research* **60**(3) 682–699.

Zhu, Taozeng, Jingui Xie, Melvyn Sim. 2022. Joint estimation and robustness optimization. *Management Science* **68**(3) 1659–1677.

Appendix A Additional Results for Section 6.2

We show results under the following distributions.

- Lognormal distribution: $\tilde{u}_e \sim \mathcal{L}(\mu'_e, {\sigma'_e}^2)$, where $\mu'_e = \ln(\mu_i^2/\sqrt{\mu_e^2+\sigma_e^2})$ and $\sigma'_e = \sqrt{\ln(1+\sigma_e^2/\mu_e^2)}$ $\forall e \in \mathcal{E}$;
- Uniform distribution: $\tilde{u}_e \in \mathcal{U}(\mu_e - \sqrt{3}\sigma_e, \mu_e + \sqrt{3}\sigma_e)$ $\forall e \in \mathcal{E}$;
- Two-point distribution: $\tilde{u}_e$ with probability $p_e = \frac{b_e-\mu_e}{b_e-a_e}$ at point a_e and probability $1-p_e$ at point b_e, where $a_e = \mu_e - \sigma_e/\sqrt{3}$ and $b_e = \mu_e + \sqrt{3}\sigma_e$ $\forall e \in \mathcal{E}$.

Note that the three distributions above share the same mean μ_e and variance σ_e^2.

(ρ, γ)	metric	JCC/ICC	JCC/SAA	JCC/Hult	ICC/SAA	ICC/Hult	SAA/Hult
(0.25, 0.9)	$\mathcal{A}$	0.629	0.653	0.651	0.475	0.715	0.675
(0.25, 0.9)	$\mathcal{M}$	0.661	0.715	0.681	0.585	0.810	0.669
(0.25, 0.9)	$\mathcal{Q}$	0.610	0.677	0.644	0.509	0.727	0.650
(0.25, 0.95)	$\mathcal{A}$	0.662	0.590	0.633	0.482	0.648	0.630
(0.25, 0.95)	$\mathcal{M}$	0.660	0.695	0.629	0.606	0.811	0.640
(0.25, 0.95)	$\mathcal{Q}$	0.636	0.623	0.608	0.538	0.693	0.597
(0.5, 0.9)	$\mathcal{A}$	0.511	0.581	0.497	0.581	0.771	0.559
(0.5, 0.9)	$\mathcal{M}$	0.569	0.643	0.550	0.627	0.855	0.585
(0.5, 0.9)	$\mathcal{Q}$	0.529	0.602	0.499	0.581	0.770	0.553
(0.5, 0.95)	$\mathcal{A}$	0.504	0.647	0.510	0.613	0.798	0.533
(0.5, 0.95)	$\mathcal{M}$	0.585	0.743	0.567	0.652	0.881	0.565
(0.5, 0.95)	$\mathcal{Q}$	0.528	0.689	0.519	0.626	0.811	0.529

Table 7 Out-of-sample performance under the lognormal distribution.

$(\rho,\, \gamma)$	metric	JCC/ICC	JCC/SAA	JCC/Hult	ICC/SAA	ICC/Hult	SAA/Hult
(0.25, 0.9)	$\mathcal{A}$	0.630	0.679	0.638	0.497	0.710	0.624
(0.25, 0.9)	$\mathcal{M}$	0.651	0.737	0.648	0.570	0.801	0.636
(0.25, 0.9)	$\mathcal{Q}$	0.616	0.672	0.615	0.500	0.704	0.616
(0.25, 0.95)	$\mathcal{A}$	0.583	0.676	0.594	0.522	0.736	0.619
(0.25, 0.95)	$\mathcal{M}$	0.599	0.728	0.603	0.590	0.850	0.629
(0.25, 0.95)	$\mathcal{Q}$	0.574	0.686	0.584	0.549	0.756	0.606
(0.5, 0.9)	$\mathcal{A}$	0.613	0.680	0.605	0.529	0.751	0.606
(0.5, 0.9)	$\mathcal{M}$	0.636	0.709	0.616	0.580	0.841	0.609
(0.5, 0.9)	$\mathcal{Q}$	0.613	0.678	0.601	0.530	0.759	0.605
(0.5, 0.95)	$\mathcal{A}$	0.540	0.629	0.574	0.485	0.772	0.574
(0.5, 0.95)	$\mathcal{M}$	0.590	0.691	0.602	0.543	0.842	0.586
(0.5, 0.95)	$\mathcal{Q}$	0.554	0.653	0.586	0.504	0.786	0.574

Table 8 Out-of-sample performance under the uniform distributions.

$(\rho,\, \gamma)$	metric	JCC/ICC	JCC/SAA	JCC/Hult	ICC/SAA	ICC/Hult	SAA/Hult
(0.25, 0.9)	$\mathcal{A}$	0.640	0.674	0.617	0.449	0.746	0.666
(0.25, 0.9)	$\mathcal{M}$	0.669	0.705	0.634	0.590	0.845	0.617
(0.25, 0.9)	$\mathcal{Q}$	0.641	0.701	0.622	0.504	0.770	0.634
(0.25, 0.95)	$\mathcal{A}$	0.603	0.572	0.571	0.499	0.755	0.652
(0.25, 0.95)	$\mathcal{M}$	0.647	0.653	0.637	0.635	0.897	0.656
(0.25, 0.95)	$\mathcal{Q}$	0.614	0.618	0.595	0.556	0.798	0.637
(0.5, 0.9)	$\mathcal{A}$	0.618	0.655	0.574	0.570	0.744	0.554
(0.5, 0.9)	$\mathcal{M}$	0.635	0.697	0.635	0.598	0.845	0.611
(0.5, 0.9)	$\mathcal{Q}$	0.632	0.668	0.595	0.563	0.760	0.572
(0.5, 0.95)	$\mathcal{A}$	0.538	0.579	0.538	0.496	0.737	0.585
(0.5, 0.95)	$\mathcal{M}$	0.570	0.634	0.566	0.594	0.844	0.611
(0.5, 0.95)	$\mathcal{Q}$	0.556	0.608	0.558	0.524	0.775	0.599

Table 9 Out-of-sample performance under the two-point distributions.

Appendix B Bisection Search Algorithm

We provide the bisection search algorithm and show its efficiency. For ease of expression, we suppress the subscript $ikjm$.

Bisection Search

Input: search precision ε.

1. RHS $\leftarrow \theta\|\boldsymbol{a}\|_*/\sqrt{\boldsymbol{a}^\top\boldsymbol{\Sigma}\boldsymbol{a}}$.
2. **Interval Initialization.**

   $\underline{\eta} \leftarrow \Phi_g^{-1}(\gamma)$, $\bar{\eta} \leftarrow \underline{\eta}$

   While $f(\bar{\eta}) <$ RHS do

   $\underline{\eta} \leftarrow \bar{\eta}$, $\bar{\eta} \leftarrow 2\bar{\eta}$
3. **Bisection Search.**

   While $\bar{\eta} - \underline{\eta} > \varepsilon$ do

   $\hat{\eta} \leftarrow (\underline{\eta} + \bar{\eta})/2$

   If $f(\hat{\eta}) <$ RHS then

   $\underline{\eta} \leftarrow \hat{\eta}$

   Else

   $\bar{\eta} \leftarrow \hat{\eta}$

Output: $\bar{\eta}$ as the final solution for $\eta^\star$.

The algorithm will find a solution $\bar{\eta}$ with precision ε (*i.e.*, $|\bar{\eta} - \eta^\star| \le \varepsilon$) in at most

$$\left\lceil \log_2 \frac{\eta^\star}{\Phi_g^{-1}(\gamma)} \right\rceil + \left\lceil \log_2 \frac{2\eta^\star - \Phi_g^{-1}(\gamma)}{\varepsilon} \right\rceil$$

iterations, reasoning as follows.

The new risk threshold introduced is $\gamma^\star = \Phi_g(\eta^\star) \ge \gamma$ with

$$\eta^\star = \inf_{\eta} \left\{ \eta \ge \Phi_g^{-1}(\gamma) : f(\eta) \ge \frac{\theta\|\boldsymbol{a}\|_*}{\sqrt{\boldsymbol{a}^\top\boldsymbol{\Sigma}\boldsymbol{a}}} \right\} \quad \text{where} \quad f(\eta) = \eta(\Phi_g(\eta) - \gamma) - \int_{(\Phi_g^{-1}(\gamma))^2/2}^{\eta^2/2} \kappa g(z)\mathrm{d}z.$$

Note that the function $f(\eta)$ is monotonically increasing in η on the interval $\eta \ge \Phi_g^{-1}(\gamma)$, since for any $\eta \ge \Phi_g^{-1}(\gamma)$, we have $f'(\eta) = \Phi_g(\eta) - \gamma + \eta\phi_g(\eta) - \eta\kappa \cdot g(\eta^2/2) = \Phi_g(\eta) - \gamma \ge 0$. Therefore, there exists one and only one zero point for the equation $f(\eta) = \theta\|\boldsymbol{a}\|_*/\sqrt{\boldsymbol{a}^\top\boldsymbol{\Sigma}\boldsymbol{a}}$ on $\eta \in [\Phi_g^{-1}(\gamma), +\infty)$, which is $\bar{\eta}^\star$. Based on this, we could efficiently find $\eta^\star$ via our bisection search algorithm. The result then follows from the two observations: (i) step 2 will stop once $\bar{\eta}$ reaches $\eta^\star$, thus there are at most $\lceil \log_2(\eta^\star/\Phi_g^{-1}(\gamma)) \rceil$ iterations; and (ii) at the beginning of step 3 we have $\underline{\eta} \ge \Phi_g^{-1}(\gamma)$ and $\bar{\eta} \le 2\bar{\eta}^*$, thus the bisection search will take at most $\lceil \log_2((2\eta^\star - \Phi_g^{-1}(\gamma))/\varepsilon) \rceil$ iterations.