An exact algorithm for the probabilistic TSP via a new tractable convex representation of recourse

Yan-Ru Wang[a], Wei-Kun Chen[a], and Ivana Ljubić[b]

[a]*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China*
[b]*Department of Information Systems, Data Analytics and Operations, ESSEC Business School, 95021 Cergy-Pontoise, France*

Abstract

We consider the probabilistic traveling salesman problem (PTSP) in which customer presences are Bernoulli random variables, and the objective is to determine an a priori tour minimizing the expected traveling cost of the a posteriori tour obtained by skipping absent customers after customer presence is revealed. The existing literature has established that, given an a priori tour, the expected recourse function—the objective function of the PTSP—can be evaluated using efficient polynomial-time procedures under certain distributions of the random customer presence. However, a tractable convex representation exploiting the structure of the a priori tour and recourse policy for the expected recourse function was unknown until now, despite its importance in developing sophisticated exact algorithms for the PTSP and related problems. In this paper, we fill this research gap by deriving a new tractable convex piecewise-affine representation for the expected recourse function in the PTSP with an *arbitrary* distribution of the random customer presence. Building upon this convex representation, we develop new mixed integer linear programming formulations for the PTSP, and design an exact algorithm in which the exponentially many linear inequalities are separated dynamically. Extensive computational results on benchmark instances with both independent and correlated customer presences demonstrate that, thanks to the strong linear programming relaxation bound, the proposed exact algorithm significantly outperforms the state-of-the-art integer L-shaped method. Using the proposed algorithm, PTSP instances with up to 300 vertices can now be solved to optimality or near-optimality within a two-hour time limit. Moreover, when run under a 600-second time budget, it can provide much better solutions than a leading heuristic algorithm in the literature.

Keywords: Probabilistic traveling salesman problem; stochastic routing; convex recourse representation; mixed integer linear programming; exact algorithm

1 Introduction

The traveling salesman problem (TSP) is one of the fundamental combinatorial optimization problems. Given an undirected graph $G = (V, E)$ with a non-negative cost for each edge in E, the TSP aims to find a tour that visits each vertex (or customer) in V exactly once while minimizing the total traveling cost. The TSP arises in, or serves as a building block for, a wide variety of applications, including last-mile delivery

Email addresses: `wangyanru@bit.edu.cn` (Yan-Ru Wang), `chenweikun@bit.edu.cn` (Wei-Kun Chen), and `ivana.ljubic@essec.edu` (Ivana Ljubić).

(Stroh et al. 2022), warehouse automation and order picking problems (Löffler et al. 2022), drone-based logistics problems (Blufstein et al. 2024), and manufacturing and scheduling (Charitopoulos et al. 2017). For detailed discussions of the TSP, we refer to Applegate et al. (2011).

In many applications, however, the set of customers requiring service is uncertain when the tour is planned. For example, in last-mile delivery, customers may be unavailable to receive service, cancel their orders, or change their delivery requirements (Özarık et al. 2021); in healthcare home visits, patients may be unavailable (Bazirha et al. 2023). Ignoring customer presence uncertainty may result in routes that include absent customers, which may substantially increase operational costs for service providers and reduce operational efficiency (Florio et al. 2018, Özarık et al. 2021). This has motivated research on stochastic routing problems, in which customer presence uncertainty is directly incorporated into the routing decision process (Jaillet 1985, Bertsimas et al. 1990, Gendreau et al. 1996, 2014).

To account for the uncertainty in customer presence, a straightforward approach is to adopt the *reoptimization strategy*: after customer presence is revealed, one could solve a deterministic TSP based on the present customers. However, this approach has two critical drawbacks. First, it can be computationally expensive, since each realization entails solving an NP-hard deterministic TSP. Second, it may undermine route consistency by producing substantially different visiting sequences across realizations; consistent routes are often desirable in delivery operations because they improve service predictability and driver familiarity (Campbell and Thomas 2008, Voccia et al. 2013, Kovacs et al. 2014, Song et al. 2020). Another approach to addressing customer presence uncertainty is to adopt the *a priori optimization strategy* where (i) one first specifies a planned, or *a priori*, tour over all potential customers, and (ii) once customer presence is revealed, the tour is executed by visiting only the customers that are present, while preserving their relative order in the a priori tour. The resulting *a posteriori* tour is therefore obtained by skipping absent customers from the a priori tour. Under mild conditions such as the triangle inequality, the traveling cost of the *a posteriori* tour is no larger than that of the a priori tour (Jaillet 1985, Bertsimas et al. 1990). The a priori optimization strategy follows this recourse policy that specifies the nature of the adjustments to be made, and can effectively bypass the two drawbacks arising in the reoptimization strategy (Gendreau et al. 2014, 2016).

Applying the a priori optimization strategy to the TSP with stochastic customers gives rise to the probabilistic traveling salesman problem (PTSP), introduced by Jaillet (1985, 1988). Given a probability distribution of the random customer presence and a graph with a non-negative cost for each edge, the PTSP seeks an a priori tour that minimizes the expected traveling cost of the induced a posteriori tour. When all customer presences are independent, the PTSP is called the *independent PTSP*, and it has been widely investigated in the literature; see Jaillet (1985, 1988) and further references below. When correlations between customer presences arise, the PTSP is called the *correlated PTSP* (Wissink 2023). In this paper, we attempt to develop a unified exact algorithm that can address both independent and correlated PTSPs within the same optimization framework.

1.1 Review of existing solution algorithms

The PTSP falls into the class of stochastic vehicle routing problems, and can be formulated as a two-stage stochastic program with recourse. We refer to Gendreau et al. (2014, 2016) and Oyola et al. (2017, 2018) and the references therein for a review of solution algorithms for stochastic vehicle routing problems, and Birge and Louveaux (2011) and Shapiro et al. (2021) and the references therein for a review of solution algorithms for two-stage stochastic programs with recourse. In the following, we review the relevant references on the solution algorithms for PTSPs.

For the independent PTSP, where all customer presences are independent, Jaillet (1985, 1988) showed that given any a priori tour of the graph, the expected recourse function (i.e., the objective function of the problem) can be evaluated in $O(|V|^2)$ time. Using this property, various heuristic algorithms have been proposed to find a suboptimal solution for independent PTSPs. These heuristic algorithms can be classified into three categories: construction heuristics, improvement heuristics, and metaheuristics. Construction heuristics generate a priori tours from scratch, including the savings-based algorithm (Jaillet 1985), space-filling curves and radial sort algorithms (Bertsimas 1988), and an insertion-based algorithm (Weiler et al. 2015). Improvement heuristics attempt to refine an existing a priori tour using probabilistic local-search variants of classical TSP moves such as 2-p-opt and 1-shift (Bertsimas and Howell 1993, Bianchi and Campbell 2007), and the empirical-estimation 2.5-opt (2.5-opt-EEs) heuristic (Birattari et al. 2008, Balaprakash et al. 2010), with candidate moves evaluated by their effect on the expected traveling cost of the induced a posteriori tour. Metaheuristics such as ant colony optimization (Bianchi et al. 2002) and hybrid scatter algorithms (Liu 2007, 2008) have also been proposed to solve independent PTSPs. We refer to Wissink (2019, Chapter 4) for a detailed review of heuristic methods for independent PTSPs. For approximation algorithms with performance guarantees, Shmoys and Talwar (2008) gave the first constant-factor results for the independent PTSP, with randomized and deterministic guarantees of 4 and 8, respectively; Blauth et al. (2025) recently improved these bounds to below 3.1 and 5.9.

Prior to this work, correlated PTSPs had only been studied by Wissink (2023) where the joint distribution of customer presence is estimated using Archimedean and discrete vine (D-vine) elliptical copula-based models. Wissink (2023) showed that the expected recourse cost of a given a priori tour can also be evaluated in polynomial time, which enables the 2.5-opt-EEs heuristic and probabilistic ant-colony optimization for the independent PTSP to be adapted to the correlated PTSP.

The current state-of-the-art exact algorithm for PTSPs is the integer L-shaped method, first proposed by Laporte et al. (1994) for independent PTSPs and later extended by Wissink (2023) to correlated PTSPs. The integer L-shaped method of Laporte et al. (1994) and Wissink (2023) is essentially a linear programming (LP) based branch-and-cut (B&C) algorithm in which integer L-shaped cuts along with several lower-bounding inequalities are used to approximate the "black-box" expected recourse function. However, the LP relaxation bound (at the root node) is usually very poor, particularly due to the weak integer L-shaped cuts, which forces the integer L-shaped method to explore a huge search tree, even for small-sized instances. Indeed, due to this intrinsic difficulty, only independent PTSP instances with up to 50 customers and correlated PTSP instances with up to 26 vertices can be solved to optimality by Laporte et al. (1994) and Wissink (2023), respectively.

1.2 Contributions and outline

To the best of our knowledge, apart from the standard but weak integer L-shaped representation, no sophisticated and tractable convex representation of the expected recourse function in the a priori tour decision variables was known until now. The motivation of this paper is to fill this research gap by developing the *first non-standard* tractable convex representation of the expected recourse function for the PTSP with an arbitrary distribution of the random customer presence and by deriving a computationally efficient LP-based B&C algorithm that is capable of finding a global solution for the PTSP. Our algorithm does not require sampling possible scenarios; instead, it guarantees finding the true optimum for the PTSP over the full probability space, while only working on the space of first-stage variables. The main contributions of the paper are summarized as follows.

- We first derive a tractable convex piecewise-affine representation of the "black-box" expected recourse function under an arbitrary joint distribution of customer presence. The key idea is to express the expected recourse function as an edge-cost-weighted sum of the expected traversal counts (ETCs) per edge. For each edge, the ETC function depends on the two paths connecting its endpoints in the a priori tour. By exploiting this tour structure and the shortcutting recourse policy, we show that each ETC function can be expressed as the maximum of finitely many affine functions of the first-stage tour variables. This yields, to the best of our knowledge, the first exact convex representation of the expected recourse function in the space of a priori tour decisions.

- This representation has direct algorithmic consequences. It allows the expected recourse function to be embedded exactly into mixed integer linear programming (MILP) formulations, without sampling or scenario enumeration. In particular, we first develop a baseline formulation based on a family of ETC inequalities. Then, we strengthen the formulation by enhancing the ETC inequalities. The resulting strengthened formulation provides a substantially tighter LP relaxation and therefore serves as the basis for an exact LP-based B&C algorithm. In this algorithm, the exponentially many inequalities, including the enhanced ETC inequalities, are generated dynamically. We further show that, under mild computational oracle conditions satisfied, for example, by standard independent distributions and the correlated distributions considered in the PTSP literature, the separation of the enhanced ETC inequalities at tour solutions can be carried out in polynomial time.

- The resulting exact algorithm changes the computational frontier of exact PTSP solution. Rather than relying on integer L-shaped cuts that mainly cut off the incumbent first-stage solution (Laporte et al. 1994, Wissink 2023), our algorithm uses ETC inequalities derived from the structural interaction between the a priori tour and the shortcutting recourse policy. This leads to much stronger lower bounds and substantially smaller enumeration trees. Computational experiments on benchmark instances with both independent and correlated customer presences show that the proposed B&C algorithm significantly outperforms the state-of-the-art integer L-shaped method (Laporte et al. 1994, Wissink 2023). Using the newly proposed methodology, PTSP instances with up to 300 vertices can be solved to optimality or near-optimality within a two-hour time limit. Moreover, when run under a restricted time budget, the proposed algorithm can provide a much better solution than a state-of-the-art heuristic algorithm in Wissink (2023).

Although our analysis is developed for the PTSP, the modeling principle is more broadly applicable. The convexification relies only on the facts that the first-stage solution defines a tour and that the recourse policy preserves the relative order of present customers while shortcutting absent ones. This suggests that the proposed convex representation can serve as a foundation for exact algorithms for related a priori stochastic routing problems, including variants with deadlines (Campbell and Thomas 2008, Weyland et al. 2012), time windows (Voccia et al. 2013), and facility-location decisions (Bertsimas 1989, Bertsimas et al. 1990).

This paper is organized as follows. Section 2 presents the problem formulation for the PTSP. Section 3 develops a convex piecewise-affine representation for the expected recourse function. Section 4 presents new MILP formulations for the PTSP. Section 5 describes the exact B&C algorithm and its implementation details. Section 6 reports the computational results. Finally, Section 7 draws conclusions and discusses directions for future research.

2 Problem formulation

In this section, we first introduce the PTSP and then present its mathematical formulation.

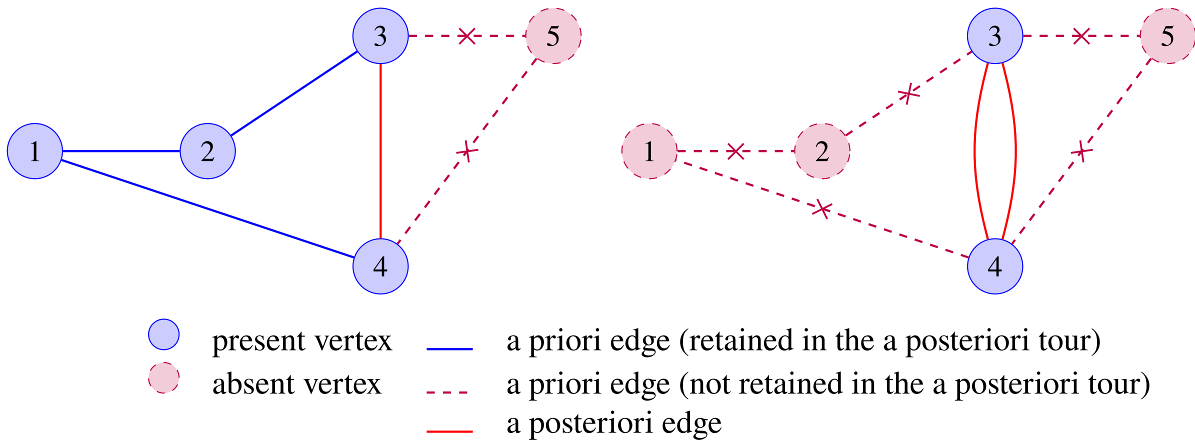


Figure 1: An illustrative example for the computation of $\Phi^{\xi}(\sigma)$ where the a priori tour is induced by $\sigma = (1, 2, 3, 5, 4)$. For realization $\xi^1 = (1, 1, 1, 1, 0)$ in the left graph, the a posteriori tour is $(1, 2, 3, 4, 1)$, and $\Phi^{\xi^1}(\sigma) = c_{12} + c_{23} + c_{34} + c_{41}$. For realization $\xi^2 = (0, 0, 1, 1, 0)$ in the right graph, the a posteriori tour is $(3, 4, 3)$, and $\Phi^{\xi^2}(\sigma) = c_{34} + c_{43} = 2c_{34}$.

2.1 Problem definition

The classic TSP is defined as follows. Let $G = (V, E)$ be a complete undirected graph, where $V = \{1, \ldots, n\}$ ($n \in \mathbb{Z}_+$ with $n \geq 3$) denotes the set of vertices (also referred to as customers) and $E = \{\{i, j\} : i \in V,\ j \in V,\ i \neq j\}$ denotes the set of edges, where $\{i, j\}$ represents the undirected edge connecting customers i and j. Each edge $\{i, j\} \in E$ is associated with a non-negative traveling cost c_{ij} ($= c_{ji}$). A tour is defined as a sequence of customers $\sigma(1), \ldots, \sigma(n), \sigma(1)$, where $\sigma = (\sigma(1), \ldots, \sigma(n))$ is a permutation of V; that is, the tour visits each customer exactly once in the order specified by σ and returns to the starting customer $\sigma(1)$. Throughout, we slightly abuse notation and use σ to denote both the permutation of V and the corresponding tour in G. Online Appendix A summarizes all the notation. The objective of the TSP is to find a tour that minimizes the total traveling cost:

$$\min_{\sigma \in \Pi} \Phi(\sigma) := \sum_{\ell=1}^{n} c_{\sigma(\ell)\sigma(\ell+1)}, \tag{TSP}$$

where $\sigma(n + 1) := \sigma(1)$ and Π denotes the set of all permutations of V.

The PTSP, introduced by Jaillet (1985, 1988), is a stochastic version of the TSP, in which customers are present with predefined probabilities and only present vertices must be visited (once customer presence is revealed). In contrast to the deterministic TSP, the PTSP explicitly accounts for the uncertainty in customer presence, making it more suitable for many real-world applications such as job scheduling (Bertsimas et al. 1990), last-mile delivery (Tang and Miller-Hooks 2007, Wissink 2023), online services (Campbell 2006), and healthcare home visits (Bazirha et al. 2023). The PTSP is a two-stage stochastic program with complete recourse (Laporte and Louveaux 1993). In the first stage, an *a priori tour* $\sigma = (\sigma(1), \ldots, \sigma(n))$ over the entire vertex/customer set V is constructed before any customer presence information is revealed. In the second stage, the customer presence information becomes available, and present customers are visited in the same order as they appear on the a priori tour, while absent customers are simply skipped. Note that under the common assumption that the traveling costs $\{c_{ij}\}_{\{i,j\}\in E}$ satisfy the triangle inequality (i.e., for any distinct $i, j, k \in V$, $c_{ij} \leq c_{ik} + c_{kj}$), the traveling cost of the modified tour, referred to as the *a posteriori tour*, is no larger than that of the a priori tour (Jaillet 1985). Let $\xi = (\xi_1, \ldots, \xi_n)$ be a vector of Bernoulli random variables with an arbitrary joint distribution, where $\xi_i = 1$ if customer i is present and $\xi_i = 0$ otherwise, and let $V^\xi := \{i \in V : \xi_i = 1\}$ be the set of present customers. Then, the a posteriori tour can be denoted as $(\sigma(k_1), \ldots, \sigma(k_{n_\xi}))$, where $n_\xi := |V^\xi|$, $k_1, \ldots, k_{n_\xi} \in \mathbb{Z}_+$ satisfy $1 \leq k_1 < \cdots < k_{n_\xi} \leq n$, and $V^\xi = \{\sigma(k_1), \ldots, \sigma(k_{n_\xi})\}$. The total traveling cost of the a posteriori tour is then computed as

$$\Phi^\xi(\sigma) := \begin{cases} \sum_{h=1}^{n_\xi} c_{\sigma(k_h)\sigma(k_{h+1})}, & \text{if } n_\xi \geq 2; \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $k_{n_\xi+1} := k_1$. An illustrative example of the a posteriori tour and the total traveling cost under two scenarios is presented in Figure 1.

Given the probability distribution of the random vector ξ, the objective of the PTSP is to determine an a priori tour that minimizes the expected cost of the induced a posteriori tour:

$$\min_{\sigma\in\Pi} \mathbb{E}_\xi[\Phi^\xi(\sigma)] = \sum_{\xi\in\{0,1\}^n} \mathbb{P}(\xi)\Phi^\xi(\sigma), \tag{PTSP}$$

where $\mathbb{P}(\xi)$ is the probability of realization ξ. Laporte et al. (1994) considered the PTSP with a depot, i.e., a customer that is always present. This PTSP, however, can be regarded as a special case of problem (PTSP) for which $\mathbb{P}(\xi_i = 1) = 1$ holds for some customer $i \in V$.

2.2 Mathematical formulation

Let $\delta(i) = \{\{k, \ell\} \in E : i \in \{k, \ell\}\}$ denote the set of edges incident to vertex $i \in V$ and $E(S) = \{\{i, j\} \in E : i, j \in S\}$ denote the set of edges with both endpoints in $S \subseteq V$. We introduce, for each $\{i, j\} \in E$, a binary variable x_{ij} denoting whether edge $\{i, j\} \in E$ is included in the a priori tour. Then, problem (PTSP) can be formulated as follows:

$$\min_x \phi(x) := \mathbb{E}_\xi[\Phi^\xi(x)] \tag{PTSP'}$$

$$\text{s.t.} \sum_{\{i,j\}\in\delta(i)} x_{ij} = 2, \ \forall\, i \in V, \tag{2a}$$

$$\sum_{\{i,j\}\in E(S)} x_{ij} \leq |S| - 1, \ \forall\, S \subseteq V, \ 3 \leq |S| \leq n - 3, \tag{2b}$$

$$x_{ij} \in \{0, 1\}, \ \forall\, \{i, j\} \in E. \tag{2c}$$

Constraints (2a), (2b), and (2c) are the degree, subtour elimination, and integrality constraints, respectively. These constraints ensure that a tour $\sigma^x \in \Pi$ in graph G defined by a solution x must visit every vertex in V exactly once and return to the starting vertex; in other words, σ^x is a permutation such that $x_{ij} = 1$ for $\{i, j\} \in \{\{\sigma^x(\ell), \sigma^x(\ell + 1)\}\}_{\ell=1}^n$ and $x_{ij} = 0$ otherwise. For convenience, we use

$$\mathcal{T} := \{x : (2a),\ (2b),\ (2c)\}$$

to denote the set of feasible incidence vectors of a priori tours. Note that other formulations, such as flow-based formulations or formulations based on Miller-Tucker-Zemlin subtour elimination constraints, can also be used to model $\mathcal{T}$ (Padberg and Sung 1991). In the objective function, by abuse of notation, we write $\Phi^{\xi}(x) := \Phi^{\xi}(\sigma^x)$ for $x \in \mathcal{T}$.

One realistic assumption for the PTSP is that for any given $x \in \mathcal{T}$, there exists an efficient oracle (i.e., an efficient polynomial-time algorithm) to compute the expected recourse function $\phi(x)$ exactly. Two examples where this assumption holds are as follows. First, if all customer presences are independent from each other, then for any given $x \in \mathcal{T}$, $\phi(x)$ can be computed in $O(n^2)$ time (Jaillet 1988, Bertsimas and Howell 1993). Second, if customer presences are correlated and the probabilities of jointly correlated customer presences can be estimated using the Archimedean and D-vine elliptical copula models (Panagiotelis et al. 2012), then for any given $x \in \mathcal{T}$, $\phi(x)$ can be computed in $O(n^2)$ and $O(n^4)$ time, respectively (Wissink 2023).

Despite the existence of these efficient evaluation procedures, the value function ϕ has until now been treated as a "black-box" function—the function values are evaluated at the queried point, but to the best of our knowledge, an amenable convex representation of function ϕ that exploits the structure of the a priori solution x was unknown until now. This lack of structure posed a major challenge for the development of sophisticated exact solution methods in the past.

In the existing literature, the only exact algorithm for PTSPs is the integer L-shaped method (Laporte et al. 1994, Wissink 2023). This method first introduces a continuous variable ζ to represent $\phi(x)$ and rewrites problem (PTSP') as the following epigraph form:

$$\min_{x,\zeta} \ \{\zeta : \zeta \geq \phi(x), \ x \in \mathcal{T}\}.$$

Then, it solves the resulting formulation using a B&C algorithm in which the integer L-shaped cuts:

$$\zeta \geq \frac{1}{2}\left(\phi(\bar{x}) - \mathrm{LB}\right)\left(\sum_{\{i,j\}\in E(\bar{x})} x_{ij} - n\right) + \phi(\bar{x}), \ \forall\, \bar{x} \in \mathcal{T}, \tag{3}$$

are used to replace constraints $\zeta \geq \phi(x)$ and are separated at the nodes of the search tree. Here, LB is a global lower bound for ϕ, where sophisticated procedures for its computation under independent and certain correlated customer presences were provided by Laporte et al. (1994) and Wissink (2023), and $E(\bar{x}) = \{\{i,j\} \in E : \bar{x}_{ij} = 1\}$. The validity of the integer L-shaped cut (3) can be seen as follows. For any tour defined by $x \in \mathcal{T}$ other than that defined by $\bar{x}$ (i.e., $x \in \mathcal{T}\backslash\{\bar{x}\}$), at least two edges in $E(\bar{x})$ are not included in the tour defined by x. This indicates that $\sum_{\{i,j\}\in E(\bar{x})} x_{ij} - n \leq -2$, and therefore $\frac{1}{2}\left(\phi(\bar{x}) - \mathrm{LB}\right)\left(\sum_{\{i,j\}\in E(\bar{x})} x_{ij} - n\right) + \phi(\bar{x}) \leq \frac{1}{2}\left(\phi(\bar{x}) - \mathrm{LB}\right)\cdot(-2) + \phi(\bar{x}) = \mathrm{LB} \leq \phi(x)$. For more details on the integer L-shaped method for the PTSP, we refer to Laporte et al. (1994) and Wissink (2023).

Unfortunately, the integer L-shaped cut (3) does not use information about solutions other than $\bar{x}$; it only uses the fact that all other solutions have n edges and differ from $\bar{x}$ by at least two edges. Indeed, the integer L-shaped optimality cut (3) can only cut off the current solution $\bar{x}$ but essentially no other solutions in $\mathcal{T}\backslash\{\bar{x}\}$. This leads to weak cuts, thereby often forcing the B&C algorithm to explore a huge search tree. In the next two sections, we will exploit the structure of the a priori tour and recourse policy to develop a new tractable convex representation for function ϕ, which enables us to derive novel MILP reformulations for the PTSP that are much more computationally amenable to a B&C algorithm.

3 A convex representation of the expected recourse function

In this section, we develop a new tractable representation of the expected recourse function $\mathbb{E}_{\xi}[\Phi^{\xi}(\cdot)]$ in terms of the first-stage edge variables x. To do this, we first decompose the "black-box" expected recourse function into an edge-cost-weighted sum of relatively simple component functions, each of which characterizes the ETC of a given edge in the a posteriori tour. Then, by exploiting the structure of the solution in the underlying formulation, we show that each ETC function can be represented as the maximum of finitely many affine functions. Combining the two steps enables us to provide, to the best of our knowledge, the first non-standard tractable convex (piecewise-affine) representation for the expected recourse function, which paves the way for developing MILP-based algorithms for the PTSP.

3.1 Disaggregating the expected recourse function

To facilitate the reformulation, we first introduce some notation. Let $x \in \mathcal{T}$ be a tour solution and σ^x be the permutation defined by x. We denote by $\pi_i = (\sigma^x)^{-1}(i)$ the position (or visiting order) of customer $i \in V$, starting from customer $\sigma^x(1)$. For two distinct customers $i, j \in V$, the path from $i(= \sigma^x(\pi_i))$ to $j(= \sigma^x(\pi_j))$ along tour σ^x is $(\sigma^x(\pi_i), \sigma^x(\pi_i + 1), \ldots, \sigma^x(\pi_i + m))$, where $m := (\pi_j - \pi_i) \bmod n$ is the number of edges along the path from i to j in tour σ^x. Throughout, we denote $\sigma^x(k) := \sigma^x(k \bmod n)$ for $k \in \{n+1, \ldots, 2n-1\}$, where mod is the modulo operator. For notational simplicity, we use

$$S_{ij}^x := \{\sigma^x(\pi_i + k) \, : \, k = 1, \ldots, m-1\} \tag{4}$$

to denote the set of intermediate customers on the path $(\sigma^x(\pi_i), \sigma^x(\pi_i + 1), \ldots, \sigma^x(\pi_i + m))$ (without including i and j). Similarly, letting $m' := (\pi_i - \pi_j) \bmod n$, then

$$T_{ij}^x := \left\{\sigma^x(\pi_j + k) \, : \, k = 1, \ldots, m'-1\right\} \tag{5}$$

denotes the set of intermediate customers on the path $(\sigma^x(\pi_j), \sigma^x(\pi_j + 1), \ldots, \sigma^x(\pi_j + m'))$ (without including i and j).

Given a realization $\xi \in \{0,1\}^n$, a customer i with $\xi_i = 1$ will be visited while a customer i with $\xi_i = 0$ will be skipped in the a posteriori tour. Let $\Psi_{ij}^\xi(x)$ denote the traversal count of edge $\{i, j\} \in E$ in the a posteriori tour induced by x under realization $\xi \in \{0,1\}^n$. If $\xi_i = \xi_j = 1$ and $\xi_k = 0$ for all $k \in S_{ij}^x \cup T_{ij}^x$, then $\Psi_{ij}^\xi(x) = 2$; if $\xi_i = \xi_j = 1$ and either $\xi_k = 0$ holds for all $k \in S_{ij}^x$ or $\xi_k = 0$ holds for all $k \in T_{ij}^x$ but not both, then $\Psi_{ij}^\xi(x) = 1$; otherwise, $\Psi_{ij}^\xi(x) = 0$. In other words, the traversal count of edge $\{i, j\}$ in the a posteriori tour is:

$$\Psi_{ij}^\xi(x) = \begin{cases} 2, & \text{if } \xi \in \Xi_{ij}^x \cap \Xi_{ji}^x; \\ 1, & \text{if } \xi \in (\Xi_{ij}^x \cup \Xi_{ji}^x) \backslash (\Xi_{ij}^x \cap \Xi_{ji}^x); \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where Ξ_{ij}^x and Ξ_{ji}^x are defined as

$$\Xi_{ij}^x := \left\{\xi \in \{0,1\}^n : \xi_i = \xi_j = 1, \ \xi_k = 0, \ \forall \, k \in S_{ij}^x\right\}, \tag{7a}$$

$$\Xi_{ji}^x := \left\{\xi \in \{0,1\}^n : \xi_i = \xi_j = 1, \ \xi_k = 0, \ \forall \, k \in T_{ij}^x\right\}. \tag{7b}$$

Note that after the a priori tour x is fixed, $\Psi_{ij}^\xi(x)$ is a discrete random variable, with probabilities $\mathbb{P}(\Psi_{ij}^\xi(x) = 2)$, $\mathbb{P}(\Psi_{ij}^\xi(x) = 1)$, and $\mathbb{P}(\Psi_{ij}^\xi(x) = 0)$ being $\mathbb{P}(\Xi_{ij}^x \cap \Xi_{ji}^x)$, $\mathbb{P}((\Xi_{ij}^x \cup \Xi_{ji}^x) \backslash (\Xi_{ij}^x \cap \Xi_{ji}^x))$, and $1 - \mathbb{P}(\Xi_{ij}^x \cup \Xi_{ji}^x)$, respectively, where for $\Xi \subseteq \{0,1\}^n$, we let $\mathbb{P}(\Xi) := \sum_{\xi \in \Xi} \mathbb{P}(\xi)$. These probabilities are also called persistence of $\Psi_{ij}^\xi(x)$ at values 2, 1, and 0 in the literature (Bertsimas et al. 2006, Natarajan et al. 2009, Shen et al. 2010).

Example 1. *Consider the example in Figure 1. Let x be a solution inducing an a priori tour $\sigma^x = (1,2,3,5,4)$. For $i = 3$ and $j = 4$, we have $i = \sigma^x(3)$ and $j = \sigma^x(5)$. The two paths between customers 3 and 4 along σ^x are $(3,5,4)$ and $(4,1,2,3)$. For realization $\xi^1 = (1,1,1,1,0)$, customer 5 is skipped along the path $(3,5,4)$ (as $\xi^1_5 = 0$) while all customers along the path $(4,1,2,3)$ are present in the a posteriori tour. Thus, the traversal count of edge $\{3,4\}$ under realization ξ^1 is $\Psi^{\xi^1}_{34}(x) = 1$. For realization $\xi^2 = (0,0,1,1,0)$, due to the absence of customers $1,2,5$ (i.e., $\xi^2_1 = \xi^2_2 = \xi^2_5 = 0$), the paths between customers 3 and 4 in the a posteriori tour are $(3,4)$ and $(4,3)$, respectively. Therefore, the traversal count of edge $\{3,4\}$ under realization ξ^2 is $\Psi^{\xi^2}_{34}(x) = 2$.*

Using (6), the ETC of edge $\{i,j\}$ in the a posteriori tour can be written as

$$\mathbb{E}_\xi[\Psi^\xi_{ij}(x)] = \sum_{\xi\in\{0,1\}^n} \mathbb{P}(\xi)\Psi^\xi_{ij}(x) = 2\times\mathbb{P}(\Xi^x_{ij}\cap\Xi^x_{ji}) + \mathbb{P}((\Xi^x_{ij}\cup\Xi^x_{ji})\setminus(\Xi^x_{ij}\cap\Xi^x_{ji})) = \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji}). \quad (8)$$

That is, $\mathbb{E}_\xi[\Psi^\xi_{ij}(x)]$ is equal to the sum of the probabilities of (i) the realizations where customers i and j are present while all intermediate customers on the path from i to j along tour σ^x are absent and (ii) the realizations where customers i and j are present while all intermediate customers on the path from j to i along tour σ^x are absent.

The expected traveling cost $\mathbb{E}_\xi[\Phi^\xi(x)]$ of the a posteriori tour defined by $x \in \mathcal{T}$ is equivalent to the edge-cost-weighted sum of such ETCs:

$$\mathbb{E}_\xi[\Phi^\xi(x)] = \mathbb{E}_\xi\left[\sum_{\{i,j\}\in E} c_{ij}\Psi^\xi_{ij}(x)\right] = \sum_{\{i,j\}\in E} c_{ij}\mathbb{E}_\xi[\Psi^\xi_{ij}(x)] = \sum_{\{i,j\}\in E} c_{ij}\left(\mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})\right). \quad (9)$$

The above disaggregated form makes it easier to derive a tractable convex representation for the expected recourse function of the PTSP using the x variables—one just needs to derive, for each $\{i,j\}\in E$, a convex representation for the "relatively simple" ETC functions $\mathbb{E}_\xi[\Psi^\xi_{ij}(\cdot)]$ for $\{i,j\}\in E$.

3.2 Convexification of the expected traversal count function per edge

The goal of this subsection is to show that for each edge $\{i,j\}\in E$, the ETC function $\mathbb{E}_\xi[\Psi^\xi_{ij}(\cdot)]$ admits a convex representation—it can be expressed as the maximum of a collection of affine functions in $x\in\mathcal{T}$.

To proceed, we introduce some notation that is used throughout this paper. For any $\{i,j\}\in E$ and any subset $I \subseteq V\setminus\{i,j\}$, let

$$\Gamma^I_{ij} = \left\{\xi\in\{0,1\}^n : \xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k\in I\right\} \quad (10)$$

be the set of realizations for which customers i and j are present and all customers in I are absent, and $\mathbb{P}(\Gamma^I_{ij}) := \sum_{\xi\in\Gamma^I_{ij}}\mathbb{P}(\xi)$ is the associated probability. From the definitions of $\mathbb{P}(\Xi^x_{ij})$ and $\mathbb{P}(\Xi^x_{ji})$ in (7a)–(7b), it follows that $\mathbb{P}(\Xi^x_{ij}) = \mathbb{P}(\Gamma^{S^x_{ij}}_{ij})$ and $\mathbb{P}(\Xi^x_{ji}) = \mathbb{P}(\Gamma^{T^x_{ij}}_{ij})$. For $\{i,j\}\in E$, let

$$\mathcal{P}_{ij} := \{\{S,T\} : S\cup T = V\setminus\{i,j\},\ S\cap T = \varnothing\},$$

be the set of all partitions of the customers in $V\setminus\{i,j\}$. Then, for any $\{S,T\}\in\mathcal{P}_{ij}$, either $\{S,T\} = \{S^x_{ij}, T^x_{ij}\}$ or $\{S,T\}\cap\{S^x_{ij},T^x_{ij}\} = \varnothing$. This allows us to express the ETC function using indicator functions:

$$\mathbb{E}_\xi[\Psi^\xi_{ij}(x)] = \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji}) = \max_{\{S,T\}\in\mathcal{P}_{ij}}\left(\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}=S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{T^x_{ij}=T}\right), \quad (11)$$

where $\mathbf{1}_A$ is the indicator function that takes the value 1 if event A is true and 0 otherwise.

Given any solution $x \in \mathcal{T}$ and any 2-partition $\{S, T\} \in \mathcal{P}_{ij}$, it follows

$$\mathbf{1}_{S^x_{ij}=S} \le \mathbf{1}_{S^x_{ij}\subseteq S} \le \mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S}, \tag{12a}$$

$$\mathbf{1}_{T^x_{ij}=T} \le \mathbf{1}_{T^x_{ij}\subseteq T} \le \mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T}, \tag{12b}$$

where $\mathbf{1}_{A\vee B} = 1$ if at least one of events A and B occurs and $\mathbf{1}_{A\vee B} = 0$ otherwise. In the following, we will show that after replacing the indicator functions $\mathbf{1}_{S=S^x_{ij}}$ and $\mathbf{1}_{T=T^x_{ij}}$ with their upper bounds $\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S}$ and $\mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T}$, respectively, equation (11) still holds.

We start with the following lemma on the upper bound for $\mathbb{P}(\Gamma^S_{ij})+\mathbb{P}(\Gamma^T_{ij})$ for any 2-partition $\{S, T\} \in \mathcal{P}_{ij}$.

Lemma 2. *Let $\{i, j\} \in E$ and $\{S, T\} \in \mathcal{P}_{ij}$. Then*

$$\mathbb{P}(\Gamma^S_{ij}) + \mathbb{P}(\Gamma^T_{ij}) \le \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}) + \mathbb{P}(\Gamma^{\varnothing}_{ij}). \tag{13}$$

Proof. Proof. Observe that

$$\begin{aligned}
\mathbb{P}(\Gamma^{\varnothing}_{ij}) &= \mathbb{P}(\xi_i = \xi_j = 1) \\
&= \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in S) + \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in T) \\
&\qquad - \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in S \cup T) \\
&\qquad + \mathbb{P}(\xi_i = \xi_j = 1,\ \exists\,(k_1, k_2) \in S \times T \text{ with } \xi_{k_1} = \xi_{k_2} = 1) \\
&\overset{\text{(a)}}{=} \mathbb{P}(\Gamma^S_{ij}) + \mathbb{P}(\Gamma^T_{ij}) - \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}) + \mathbb{P}(\xi_i = \xi_j = 1,\ \exists\,(k_1, k_2) \in S \times T \text{ with } \xi_{k_1} = \xi_{k_2} = 1) \\
&\overset{\text{(b)}}{\ge} \mathbb{P}(\Gamma^S_{ij}) + \mathbb{P}(\Gamma^T_{ij}) - \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}),
\end{aligned}$$

where (a) follows from $S \cup T = V\backslash\{i, j\}$ and the definition of Γ^I_{ij} in (10); and (b) follows from $\mathbb{P}(\xi_i = \xi_j = 1,\ \exists\,(k_1, k_2) \in S \times T \text{ with } \xi_{k_1} = \xi_{k_2} = 1) \ge 0$. □ □

Proposition 3. *Let $\{i, j\} \in E$ and $x \in \mathcal{T}$. Then*

$$\mathbb{E}_\xi[\Psi^\xi_{ij}(x)] = \max_{\{S,T\}\in\mathcal{P}_{ij}} \left(\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T}\right), \tag{14}$$

where the maximum is attained at $\{S, T\} = \{S^x_{ij}, T^x_{ij}\}$.

Proof. Proof. Let $\{S, T\}$ be a partition of $V\backslash\{i, j\}$. It suffices to show

$$\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T} \le \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})(= \mathbb{E}_\xi[\Psi^\xi_{ij}(x)]).$$

We consider the following three cases.

(i) $\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S} + \mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T} = 2$. If $\{S^x_{ij}, T^x_{ij}\} = \{\varnothing, V\backslash\{i, j\}\}$, then by Theorem 2, $\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T} = \mathbb{P}(\Gamma^S_{ij})+\mathbb{P}(\Gamma^T_{ij}) \le \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij})+\mathbb{P}(\Gamma^{\varnothing}_{ij}) = \mathbb{P}(\Xi^x_{ij})+\mathbb{P}(\Xi^x_{ji})$. Otherwise, since $\{S, T\}$ and $\{S^x_{ij}, T^x_{ij}\}$ are both partitions of $V\backslash\{i, j\}$, it follows $\{S, T\} = \{S^x_{ij}, T^x_{ij}\}$ (or equivalently, $(S, T) = (S^x_{ij}, T^x_{ij})$ or $(S, T) = (T^x_{ij}, S^x_{ij})$). Thus, $\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S\vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T\vee T^x_{ij}\subseteq T} = \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})$.

(ii) $\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} + \mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = 1$. Without loss of generality, we assume that $\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} = 1$ and $\mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = 0$. Hence,

$$\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = \mathbb{P}(\Gamma^S_{ij}). \tag{15}$$

From $\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} = 1$, it follows $S^x_{ij} \subseteq S$ or $T^x_{ij} \subseteq S$. If $S^x_{ij} \subseteq S$, then,

$$\mathbb{P}(\Gamma^S_{ij}) = \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in S) \le \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in S^x_{ij}) = \mathbb{P}(\Xi^x_{ij});$$

otherwise, $T^x_{ij} \subseteq S$ and

$$\mathbb{P}(\Gamma^S_{ij}) = \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in S) \le \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in T^x_{ij}) = \mathbb{P}(\Xi^x_{ji}).$$

In both cases, it follows that $\mathbb{P}(\Gamma^S_{ij}) \le \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})$. This, together with (15), implies $\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = \mathbb{P}(\Gamma^S_{ij}) \le \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})$.

(iii) $\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} + \mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = 0$. Then, $\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} = 0$ and $\mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = 0$, and thus $\mathbb{P}(\Gamma^S_{ij})\mathbf{1}_{S^x_{ij}\subseteq S \vee T^x_{ij}\subseteq S} + \mathbb{P}(\Gamma^T_{ij})\mathbf{1}_{S^x_{ij}\subseteq T \vee T^x_{ij}\subseteq T} = 0 \le \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji})$. □

□

Theorem 3 implies that, to derive a closed formula for the ETC function $\mathbb{E}_\xi[\Psi^\xi_{ij}(\cdot)]$, it suffices to consider the indicator terms $\mathbf{1}_{S^x_{ij}\subseteq I \vee T^x_{ij}\subseteq I}$ for $I \subseteq V\backslash\{i,j\}$. Note that when $I = V\backslash\{i,j\}$, $\mathbf{1}_{S^x_{ij}\subseteq I \vee T^x_{ij}\subseteq I}$ is a constant.

Lemma 4. *Let $\{i,j\} \in E$ and $x \in \mathcal{T}$. Then,* $\mathbf{1}_{S^x_{ij}\subseteq V\backslash\{i,j\} \vee T^x_{ij}\subseteq V\backslash\{i,j\}} = 1$.

Proof. Proof. The proof follows from the fact that S^x_{ij} and T^x_{ij} are subsets of $V\backslash\{i,j\}$. □ □

To consider the case $I \subsetneq V\backslash\{i,j\}$, we first provide the following remark on the structure of the solution $x \in \mathcal{T}$.

Remark 5. *Let $x \in \mathcal{T}$ and $S \subsetneq V$ be such that $|S| \ge 2$. Then, inequality $\sum_{\{k,\ell\}\in E(S)} x_{k\ell} \le |S| - 1$ holds. Moreover, the equality holds if and only if the edges in $\{\{k,\ell\} \in E(S) : x_{k\ell} = 1\}$ form a path passing through all vertices in S.*

Indeed, summing the degree constraints (2a) over all vertices in S gives

$$2\sum_{\{k,\ell\}\in E(S)} x_{k\ell} + \sum_{\{k,\ell\}\in\delta(S)} x_{k\ell} = 2|S|,$$

where $\delta(S) := \{\{k,\ell\} \in E : k \in S,\ \ell \in V\backslash S\}$. Since $x \in \mathcal{T}$ is the incidence vector of a Hamiltonian cycle and $\emptyset \subsetneq S \subsetneq V$, at least two selected edges must cross the cut $\delta(S)$, i.e., $\sum_{\{k,\ell\}\in\delta(S)} x_{k\ell} \ge 2$. Therefore, it must follow that $\sum_{\{k,\ell\}\in E(S)} x_{k\ell} \le |S| - 1$. Moreover, the equality holds if and only if $\sum_{\{k,\ell\}\in\delta(S)} x_{k\ell} = 2$. From $x \in \mathcal{T}$, the latter is equivalent to saying that the edges in $\{\{k,\ell\} \in E(S) : x_{k\ell} = 1\}$ form a path passing through all vertices in S. Theorem 5 enables us to provide linear lower bound approximations for the indicator function $\mathbf{1}_{S^x_{ij}\subseteq I \vee T^x_{ij}\subseteq I}$ when $I \subsetneq V\backslash\{i,j\}$, and a sufficient condition under which the lower bound is tight.

Lemma 6. *Let $\{i,j\} \in E$, $I \subsetneq V\backslash\{i,j\}$, and $x \in \mathcal{T}$. Then,*

$$\mathbf{1}_{S_{ij}^x \subseteq I \vee T_{ij}^x \subseteq I} \geq \sum_{\{k,\ell\} \in E(I \cup \{i,j\})} x_{k\ell} - |I|. \tag{16}$$

Moreover, if $I = S_{ij}^x$ or $I = T_{ij}^x$, then $\sum_{\{k,\ell\} \in E(I \cup \{i,j\})} x_{k\ell} - |I| = 1$ and inequality (16) holds at equality.

Proof. Proof. If $\sum_{\{k,\ell\} \in E(I \cup \{i,j\})} x_{k\ell} - |I| \leq 0$, then inequality (16) holds trivially. Otherwise, by Theorem 5, $\sum_{\{k,\ell\} \in E(I \cup \{i,j\})} x_{k\ell} - |I| = 1$ and the edges in $\{\{k,\ell\} \in E(I \cup \{i,j\}) : x_{k\ell} = 1\}$ form a path passing through all customers in $I \cup \{i,j\}$. Since this path is a contiguous segment of the Hamiltonian cycle and contains both i and j, it contains the i-to-j segment in one of the two directions of the tour. Therefore, all intermediate vertices on one of the two i-to-j segments must belong to I, i.e., either $S_{ij}^x \subseteq I$ or $T_{ij}^x \subseteq I$. □ □

We are now ready to present the main result of this subsection, that is, the probability function $\mathbb{P}(\Xi_{ij}^x) + \mathbb{P}(\Xi_{ji}^x)$ can be written as the maximum of a collection of affine functions in x.

Proposition 7. *Given $\{i,j\} \in E$ and $x \in \mathcal{T}$, it follows*

$$\begin{aligned}\mathbb{E}_\xi[\Psi_{ij}^\xi(x)] = \max \Bigg\{ & \mathbb{P}(\Gamma_{ij}^{V\backslash\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) x_{ij}, \\ & \max_{\substack{\{S,T\} \in \mathcal{P}_{ij}: \\ \varnothing \notin \{S,T\}}} \left\{ \mathbb{P}(\Gamma_{ij}^S) \left(\sum_{\{k,\ell\} \in E(S \cup \{i,j\})} x_{k\ell} - |S| \right) + \mathbb{P}(\Gamma_{ij}^T) \left(\sum_{\{k,\ell\} \in E(T \cup \{i,j\})} x_{k\ell} - |T| \right) \right\} \Bigg\},\end{aligned} \tag{17}$$

where if $x_{ij} = 1$, the maximum of (17) is attained at $\mathbb{P}(\Gamma_{ij}^{V\backslash\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) x_{ij}$; otherwise, the maximum of (17) is attained at

$$\mathbb{P}(\Gamma_{ij}^S) \left(\sum_{\{k,\ell\} \in E(S \cup \{i,j\})} x_{k\ell} - |S| \right) + \mathbb{P}(\Gamma_{ij}^T) \left(\sum_{\{k,\ell\} \in E(T \cup \{i,j\})} x_{k\ell} - |T| \right),$$

with $\{S,T\} = \{S_{ij}^x, T_{ij}^x\}$.

Proof. Proof. Let $\{S,T\} \in \mathcal{P}_{ij}$. Then, by Theorem 3, it follows

$$\mathbb{E}_\xi[\Psi_{ij}^\xi(x)] = \mathbb{P}(\Xi_{ij}^x) + \mathbb{P}(\Xi_{ji}^x) \geq \mathbb{P}(\Gamma_{ij}^S) \mathbf{1}_{S_{ij}^x \subseteq S \vee T_{ij}^x \subseteq S} + \mathbb{P}(\Gamma_{ij}^T) \mathbf{1}_{S_{ij}^x \subseteq T \vee T_{ij}^x \subseteq T}. \tag{18}$$

If $\varnothing \in \{S,T\}$, then $\{S,T\} = \{\varnothing, V\backslash\{i,j\}\}$. Without loss of generality, we assume that $S = \varnothing$ and $T = V\backslash\{i,j\}$. From Theorem 4 and $T = V\backslash\{i,j\}$, we obtain $\mathbf{1}_{S_{ij}^x \subseteq T \vee T_{ij}^x \subseteq T} = 1$; and from $S = \varnothing$ and Theorem 6, we obtain $\mathbf{1}_{S_{ij}^x \subseteq S \vee T_{ij}^x \subseteq S} \geq \sum_{\{k,\ell\} \in E(S \cup \{i,j\})} x_{k\ell} - |S| = x_{ij}$. Thus,

$$\mathbb{P}(\Gamma_{ij}^S) \mathbf{1}_{S_{ij}^x \subseteq S \vee T_{ij}^x \subseteq S} + \mathbb{P}(\Gamma_{ij}^T) \mathbf{1}_{S_{ij}^x \subseteq T \vee T_{ij}^x \subseteq T} \geq \mathbb{P}(\Gamma_{ij}^{\varnothing}) x_{ij} + \mathbb{P}(\Gamma_{ij}^{V\backslash\{i,j\}}). \tag{19}$$

Otherwise, $\varnothing \notin \{S,T\}$ and $V\backslash\{i,j\} \notin \{S,T\}$. From Theorem 6, it follows

$$\begin{aligned}&\mathbb{P}(\Gamma_{ij}^S) \mathbf{1}_{S_{ij}^x \subseteq S \vee T_{ij}^x \subseteq S} + \mathbb{P}(\Gamma_{ij}^T) \mathbf{1}_{S_{ij}^x \subseteq T \vee T_{ij}^x \subseteq T} \\ &\quad \geq \mathbb{P}(\Gamma_{ij}^S) \left(\sum_{\{k,\ell\} \in E(S \cup \{i,j\})} x_{k\ell} - |S| \right) + \mathbb{P}(\Gamma_{ij}^T) \left(\sum_{\{k,\ell\} \in E(T \cup \{i,j\})} x_{k\ell} - |T| \right).\end{aligned} \tag{20}$$

By (18)–(20) and the arbitrary selection of the partition $\{S, T\} \in \mathcal{P}_{ij}$, we obtain $\mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji}) \geq g(x)$, where $g(x)$ denotes the right-hand side of (17).

Now, if $x_{ij} = 1$, then $\{S^x_{ij}, T^x_{ij}\} = \{\varnothing, V\backslash\{i, j\}\}$ and the right-hand side of (19) becomes

$$\mathbb{P}(\Gamma^{\varnothing}_{ij})x_{ij} + \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}) = \mathbb{P}(\Gamma^{\varnothing}_{ij}) + \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}) = \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji}).$$

Otherwise, $x_{ij} = 0$ and the right-hand side of (20) with $\{S, T\} = \{S^x_{ij}, T^x_{ij}\}$ becomes

$$\mathbb{P}(\Gamma^S_{ij})\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma^T_{ij})\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right) \overset{(a)}{=} \mathbb{P}(\Gamma^S_{ij}) + \mathbb{P}(\Gamma^T_{ij}) = \mathbb{P}(\Xi^x_{ij}) + \mathbb{P}(\Xi^x_{ji}),$$

where (a) follows from the second part of Theorem 6. This completes the proof. □ □

3.3 Convexification of the expected recourse function

As a direct consequence of Theorem 7, we can provide a closed formula for the expected recourse function $\mathbb{E}_\xi[\Phi^\xi(\cdot)]$ (in (9)) of the PTSP:

Theorem 8. *Let $x \in \mathcal{T}$. Then*

$$\begin{aligned}\mathbb{E}_\xi[\Phi^\xi(x)] = \sum_{\{i,j\}\in E} c_{ij} \max \Bigg\{ & \mathbb{P}(\Gamma^{V\backslash\{i,j\}}_{ij}) + \mathbb{P}(\Gamma^{\varnothing}_{ij})x_{ij}, \\ & \max_{\substack{\{S,T\}\in\mathcal{P}_{ij}:\\ \varnothing\notin\{S,T\}}} \left\{\mathbb{P}(\Gamma^S_{ij})\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma^T_{ij})\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right)\right\}\Bigg\}.\end{aligned} \tag{21}$$

Two remarks on Theorem 8 are in order. First, Theorem 8 establishes a new tractable representation of the expected recourse function as a convex piecewise-affine function—the edge-cost-weighted sum of the maxima over collections of affine functions—in x. The convexity and piecewise affinity in representation (21) are crucial, as they enable the development of new MILP formulations for the PTSP along with B&C algorithms; see Section 4 further ahead.

Second, from a computational perspective, the use of the representation in (21) requires the (efficient) computation of the joint probability $\mathbb{P}(\Gamma^I_{ij})$ for $I \subseteq V\backslash\{i, j\}$, i.e., the probability of realizations where i and j are present while all vertices in I are absent. This requirement is not restrictive in many cases. As an example, $\mathbb{P}(\Gamma^I_{ij})$ can be computed in linear time when $\{\xi_i\}_{i\in V}$ are independent Bernoulli random variables with $\mathbb{P}(\xi_i = 1) = p_i$ for $i \in V$ (Jaillet 1985, Laporte et al. 1994, Voccia et al. 2013). In this case, $\mathbb{P}(\Gamma^I_{ij})$ can be computed as:

$$\mathbb{P}(\Gamma^I_{ij}) = \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\ k \in I) = \mathbb{P}(\xi_i = 1) \times \mathbb{P}(\xi_j = 1) \times \prod_{k\in I} \mathbb{P}(\xi_k = 0) = p_i p_j \prod_{k\in I}(1 - p_k).$$

For the PTSP under correlated distributions of ξ, where Archimedean or D-vine elliptical copula-based models can be used to estimate the joint probabilities, $\mathbb{P}(\Gamma^I_{ij})$ can also be computed in polynomial time (Wissink 2023); for completeness, we provide the details in Online Appendix B.

4 MILP formulation for the PTSP over the full probability space

In this section, we first exploit the convex piecewise-affine characterization of function $\mathbb{E}_\xi[\Phi^\xi(\cdot)]$ developed in Section 3 to develop a new MILP formulation for the PTSP, in which an exponential number of linear constraints are used to model the nonlinear max terms in the objective function. Then, we derive a class of strengthened valid inequalities that yield a much stronger LP relaxation bound, making the resulting MILP formulation much more amenable to algorithmic design. We show that under mild conditions, separating the exponentially many constraints (characterizing the max terms in the objective function) at a given tour solution can be conducted in polynomial time.

4.1 An MILP formulation for the PTSP based on the convex representation of $\mathbb{E}_\xi[\Phi^\xi(\cdot)]$

Using the convex piecewise-affine representation (21) of the expected recourse function $\mathbb{E}_\xi[\Phi^\xi(\cdot)]$ in Theorem 8, we can develop an MILP formulation for the PTSP. In particular, by introducing, for each edge $\{i,j\} \in E$, a continuous variable θ_{ij} representing the max term (17), we obtain the following MILP formulation for the PTSP:

$$\min_{x\in\mathcal{T},\,\theta\in\mathbb{R}_+^{|E|}} \sum_{\{i,j\}\in E} c_{ij}\theta_{ij} \tag{MILP}$$

$$\text{s.t.} \quad \theta_{ij} \geq \mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right),$$
$$\forall\, \{i,j\} \in E,\ \forall\, \{S,T\} \in \mathcal{P}_{ij} : \varnothing \notin \{S,T\}, \tag{22}$$
$$\theta_{ij} \geq \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing})x_{ij},\ \forall\, \{i,j\} \in E. \tag{23}$$

Formulation (MILP) enables the use of MILP-based algorithms for finding a globally optimal solution or a high-quality solution with an optimality guarantee for PTSPs. It can be applied to PTSPs with general distributions of the Bernoulli random vector ξ, provided that an efficient oracle is available for computing $\mathbb{P}(\Gamma_{ij}^I)$ for any $\{i,j\} \in E$ and $I \subseteq V \setminus \{i,j\}$.

However, the formulation (MILP) contains an exponential number of ETC constraints (22), making it impractical to solve directly by using state-of-the-art MILP solvers such as Gurobi, Cplex, and Xpress. Nevertheless, these constraints can be added in a dynamic fashion rather than explicitly enforced at the beginning. The following result shows that the ETC inequalities (22) can be efficiently separated, provided that probabilities $\mathbb{P}(\Gamma_{ij}^I)$ for any $I \subseteq V\setminus\{i,j\}$ and $\{i,j\} \in E$ can be evaluated by an efficient computational oracle.

Theorem 9. *Suppose that, given any $\{i,j\} \in E$ and $I \subseteq V\setminus\{i,j\}$, the probability $\mathbb{P}(\Gamma_{ij}^I)$ can be computed in polynomial time. Then, given any point $(\bar{x},\bar{\theta}) \in \mathcal{T} \times \mathbb{R}_+^{|E|}$ satisfying (23), one can find an ETC inequality (22) violated by $(\bar{x},\bar{\theta})$ or certify that none exists, in polynomial time.*

Proof. Proof. It suffices to consider the separation of ETC inequalities (22) for each fixed $\{i,j\} \in E$. Let $\bar{S} = S_{ij}^{\bar{x}}$ and $\bar{T} = T_{ij}^{\bar{x}}$. From Theorem 7 and the fact that inequality (23) holds at point $(\bar{x},\bar{\theta}) \in \mathcal{T} \times \mathbb{R}_+^{|E|}$, if $\bar{\theta}_{ij} \geq \mathbb{P}(\Xi_{ij}^{\bar{x}}) + \mathbb{P}(\Xi_{ji}^{\bar{x}}) = \mathbb{P}(\Gamma_{ij}^{\bar{S}}) + \mathbb{P}(\Gamma_{ij}^{\bar{T}})$, then no inequality in (22) is violated by $(\bar{x},\bar{\theta})$; otherwise, inequality (22) with $\{S,T\} = \{\bar{S},\bar{T}\}$ is violated by $(\bar{x},\bar{\theta})$. This, together with the fact that $\bar{S}$ and $\bar{T}$ can be computed in $\mathcal{O}(n)$ time, implies that the whole separation procedure can be conducted in $\mathcal{O}(n + f(n))$ time, where $\mathcal{O}(f(n))$ is the asymptotic polynomial upper bound for the calculation of $\mathbb{P}(\Gamma_{ij}^I)$. □ □

The efficient separation procedure for (22) enables us to develop a B&C algorithm for solving formulation (MILP); see Section 5 for the implementation details.

4.2 Enhanced expected traversal count inequalities

Next, we strengthen inequalities (22) so that a stronger LP relaxation bound can be obtained by the MILP formulation. The following lemma is useful for developing these strengthened inequalities.

Lemma 10. *Let $x \in \mathcal{T}$, $\{i,j\} \in E$, and $\{S,T\} \in \mathcal{P}_{ij}$ with $\varnothing \notin \{S,T\}$. Then*

$$\mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right) + \left(\mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^S) - \mathbb{P}(\Gamma_{ij}^T)\right) x_{ij}$$

$$\begin{cases} = \mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right), & \text{if } x_{ij} = 0; \\ \leq \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) x_{ij}, & \text{if } x_{ij} = 1. \end{cases}$$

Proof. Proof. Clearly, if $x_{ij} = 0$, the above inequality holds. Otherwise, $x_{ij} = 1$ and

$$\begin{aligned} &\mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right) + \left(\mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^S) - \mathbb{P}(\Gamma_{ij}^T)\right) x_{ij} \\ &= \mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right) + \left(\mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^S) - \mathbb{P}(\Gamma_{ij}^T)\right) \\ &\overset{(a)}{\leq} \mathbb{P}(\Gamma_{ij}^S) + \mathbb{P}(\Gamma_{ij}^T) + \left(\mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^S) - \mathbb{P}(\Gamma_{ij}^T)\right) = \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) = \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) x_{ij}, \end{aligned}$$

where (a) follows from Theorem 5. Therefore, the inequality in the statement also holds. □ □

Using Theorem 10 and (22)–(23), we can derive the following class of valid inequalities:

$$\begin{aligned} \theta_{ij} \geq\ & \mathbb{P}(\Gamma_{ij}^S)\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^T)\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})} x_{k\ell} - |T|\right) \\ & + \left(\mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^S) - \mathbb{P}(\Gamma_{ij}^T)\right) x_{ij},\ \forall\ \{i,j\} \in E,\ \{S,T\} \in \mathcal{P}_{ij} : \varnothing \notin \{S,T\}, \end{aligned} \tag{24}$$

which we refer to as enhanced ETC inequalities. Moreover, by Theorem 2, the coefficient of x_{ij} added in (24) is nonnegative; hence, since $x_{ij} \geq 0$, (24) dominates (22). Therefore, we can replace constraints (22) with their enhanced counterpart (24) in problem (MILP), obtaining an enhanced MILP formulation for the PTSP:

$$\begin{aligned} \min_{x\in\mathcal{T},\, \theta\in\mathbb{R}_+^{|E|}} \quad & \sum_{\{i,j\}\in E} c_{ij}\theta_{ij} \\ \text{s.t.} \quad & (23),\ (24). \end{aligned} \tag{E-MILP}$$

Similar to Theorem 9, the separation of the exponentially many enhanced ETC inequalities (24) at a point $(\bar{x}, \bar{\theta}) \in \mathcal{T} \times \mathbb{R}_+^{|E|}$ satisfying (23) can be conducted in an efficient fashion, enabling the use of a B&C algorithm to solve formulation (E-MILP).

Theorem 11. *Suppose that, given any $\{i,j\} \in E$ and $I \subseteq V \backslash \{i,j\}$, the probability $\mathbb{P}(\Gamma_{ij}^I)$ can be computed in polynomial time. Then, given any point $(\bar{x}, \bar{\theta}) \in \mathcal{T} \times \mathbb{R}_+^{|E|}$ satisfying (23), one can find an inequality in (24) violated by $(\bar{x}, \bar{\theta})$ or certify that none exists, in polynomial time.*

Proof. Proof. Similar to Theorem 9, it suffices to consider the separation of enhanced ETC inequalities (24) for each fixed $\{i,j\} \in E$. Let $\bar{S} = S_{ij}^{\bar{x}}$ and $\bar{T} = T_{ij}^{\bar{x}}$. From Theorem 7 and Theorem 10, if $\bar{\theta}_{ij} \geq \mathbb{P}(\Xi_{ij}^{\bar{x}}) + \mathbb{P}(\Xi_{ji}^{\bar{x}}) = \mathbb{P}(\Gamma_{ij}^{\bar{S}}) + \mathbb{P}(\Gamma_{ij}^{\bar{T}})$, then no inequality in (24) is violated by $(\bar{x}, \bar{\theta})$; otherwise, as inequality (23) holds at point $(\bar{x}, \bar{\theta}) \in \mathcal{T} \times \mathbb{R}_+^{|E|}$, $\bar{x}_{ij} = 0$ must hold, which implies that inequality (24) with $\{S, T\} = \{\bar{S}, \bar{T}\}$ is violated by $(\bar{x}, \bar{\theta})$. As $\bar{S}$ and $\bar{T}$ can be computed in $O(n)$ time, the whole separation procedure can be conducted in $O(n + f(n))$ time, where $O(f(n))$ is the asymptotic polynomial upper bound for the calculation of $\mathbb{P}(\Gamma_{ij}^I)$. □ □

Moreover, since the enhanced ETC inequalities (24) are stronger than the ETC inequalities (22), formulation (E-MILP) can provide a stronger LP relaxation bound than formulation (MILP). Indeed, as will be shown in Section 6.2, the LP relaxation of formulation (E-MILP) is usually much stronger than that of (MILP) without sacrificing the separation time, making (E-MILP) more attractive than (MILP) from a computational perspective.

5 Branch-and-cut algorithm and its implementation details

The overall algorithmic framework for solving formulation (E-MILP) (respectively, its basic version (MILP)) is an LP-based B&C algorithm in which the exponentially many enhanced ETC inequalities (24) (respectively, ETC inequalities (22)) and subtour elimination constraints (2b) are separated at the nodes of the search tree. In the following, we will present the separation procedures for the exponentially many inequalities and two acceleration techniques for the B&C algorithm. For simplicity, we only present results for the B&C algorithm based on the enhanced ETC inequalities (24) as those for the B&C algorithm based on ETC inequalities (22) can be described in a similar manner.

5.1 Separation

Let $(\bar{x}, \bar{\theta}) \in [0,1]^{|E|} \times \mathbb{R}_+^{|E|}$ be an LP solution encountered at a node of the search tree. We first describe the separation of the enhanced ETC inequalities (24) at $(\bar{x}, \bar{\theta})$. If $\bar{x} \in \mathcal{T}$, the separation of the enhanced ETC inequalities (24) can be performed efficiently as discussed in Theorem 11. Hence, in the following, we only discuss the implementation for the case $\bar{x} \notin \mathcal{T}$. In this case, we adopt a heuristic separation procedure as follows. We first construct a solution $\hat{x}$ that defines a tour in graph G from $\bar{x}$ by using a greedy procedure. Specifically, we sort all edges $\{i,j\} \in E$ as $\{i^1, j^1\}, \ldots, \{i^{|E|}, j^{|E|}\}$ in a non-increasing order of their LP values, that is, $\bar{x}_{i^1 j^1} \geq \cdots \geq \bar{x}_{i^{|E|} j^{|E|}}$, and set $\hat{E} \leftarrow \varnothing$. Then, for each $h = 1, \ldots, |E|$, we iteratively update $\hat{E} \leftarrow \hat{E} \cup \{\{i^h, j^h\}\}$ if doing so neither creates a subtour nor increases the degree of any vertex beyond two. This procedure is repeated until all vertices have degree two, yielding a Hamiltonian tour (note that as G is a complete graph, such a tour always exists). Letting $\hat{x}$ be the incidence vector of $\hat{E}$, i.e., $\hat{x}_{ij} = 1$ if $\{i,j\} \in \hat{E}$ and $\hat{x}_{ij} = 0$ otherwise, then $\hat{x} \in \mathcal{T}$. For each $\{i,j\} \in E$ with $\hat{x}_{ij} = 0$, letting $\hat{S} := S_{ij}^{\hat{x}}$ and $\hat{T} := T_{ij}^{\hat{x}}$, we then check whether the inequality (24) with $\{S, T\} = \{\hat{S}, \hat{T}\}$ is violated by point $(\bar{x}, \bar{\theta})$. In

Algorithm 1: Separation of inequalities (24) and (2b)

Input: An LP relaxation solution $(\bar{x}, \bar{\theta}) \in [0,1]^{|E|} \times \mathbb{R}_+^{|E|}$ satisfying (2a) and (23)

Output: A set C of violated inequalities.

1 Apply the Stoer-Wagner algorithm to find subtour elimination constraints (2b) violated by $(\bar{x}, \bar{\theta})$;

2 Set $C \leftarrow$ the set of violated subtour elimination constraints;

3 **if** $C = \varnothing$ **then**

4   If $\bar{x} \in \mathcal{T}$, set $\hat{x} \leftarrow \bar{x}$; otherwise, construct a Hamiltonian tour $\hat{x}$ from $\bar{x}$ using a greedy procedure;

5   **foreach** $\{i,j\} \in E$ *with* $\hat{x}_{ij} = 0$ **do**

6   Set $\hat{S} \leftarrow S_{ij}^{\hat{x}}$ and $\hat{T} \leftarrow T_{ij}^{\hat{x}}$;

7   If (25) holds, add the violated enhanced ETC inequality (24) with $\{S,T\} = \{\hat{S}, \hat{T}\}$ to C;

8 **return** C;

particular, if $\hat{x}_{ij} = 0$ and

$$
\begin{aligned}
\bar{\theta}_{ij} < \mathbb{P}(\Gamma_{ij}^{\hat{S}}) \left(\sum_{\{k,\ell\} \in E(\hat{S} \cup \{i,j\})} \bar{x}_{k\ell} - |\hat{S}| \right) + \mathbb{P}(\Gamma_{ij}^{\hat{T}}) \left(\sum_{\{k,\ell\} \in E(\hat{T} \cup \{i,j\})} \bar{x}_{k\ell} - |\hat{T}| \right) \\
+ \left(\mathbb{P}(\Gamma_{ij}^{V \setminus \{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing}) - \mathbb{P}(\Gamma_{ij}^{\hat{S}}) - \mathbb{P}(\Gamma_{ij}^{\hat{T}}) \right) \bar{x}_{ij},
\end{aligned}
\tag{25}
$$

then inequality (24) with $\{S,T\} = \{\hat{S}, \hat{T}\}$ is violated by $(\bar{x}, \bar{\theta})$.

We next describe the separation of the subtour elimination constraints (2b). For any $S \subseteq V$, using the degree constraints (2a), we obtain

$$
\sum_{\{i,j\} \in E(S)} x_{ij} = \frac{1}{2} \left(\sum_{i \in S} \sum_{j : \{i,j\} \in E} x_{ij} - \sum_{\{i,j\} \in \delta(S)} x_{ij} \right) = |S| - \frac{1}{2} \sum_{\{i,j\} \in \delta(S)} x_{ij},
$$

where $\delta(S) := \{\{i,j\} \in E : i \in S,\, j \notin S\}$ for $S \subseteq V$. Thus, the subtour elimination constraint (2b) associated with S is violated by $(\bar{x}, \bar{\theta})$ if and only if $\sum_{\{i,j\} \in \delta(S)} \bar{x}_{ij} < 2$. The latter is equivalent to the minimum cut problem that can be solved by the efficient polynomial-time Stoer-Wagner algorithm (Stoer and Wagner 1997); see also Padberg and Rinaldi (1990) and Applegate et al. (2011) for other algorithms.

The overall separation procedure for inequalities (24) and (2b) is summarized in Algorithm 1. Given an LP solution $(\bar{x}, \bar{\theta})$ at a node of the B&C tree, the procedure first separates the subtour elimination constraints (2b) using the Stoer-Wagner algorithm. If no violated subtour elimination constraint exists, the procedure proceeds to the separation of the enhanced ETC inequalities (24); see Steps 4–7. This guarantees that, if $\bar{x}$ is integral, then $\bar{x} \in \mathcal{T}$ must hold when separating the enhanced ETC inequalities.

Note that since there are a finite number of inequalities (2b) and (24), the exact separation of inequalities (2b) and (24) at points with integer vectors $\bar{x} \in \mathcal{T}$ is sufficient to ensure the convergence and correctness of the B&C algorithm for problem (E-MILP). Nevertheless, to accelerate the convergence of the B&C algorithm, we additionally perform separation of inequalities (2b) and (24) at fractional solutions encountered at the root node of the B&C search tree.

5.2 Two acceleration techniques

Next, we present two acceleration techniques to further improve the computational performance of the proposed B&C algorithm. In the first acceleration technique, we introduce a variable η_{ij} for each $\{i,j\} \in E$ and perform the following variable substitutions:

$$\theta_{ij} = \eta_{ij} + \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing})x_{ij}, \quad \forall\, \{i,j\} \in E, \tag{26}$$

in the two inequalities (23) and (24) in the B&C algorithm. This is equivalent to applying a B&C algorithm to the MILP reformulation (obtained by applying the variable substitutions in (26)):

$$\min_{x\in\mathcal{T},\eta} \sum_{\{i,j\}\in E} c_{ij}\left[\eta_{ij} + \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}}) + \mathbb{P}(\Gamma_{ij}^{\varnothing})x_{ij}\right] \tag{E-MILP'}$$

$$\text{s.t. } \eta_{ij} \geq \mathbb{P}(\Gamma_{ij}^{S})\left(\sum_{\{k,\ell\}\in E(S\cup\{i,j\})\setminus\{\{i,j\}\}} x_{k\ell} - |S|\right) + \mathbb{P}(\Gamma_{ij}^{T})\left(\sum_{\{k,\ell\}\in E(T\cup\{i,j\})\setminus\{\{i,j\}\}} x_{k\ell} - |T|\right)$$
$$- \mathbb{P}(\Gamma_{ij}^{V\setminus\{i,j\}})(1-x_{ij}),\ \forall\, \{i,j\} \in E,\ \forall\, \{S,T\} \in \mathcal{P}_{ij} : \varnothing \notin \{S,T\}, \tag{27a}$$

$$\eta_{ij} \geq 0,\ \forall\, \{i,j\} \in E. \tag{27b}$$

Here, inequalities (2b) and (27a) can be separated dynamically. Moreover, compared with the adapted B&C algorithm based on inequalities (23) and the enhanced ETC inequalities (24) that requires adding $|E|$ linear constraints (23) at the beginning, the B&C algorithm based on inequalities (27a) and (27b) only needs to add $|E|$ bound constraints rather than $|E|$ linear constraints. Based on our preliminary experiments, this alleviates the computational overhead associated with the LP relaxation, thereby accelerating the overall B&C algorithm.

In the second technique, we follow Laporte et al. (1994) and Wissink (2023) to add the following inequality at the beginning of the B&C algorithm:

$$\sum_{\{i,j\}\in E} c_{ij}\theta_{ij} \ \geq \sum_{\{i,j\}\in E} b_{ij}x_{ij}, \tag{28}$$

where, for each $\{i,j\} \in E$, b_{ij} stands for a conservative estimate of the expected contribution of edge $\{i,j\}$:

$$\begin{aligned} b_{ij} := \mathbb{P}(\xi_i = \xi_j = 1)c_{ij} &+ \frac{1}{2}\mathbb{P}\left(\xi_i = 1,\ \xi_j = 0,\ \sum_{k\in V\setminus\{i,j\}} \xi_k \geq 1\right) \min_{k\in V\setminus\{i,j\}} c_{ik} \\ &+ \frac{1}{2}\mathbb{P}\left(\xi_i = 0,\ \xi_j = 1,\ \sum_{k\in V\setminus\{i,j\}} \xi_k \geq 1\right) \min_{k\in V\setminus\{i,j\}} c_{kj}. \end{aligned} \tag{29}$$

The above inequality can be interpreted as follows. For an a priori edge $\{i,j\} \in E$, if both endpoints are present, its contribution is c_{ij}; if either i or j is present but not both and at least one other customer is present, the contribution is bounded below by the minimum edge cost of reconnecting from the present endpoint i or j, namely, $\min_{k\neq i,j} c_{ik}$ or $\min_{k\neq i,j} c_{kj}$, respectively; otherwise, the contribution is zero. The factor $1/2$ symmetrically allocates the cost of the resulting a posteriori edge, ensuring that the bound does not overestimate the total contribution.

Remark 12. *Inequality (28) differs slightly from the inequality proposed by Laporte et al. (1994) and Wissink (2023), namely,*

$$\sum_{\{i,j\}\in E} c_{ij}\theta_{ij} \geq \sum_{\{i,j\}\in E} \hat{b}_{ij}x_{ij} \tag{30}$$

where, for each edge $\{i, j\} \in E$, the coefficient $\hat{b}_{ij}$ is defined as

$$\hat{b}_{ij} := \mathbb{P}(\xi_i = \xi_j = 1)c_{ij} + \frac{1}{2}\mathbb{P}(\xi_i = 1,\ \xi_j = 0) \min_{k\neq i,j} c_{ik} + \frac{1}{2}\mathbb{P}(\xi_i = 0,\ \xi_j = 1) \min_{k\neq i,j} c_{kj}, \tag{31}$$

Although $\hat{b}_{ij} \geq b_{ij}$, replacing b_{ij} with $\hat{b}_{ij}$ does not generally preserve validity. Indeed, in Online Appendix C, we present a counterexample demonstrating that inequality (30) is generally not a valid inequality for the PTSP.

To end this subsection, we would like to highlight the advantage of the proposed B&C algorithm over the state-of-the-art integer L-shaped method of Laporte et al. (1994) and Wissink (2023). In particular, the integer L-shaped method is also a B&C algorithm based on integer L-shaped cuts, whose derivation only uses the function value of a point $\bar{x}$, thereby often leading to weak cuts and forcing the B&C algorithm to explore a huge search tree. In contrast, the proposed B&C algorithm is based on inequalities (23) and the enhanced ETC inequalities (24) (or their adapted version (27a) and (27b)), which exploit the structure of the a priori solution $x \in \mathcal{T}$ and the shortcutting recourse policy, thereby usually leading to much stronger cuts. Indeed, as will be shown in Section 6.1, the proposed inequalities (23) and the enhanced ETC inequalities (24) are usually much stronger than the integer L-shaped cuts in terms of providing a substantially tighter LP relaxation bound. With this advantage, it can be expected that the proposed B&C algorithm based on inequalities (23) and (24) is much more computationally efficient than the integer L-shaped method of Laporte et al. (1994) and Wissink (2023); see Section 6.1 further ahead.

6 Computational results

In this section, we present computational results to demonstrate the efficiency of the proposed B&C algorithm. To do this, we first perform numerical experiments comparing the computational performance of our proposed B&C algorithm based on formulation (E-MILP) with the state-of-the-art integer L-shaped method of Laporte et al. (1994) and Wissink (2023) on a testset of PTSP instances with up to 100 vertices. Then, we report computational results to demonstrate the advantage of the enhanced ETC inequalities (24). Subsequently, we study the scalability of our proposed algorithm on a testset of PTSP instances with the number of customers ranging from 101 to 300. Finally, we compare the performance of the proposed B&C algorithm with the state-of-the-art heuristic algorithm of Wissink (2023). Note that we do not investigate managerial insights such as the gap between the TSP and PTSP solutions and the impact of correlations among customer presences, as they were investigated in previous work (Jaillet 1985, Wissink 2023).

The proposed B&C algorithm was implemented in Julia 1.11.6 using Cplex 20.1.0. We set parameters of Cplex to run the code with a time limit of 7200 seconds and a relative MIP gap tolerance of 0%. Except for the ones mentioned above, other parameters of Cplex were set to their default values. All computational experiments were performed on a cluster of Intel(R) Xeon(R) Gold 6140 CPU @ 2.30 GHz computers.

Benchmark instances. In our computational experiments, we construct PTSP instances based on 48 deterministic TSP instances (available from http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/) satisfying the triangle inequality from TSPLIB with up to 300 vertices; see Tables 1 and 3. The underlying graph $G = (V, E)$ and the associated edge costs $\{c_{ij}\}_{\{i,j\}\in E}$ are directly adopted from the corresponding TSP instances. For each TSP instance, we construct two independent PTSP instances, corresponding to two different levels of uncertainty. Specifically, each customer presence probability p_i is randomly drawn from $[0.95, 1]$ and $[0.90, 1]$, respectively. For each set of customer presence probabilities corresponding to an independent PTSP instance, we follow Wissink (2023) to further construct 6 correlated

PTSP instances using Clayton, Gumbel, and truncated D-vine pair Gaussian copulas, described in Online Appendix B, with two dependence levels {low, high} for customer presence; see Online Appendix D for the details. In total, we generate $48 \times 2 \times (1 + 6) = 672$ PTSP instances. These instances are divided into two testsets T1 and T2 according to their size. The first testset T1 contains 280 PTSP instances with at most 100 vertices, and the second testset T2 consists of 392 PTSP instances with the number of vertices ranging from 101 to 300.

6.1 Comparison with the state-of-the-art integer L-shaped method of Laporte et al. (1994), Wissink (2023)

We first compare the performance of the proposed B&C algorithm based on formulation (E-MILP), denoted by `E-MILP`, with the state-of-the-art integer L-shaped method of Laporte et al. (1994) and Wissink (2023), denoted by `ILSM`, where the valid inequality (28) was adapted to $\zeta \geq \sum_{\{i,j\}\in E} b_{ij}x_{ij}$ and implemented in `ILSM`. Tables 1 and 2 summarize the computational results of `ILSM` and `E-MILP` on instances in testset T1, where customer presence probabilities p_i are drawn from the intervals [0.95, 1] and [0.90, 1], respectively. For each row of Tables 1 and 2, we report the computational results of the 7 and 20 PTSP instances, grouped by the TSP instance and the distribution of ξ, respectively. Specifically, for each of `ILSM` and `E-MILP`, we report the number of instances solved to optimality within the time limit of 7200 seconds (`# S`), the average CPU time in seconds (`T [s]`), the average number of explored nodes (`# N`), the average optimality gap in percentage returned by CPLEX at termination (`G [%]`), and the average LP relaxation gap at the root node (`rG [%]`), computed as $\frac{o^* - o_{\text{root}}}{o^*} \times 100$, where o_{root} and o^* are the LP relaxation bound obtained at the root node and the objective value of the optimal solution (or the best incumbent) of the PTSP instance, respectively. At the end of the table, we report the summarized results for all instances.

From Table 1, we first observe in column `rG [%]` that, for instances in testset T1 with $p_i \in [0.95, 1]$, `E-MILP` provides much stronger LP relaxation bounds at the root node compared to `ILSM`. In particular, the average LP relaxation gap under `E-MILP` is only 0.8%, whereas that under `ILSM` is 1.9%. This confirms the advantage of the newly proposed MILP formulation (E-MILP) based on inequalities (23) and (24) for the PTSP—it can provide a much stronger LP relaxation bound than the MILP formulation based on the integer L-shaped optimality cuts (3). Due to this advantage, `E-MILP` significantly outperforms `ILSM`. More specifically, for PTSP instances with $p_i \in [0.95, 1]$ in testset T1, due to the huge search tree, `ILSM` can only solve 56 PTSP instances with up to 52 vertices within the time limit of 7200 seconds; in contrast, the proposed `E-MILP` is able to solve all 140 instances to optimality with an average CPU time of 44.2 seconds and an average number of explored nodes of 5720.

Next, we compare the performance of `ILSM` and `E-MILP` on PTSP instances with $p_i \in [0.90, 1]$ in testset T1. From Tables 1 and 2, we observe that for both `ILSM` and `E-MILP`, due to a large LP relaxation gap at the root node, PTSP instances with $p_i \in [0.90, 1]$ are generally more difficult to solve than those with $p_i \in [0.95, 1]$. This is consistent with the empirical findings in Laporte et al. (1994) and Wissink (2023) where the authors observed that PTSP instances with low probabilities of the presence of customers are generally harder to solve than those with high probabilities. Nevertheless, the performance of `E-MILP` remains relatively robust, compared with that of `ILSM`. Indeed, as shown in Table 2, `ILSM` solves only instances with up to 42 vertices, accounting for 35 out of 140 instances, while `E-MILP` can still solve all but 8 instances to optimality with an average CPU time of 585.5 seconds and an average number of explored nodes of 44397.

To provide a more intuitive comparison of the performance of the proposed `E-MILP` and `ILSM`, we plot in Figure 2 the empirical cumulative distribution functions (ECDFs) of the LP relaxation gap at the root node, the number of explored nodes, the CPU time, and the optimality gap at termination under settings

Table 1: Performance comparison of settings ILSM and E-MILP on PTSP instances with $p_i \in [0.95, 1.0]$.

	# inst	ILSM					E-MILP				
		# S	T [s]	# N	G [%]	rG [%]	# S	T [s]	# N	G [%]	rG [%]
burma14	7	7	0.1	30	0.0	1.0	7	<0.1	2	0.0	<0.1
ulysses16	7	7	1.3	946	0.0	1.9	7	0.1	17	0.0	0.8
ulysses22	7	7	13.5	6298	0.0	2.0	7	0.2	12	0.0	0.4
fri26	7	7	0.7	326	0.0	1.5	7	0.2	6	0.0	0.1
bayg29	7	7	1.9	563	0.0	1.3	7	0.3	19	0.0	0.4
swiss42	7	7	22.8	7088	0.0	1.9	7	0.4	6	0.0	<0.1
att48	7	3	4638.2	228561	0.4	1.9	7	1.1	68	0.0	0.9
eil51	7	7	486.5	56691	0.0	1.5	7	1.6	145	0.0	0.5
berlin52	7	4	4324.8	198540	0.2	1.7	7	0.4	2	0.0	<0.1
st70	7	0	7200.0	472034	1.5	1.9	7	5.4	616	0.0	1.3
eil76	7	0	7200.0	374361	1.1	1.6	7	3.1	125	0.0	0.4
pr76	7	0	7200.0	472006	1.6	3.5	7	708.0	97913	0.0	2.6
gr96	7	0	7200.0	657740	1.6	2.2	7	22.3	1852	0.0	1.8
rat99	7	0	7200.0	436186	1.3	1.8	7	6.3	148	0.0	0.7
kroA100	7	0	7200.0	622731	1.6	2.1	7	15.6	1661	0.0	1.1
kroB100	7	0	7200.0	615465	1.7	2.3	7	36.4	3011	0.0	1.6
kroC100	7	0	7200.0	683667	1.5	1.9	7	7.7	240	0.0	1.0
kroD100	7	0	7200.0	653209	1.7	2.0	7	11.1	796	0.0	1.0
kroE100	7	0	7200.0	585320	1.6	2.4	7	57.3	7649	0.0	1.6
rd100	7	0	7200.0	769096	1.8	2.0	7	6.4	104	0.0	0.5
Independent	20	7	4704.2	404851	1.0	2.1	20	8.4	3561	0.0	0.9
Clayton-high	20	9	3967.2	261919	0.5	1.3	20	13.1	2799	0.0	0.7
Clayton-low	20	9	4079.4	347521	0.8	1.8	20	93.2	5178	0.0	0.9
Gumbel-high	20	9	4127.2	359142	0.8	1.8	20	100.7	8022	0.0	0.9
Gumbel-low	20	8	4674.9	407046	1.0	2.1	20	18.8	7032	0.0	0.8
Gaussian-high	20	7	4744.3	298051	1.0	2.1	20	58.3	10024	0.0	0.8
Gaussian-low	20	7	4744.2	315771	1.1	2.2	20	16.8	3421	0.0	0.8
Ave.	140	56	4434.5	342043	0.9	1.9	140	44.2	5720	0.0	0.8

E-MILP and ILSM on instances in testset T1. The results show that the red diamond lines corresponding to the proposed E-MILP are much higher than the black circle lines corresponding to ILSM, demonstrating the computational efficiency of the proposed E-MILP over ILSM. In particular, Figure 2a shows that E-MILP achieves an LP relaxation gap at the root node within 1% for about 40% of the instances in testset T1, while ILSM achieves the same for almost none of these instances; Figure 2b shows that E-MILP can solve over 80% of the instances within 10000 explored nodes, whereas only about 20% of them are solved by ILSM within the same number of explored nodes; Figure 2c shows that over 75% of the instances are solved by E-MILP within 100 seconds, whereas only about 25% of them are solved by ILSM within the same period of time; Figure 2d shows that E-MILP can solve almost all instances to optimality within the two-hour time limit, whereas ILSM terminates with an optimality gap above 1% for more than 60% of the instances.

6.2 Empirical effect of enhanced ETC inequalities

As discussed in Section 4.2, from a theoretical perspective, formulation (E-MILP) with the enhanced ETC inequalities (24) provides a stronger LP relaxation bound than formulation (MILP) with the ETC

Table 2: Performance comparison of settings `ILSM` and `E-MILP` on PTSP instances with $p_i \in [0.90, 1.0]$.

	#inst	ILSM					E-MILP				
		#S	T [s]	#N	G [%]	rG [%]	#S	T [s]	#N	G [%]	rG [%]
burma14	7	7	0.5	233	0.0	2.6	7	2.5	10	0.0	0.7
ulysses16	7	7	20.9	11363	0.0	3.8	7	0.2	75	0.0	2.0
ulysses22	7	4	4051.8	202531	0.4	4.2	7	0.4	121	0.0	1.7
fri26	7	7	15.0	7595	0.0	3.3	7	0.3	30	0.0	0.9
bayg29	7	7	108.0	24675	0.0	2.7	7	0.8	228	0.0	1.8
swiss42	7	3	5395.8	300234	1.0	3.8	7	0.8	50	0.0	0.9
att48	7	0	7200.0	634440	2.7	3.6	7	4.2	577	0.0	2.4
eil51	7	0	7200.0	453339	2.0	3.1	7	23.0	2482	0.0	1.8
berlin52	7	0	7200.0	626149	2.6	3.4	7	1.0	43	0.0	0.5
st70	7	0	7200.0	791128	3.4	3.9	7	147.1	13395	0.0	2.8
eil76	7	0	7200.0	730328	2.9	3.3	7	86.3	3392	0.0	1.6
pr76	7	0	7200.0	747758	3.8	5.6	0	7200.0	582756	0.9	4.3
gr96	7	0	7200.0	930610	3.4	4.0	7	388.6	21853	0.0	3.2
rat99	7	0	7200.0	842005	3.2	3.6	7	64.2	4176	0.0	1.8
kroA100	7	0	7200.0	924489	3.5	3.9	7	309.8	26599	0.0	2.7
kroB100	7	0	7200.0	893592	3.7	4.2	7	1203.6	79181	0.0	3.3
kroC100	7	0	7200.0	922161	3.4	3.6	7	62.0	3880	0.0	2.0
kroD100	7	0	7200.0	909130	3.6	4.0	7	271.6	20555	0.0	2.5
kroE100	7	0	7200.0	899725	3.5	4.2	6	1833.9	121758	<0.1	3.2
rd100	7	0	7200.0	907304	3.8	4.0	7	108.8	6783	0.0	1.9
Independent	20	4	5765.3	679418	2.7	4.2	19	438.4	71640	<0.1	2.3
Clayton-high	20	6	5051.9	522172	1.5	2.6	19	390.5	35055	<0.1	1.5
Clayton-low	20	6	5265.9	617654	2.1	3.5	18	1132.0	17950	0.1	2.0
Gumbel-high	20	6	5265.9	612843	2.1	3.4	19	758.1	14157	<0.1	2.0
Gumbel-low	20	5	5727.8	636403	2.5	4.0	19	482.5	72231	<0.1	2.2
Gaussian-high	20	4	5774.1	517813	2.6	4.0	19	454.2	30415	<0.1	2.3
Gaussian-low	20	4	5786.4	529273	2.8	4.3	19	442.5	69334	<0.1	2.3
Ave.	140	35	5519.6	587939	2.3	3.7	132	585.5	44397	<0.1	2.1

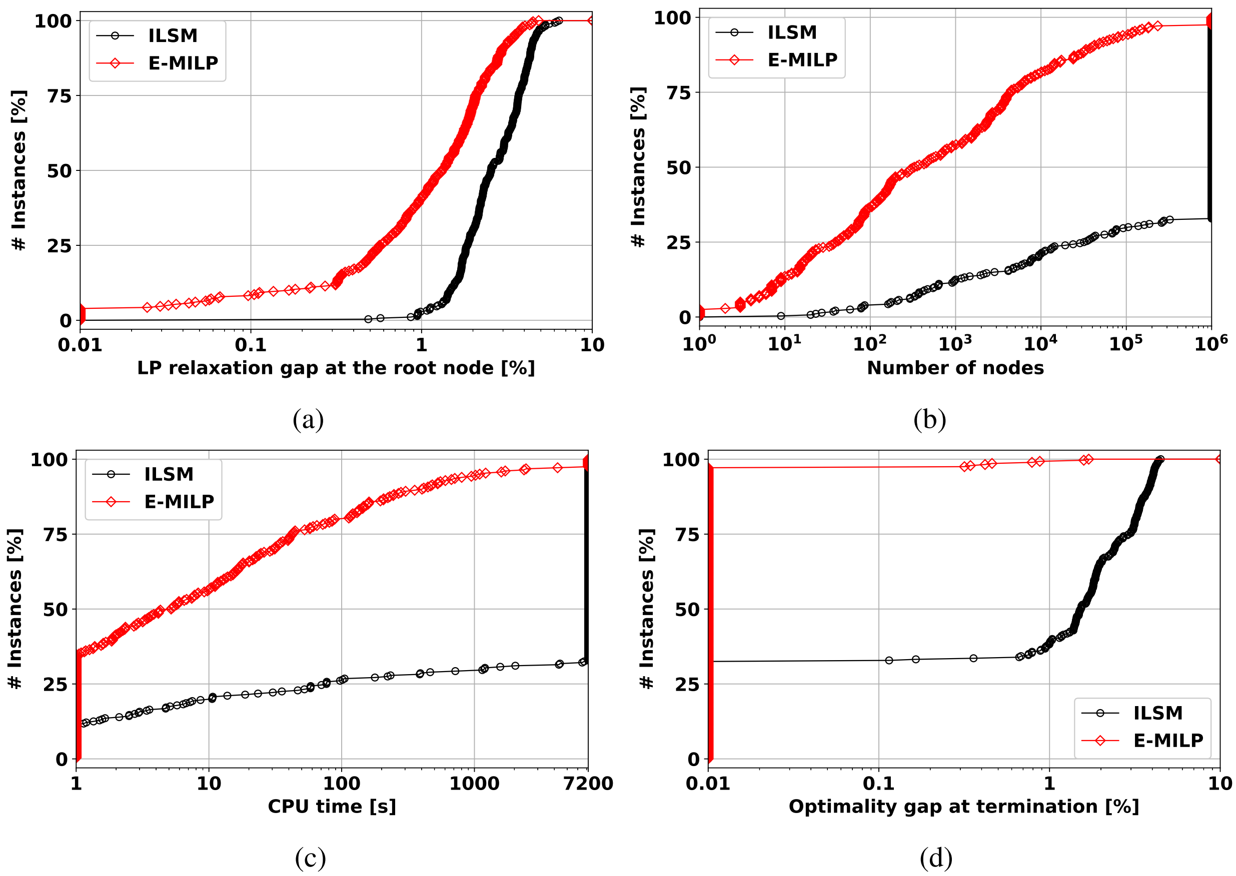


Figure 2: ECDFs of (a) the LP relaxation gap at the root node, (b) number of nodes, (c) CPU time, and (d) optimality gap at termination for `ILSM` and `E-MILP` on instances in testset T1.

inequalities (22). Here, to evaluate the effect of the enhanced ETC inequalities (24), we further compare the computational performance of the B&C algorithms based on formulation (E-MILP) and (MILP), denoted by `E-MILP` and `MILP`, respectively. The results are summarized in Figure 3, where we plot the ECDFs of the LP relaxation gap at the root node, the number of explored nodes, the CPU time, and the optimality gap at termination under settings `E-MILP` and `MILP` on instances in testset T1. Detailed results are reported in Tables E.1 and E.2 in Online Appendix E.

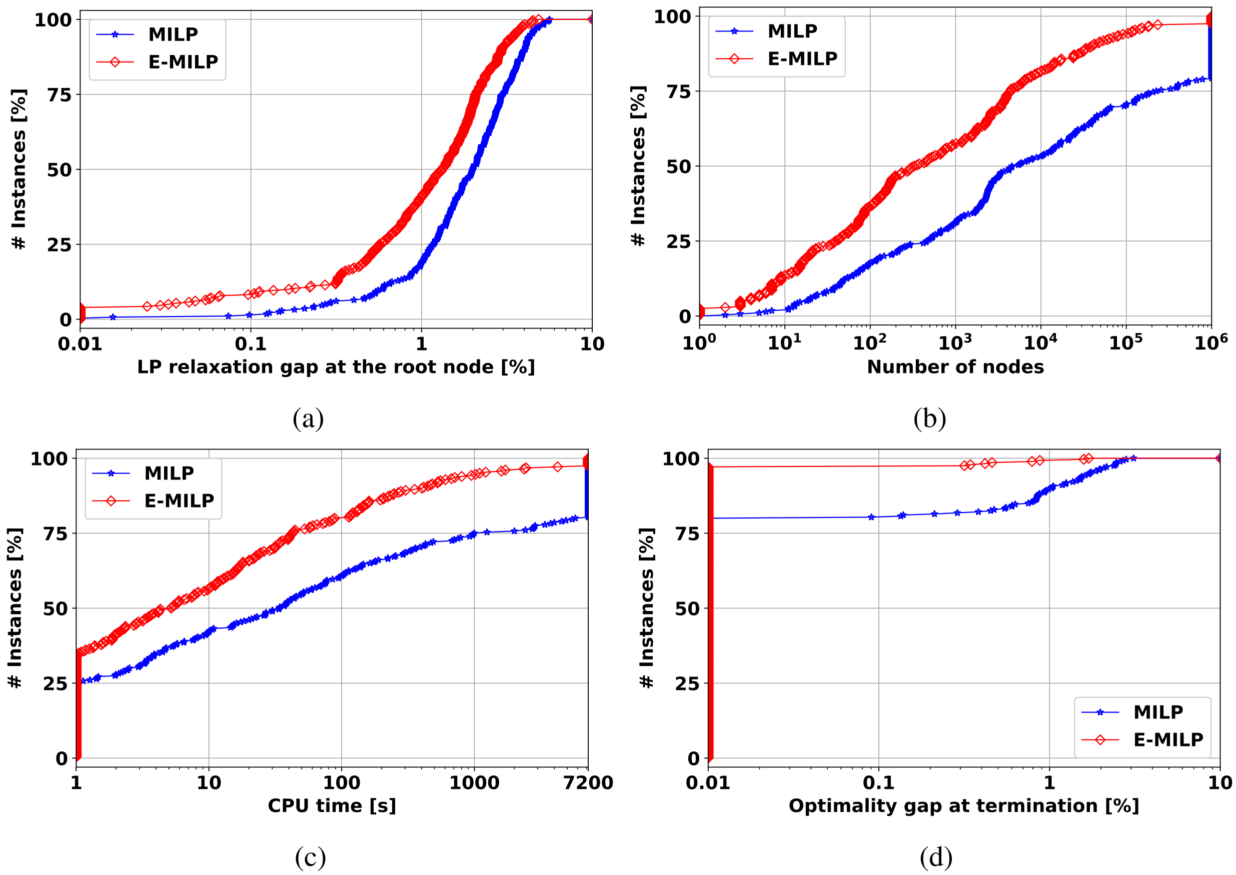


Figure 3: ECDFs of (a) the LP relaxation gap at the root node, (b) number of nodes, (c) CPU time, and (d) optimality gap at termination for `MILP` and `E-MILP` on instances in testset T1.

As shown in Figure 3a, the red diamond line corresponding to `E-MILP` is consistently higher than the blue star line corresponding to `MILP`. This confirms that formulation (E-MILP) with the enhanced ETC inequalities (24) can indeed provide a (much) stronger LP relaxation bound than formulation (MILP) with the ETC inequalities (22). As a result, `E-MILP` achieves an overall better computational performance than `MILP`. In particular, as shown in Figure 3b, over 80% of the instances are solved by `E-MILP` within 10000 nodes, whereas about 50% of the instances are solved by `MILP` within the same number of nodes; as shown in Figure 3c, over 75% of the instances are solved by `E-MILP` within 100 seconds, whereas about 60% of the instances are solved by `MILP` within the same period of time; as shown in Figure 3d, `E-MILP` can solve almost all instances to optimality, while `MILP` yields an optimality gap above 1% for nearly 10% of the instances.

6.3 Scalability of our algorithm

To assess the scalability of the proposed algorithm, we evaluate the performance of `E-MILP` on PTSP instances with vertices ranging from 101 to 300 in testset T2. The computational results are summarized in Table 3, where, for each row, we additionally report the average number of added cuts (`#C`), i.e., (2b) and (24), as well as the CPU time spent on separating these cuts (`CT [s]`), to provide more insight into the proposed `E-MILP`.

We first consider PTSP instances in testset T2 with customer presence probabilities $p_i \in [0.95, 1]$. From Table 3, we observe that PTSP instances generally become difficult to solve as the number of vertices increases. This is quite expected since with the increase of the number of vertices, the problem size of PTSP instances becomes larger and the gap between (E-MILP) and its LP relaxation generally becomes larger. Regarding the effect of customer presence probabilities, `E-MILP` solves 147 of the 196 PTSP instances in testset T2 with $p_i \in [0.95, 1]$ to optimality, whereas its performance deteriorates substantially when $p_i \in [0.90, 1]$, solving only 57 of the 196 instances to optimality. Nevertheless, for instances that cannot be solved to optimality, the average optimality gaps at termination are relatively small, usually less than 2%. This shows the effectiveness of the proposed `E-MILP` in finding a high-quality solution for the PTSP (when `E-MILP` fails to find an optimal solution). Another finding from Table 3 is that the average CPU time spent on separating cuts (`CT [s]`) for instances in the `Gaussian-high` and `Gaussian-low` groups is about one order of magnitude larger than that for other groups. This is reasonable as the computation of probabilities $\mathbb{P}(\Gamma_{ij}^{I})$ under Gaussian copulas is much more expensive than their counterparts under other copulas and the independent distribution of ξ; see Online Appendix B.

6.4 Comparison with the state-of-the-art heuristic algorithm of Wissink (2023)

In this section, we compare the performance of the proposed B&C algorithm with the state-of-the-art 2.5-opt-EEs heuristic algorithm, which was proposed by Birattari et al. (2008) for the independent PTSP, and later extended by Wissink (2023) to the correlated PTSP. The 2.5-opt-EEs algorithm is an iterative improvement algorithm that uses an empirical-estimation based approach to compute the difference in the traveling cost between two solutions. To improve diversification and obtain a better solution, an iterated local search procedure of Balaprakash et al. (2010) can be applied, which repeatedly perturbs the incumbent local optimum and applies the 2.5-opt-EEs algorithm. The resulting algorithm is referred to as `ILS-EEs`. In our experiments, we use the same implementation as in Wissink (2023).

The purpose of this experiment is to compare the quality of feasible solutions obtained within a limited computational budget. Both `ILS-EEs` and `E-MILP` are run with a time limit of 600 seconds. We quantify the quality of the solutions by the primal gap:

$$\frac{o_s - o^{7200}}{o^{7200}} \times 100,$$

where o_s is the objective value of the best solution returned within 600 seconds by method $s \in \{\texttt{ILS-EEs}, \texttt{E-MILP}\}$, and o^{7200} is the objective value of the optimal solution or the best incumbent found by `E-MILP` within a time limit of 7200 seconds (which serves as a high-quality reference solution).

Figures 4a and 4b plot the ECDFs of the primal gaps for instances with $p_i \in [0.95, 1.0]$ and $p_i \in [0.90, 1.0]$, respectively. The results show the advantage of the proposed `E-MILP` in finding a high-quality solution within a restricted time budget in addition to its ability to prove optimality. In particular, for PTSPs with $p_i \in [0.95, 1.0]$, `E-MILP` finds 600-second incumbents within a gap of 0.01% for more than 95% of the instances in testset T1 and for around 70% of the instances in testset T2. For PTSPs with $p_i \in [0.90, 1.0]$, the corresponding proportions remain about 90% for testset T1 and 50% for testset T2.

Table 3: Performance of `E-MILP` on PTSP instances in testset T2

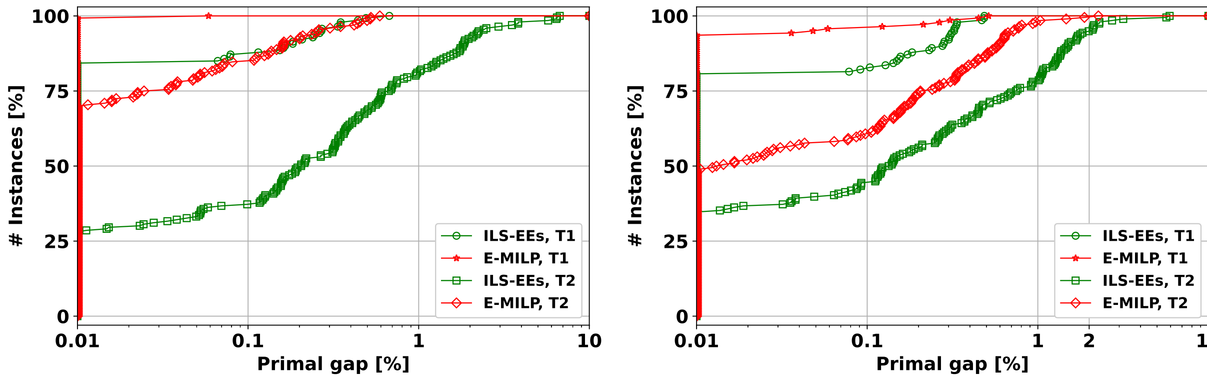
	#inst	E-MILP-[0.95,1]							E-MILP-[0.90,1]						
		#S	T[s]	#N	G[%]	rG[%]	#C	CT[s]	#S	T[s]	#N	G[%]	rG[%]	#C	CT[s]
eil101	7	7	13.0	1050	0.0	0.6	3265	5.8	7	1708.7	63114	0.0	1.9	14192	14.4
lin105	7	7	4.3	41	0.0	0.2	2913	3.2	7	22.5	1821	0.0	1.1	6585	6.2
pr107	7	7	34.0	3330	0.0	1.6	5774	7.7	1	6353.4	189894	1.2	4.4	29030	27.4
pr124	7	7	18.7	736	0.0	1.9	6597	9.7	7	159.4	7424	0.0	2.8	15890	17.7
bier127	7	7	15.9	888	0.0	0.6	5473	9.7	7	817.3	28533	0.0	1.8	16527	32.0
ch130	7	7	89.6	5700	0.0	1.1	8599	16.8	1	7018.6	528134	0.3	3.1	18588	32.3
pr136	7	7	2674.6	99663	0.0	2.3	15334	26.1	0	7200.0	330861	1.2	4.2	34120	45.5
gr137	7	7	41.7	2631	0.0	1.0	9334	16.5	5	3454.0	137115	0.1	2.5	23947	36.9
pr144	7	7	79.6	4114	0.0	1.7	12160	23.3	1	6307.5	206096	1.1	4.0	30674	50.3
ch150	7	7	164.9	6238	0.0	1.1	10436	24.7	5	4370.5	293920	0.2	2.7	21041	53.8
kroA150	7	7	94.8	4064	0.0	1.4	8430	28.4	4	4362.9	279847	0.2	3.2	18731	54.7
kroB150	7	7	499.3	41973	0.0	1.8	9604	28.9	4	5967.7	478956	0.2	3.2	17162	47.2
pr152	7	7	549.0	50126	0.0	1.9	11995	35.5	0	7200.0	173624	2.2	4.7	34239	70.0
u159	7	7	26.6	993	0.0	0.6	7756	18.1	7	282.8	18503	0.0	1.7	13980	32.2
si175	7	0	7200.0	162466	0.4	0.5	30221	108.1	0	7200.0	73578	0.9	1.0	41898	169.3
rat195	7	7	1638.8	45127	0.0	2.1	14111	82.5	0	7200.0	198220	0.7	3.4	32141	142.4
d198	7	5	3386.6	104326	<0.1	2.0	25459	111.3	0	7200.0	71510	2.2	5.0	56066	201.1
kroA200	7	1	7019.4	679146	0.3	1.7	12768	78.0	0	7200.0	367446	1.6	3.4	24843	115.9
kroB200	7	7	819.6	26409	0.0	1.6	11569	68.5	0	7200.0	371187	0.7	3.5	22058	115.0
gr202	7	7	546.3	13727	0.0	0.9	19074	84.6	1	7104.4	193482	0.4	2.4	34211	150.4
ts225	7	0	7200.0	197403	4.2	6.5	13652	151.2	0	7200.0	149047	5.2	8.6	23278	181.8
tsp225	7	5	3711.7	173069	<0.1	1.9	11976	124.5	0	7200.0	170332	1.6	4.0	26457	216.9
pr226	7	3	5867.2	180934	<0.1	1.3	33914	226.3	0	7200.0	91105	1.8	3.8	59842	310.0
gr229	7	3	4882.5	142283	0.1	1.8	22024	149.7	0	7200.0	152834	1.2	4.1	38341	223.1
gil262	7	0	7200.0	367977	0.3	1.9	20242	201.9	0	7200.0	200095	1.2	3.9	28350	255.7
pr264	7	7	229.2	4682	0.0	0.9	16441	123.2	0	7200.0	135428	0.7	2.7	50391	450.8
a280	7	4	4143.3	126742	<0.1	0.9	28746	216.6	0	7200.0	224551	0.7	2.9	29720	234.5
pr299	7	0	7200.0	266296	0.9	2.5	21570	294.2	0	7200.0	152495	2.2	4.1	33963	363.1
Independent	28	23	1585.3	133309	0.2	1.7	4602	12.9	10	5229.3	366356	0.9	3.8	9897	20.6
Clayton-high	28	21	2126.4	94260	0.2	1.3	8054	14.1	12	4866.6	146593	0.7	2.4	15284	23.9
Clayton-low	28	18	3256.3	35704	0.3	1.6	32290	21.7	5	6194.1	32903	1.3	3.1	57417	37.7
Gumbel-high	28	18	3063.2	47783	0.3	1.6	27083	16.5	5	6225.7	45402	1.2	3.3	55332	29.7
Gumbel-low	28	22	1765.0	111122	0.2	1.6	8356	13.2	8	5588.8	245016	0.9	3.7	18210	20.9
Gaussian-high	28	22	2666.0	133350	0.2	1.6	11877	236.7	8	5532.5	197728	1.1	3.5	27889	389.1
Gaussian-low	28	23	1875.3	122504	0.2	1.7	7599	253.7	9	5345.5	288291	0.9	3.8	15037	390.7
Ave.	196	147	2333.9	96862	0.2	1.6	14266	81.2	57	5568.9	188898	1.0	3.4	28438	130.4

For the remaining instances, the primal gaps are usually smaller than 1%. Thus, even under a time limit of 600 seconds, the proposed `E-MILP` can already produce feasible solutions that are very close to the best solutions found in long 7200-second runs.

The comparison also highlights the remaining improvement potential of the state-of-the-art heuristic algorithm of `ILS-EEs`. In particular, the primal gaps of `ILS-EEs` are non-negligible on a sizeable proportion of instances, especially for the larger and more difficult instances in testset T2. This is quite expected as the heuristic algorithm `ILS-EEs` uses empirical estimates to evaluate local-search moves and may become trapped in local optima, whereas `E-MILP` exploits the global structure of the PTSP through the enhanced ETC inequalities and the B&C search. Overall, these results indicate that the proposed exact algorithm is useful not only for certifying optimality or near-optimality, but also as an effective primal heuristic when run under a restricted time budget.

7 Conclusion and future work

This paper develops an exact algorithm for the PTSP with Bernoulli customer presences under a general joint distribution. The key methodological contribution is a new tractable convex piecewise-affine representation of the expected recourse function in the space of first-stage tour variables. The representation is obtained



(a) $p_i \in [0.95, 1.0]$

(b) $p_i \in [0.90, 1.0]$

Figure 4: ECDFs of the primal gap of the 600-second incumbents produced by `ILS-EEs` and `E-MILP` with respect to the best incumbent obtained by `E-MILP` within 7200 seconds.

by decomposing the expected a posteriori tour cost into an edge-cost-weighted sum of ETCs and exploiting the structure of the a priori tour and recourse policy. For each edge, the corresponding ETC is shown to be representable as the maximum of finitely many affine functions. This result provides an explicit convex description of a recourse function that has previously been treated primarily as a black box.

The convex representation leads directly to an exact MILP formulation for the PTSP over the full probability space that does not rely on scenario sampling or scenario enumeration. We further strengthen this MILP formulation through enhanced ETC inequalities. The resulting inequalities yield substantially stronger LP relaxations and can be separated dynamically within a B&C framework. Under standard oracle assumptions for evaluating the relevant joint probabilities, the separation of these inequalities at tour solutions can be performed efficiently, covering both independent and correlated PTSPs considered in the literature.

The computational results demonstrate the effectiveness of the proposed algorithm. Compared with the integer L-shaped method of Laporte et al. (1994) and Wissink (2023), the proposed exact algorithm gives stronger root-node bounds and much smaller enumeration trees. It solves instances with up to 300 vertices to optimality or near-optimality within two hours, with remaining gaps usually below 2%. Against the leading heuristic algorithm of Wissink (2023), the proposed algorithm also returns better 600-second incumbents, which are usually within 1% of the best 7200-second reference values. These results indicate that the proposed convex recourse representation provides a strong algorithmic foundation for exact optimization and high-quality primal solutions for the PTSP.

It is worth noting that the convex piecewise-affine representation (21) of the expected recourse function relies only on the condition that $x \in \mathcal{T}$ (i.e., x defines a tour in graph G) and on the shortcutting recourse policy. Specifically, for any pair of customers i and j, the a priori tour contains two internally vertex-disjoint paths connecting them, and these paths may shrink into edges between i and j in the a posteriori tour after absent customers are skipped. This favorable feature suggests that the representation in Theorem 8 is not limited to the standard PTSP, but may also apply to variants that incorporate practical considerations, such as PTSPs with deadlines (Campbell and Thomas 2008, Weyland et al. 2012) or time windows (Voccia et al. 2013), as well as the probabilistic traveling salesman facility location problem (Bertsimas 1989, Bertsimas et al. 1990). Consequently, extending the proposed exact B&C algorithm to these PTSP variants and conducting a computational investigation would be an interesting direction for future research.

References

Applegate DL, Bixby RE, Chvátal V, Cook WJ (2011) *The Traveling Salesman Problem: A Computational Study* (Princeton University Press).

Balaprakash P, Birattari M, Stützle T, Dorigo M (2010) Estimation-based metaheuristics for the probabilistic traveling salesman problem. *Computers & Operations Research* 37(11):1939–1951.

Bazirha M, Kadrani A, Benmansour R (2023) Stochastic home health care routing and scheduling problem with multiple synchronized services. *Annals of Operations Research* 320(2):573–601.

Bertsimas D (1988) *Probabilistic Combinatorial Optimization Problems*. Ph.D. thesis, Massachusetts Institute of Technology.

Bertsimas D (1989) Traveling salesman facility location problems. *Transportation Science* 23(3):184–191.

Bertsimas D, Howell LH (1993) Further results on the probabilistic traveling salesman problem. *European Journal of Operational Research* 65(1):68–95.

Bertsimas D, Jaillet P, Odoni AR (1990) A priori optimization. *Operations Research* 38(6):1019–1033.

Bertsimas D, Natarajan K, Teo CP (2006) Persistence in discrete optimization under data uncertainty. *Mathematical Programming* 108(2-3):251–274.

Bianchi L, Campbell AM (2007) Extension of the 2-p-opt and 1-shift algorithms to the heterogeneous probabilistic traveling salesman problem. *European Journal of Operational Research* 176(1):131–144.

Bianchi L, Gambardella LM, Dorigo M (2002) Solving the homogeneous probabilistic traveling salesman problem by the ACO metaheuristic. Dorigo M, Di Caro G, Sampels M, eds., *Ant Algorithms*, volume 2463 of *Lecture Notes in Computer Science*, 176–187 (Berlin, Heidelberg: Springer Berlin Heidelberg).

Birattari M, Balaprakash P, Stützle T, Dorigo M (2008) Estimation-based local search for stochastic combinatorial optimization using delta evaluations: A case study on the probabilistic traveling salesman problem. *INFORMS Journal on Computing* 20(4):644–658.

Birge JR, Louveaux F (2011) *Introduction to Stochastic Programming* (New York, NY: Springer).

Blauth J, Neuwohner M, Puhlmann L, Vygen J (2025) Improved guarantees for the a priori TSP. *Mathematics of Operations Research* 50(4):2909–2940.

Blufstein M, Lera-Romero G, Soulignac FJ (2024) Decremental state-space relaxations for the basic traveling salesman problem with a drone. *INFORMS Journal on Computing* 36(4):1064–1083.

Campbell AM (2006) Aggregation for the probabilistic traveling salesman problem. *Computers & Operations Research* 33(9):2703–2724.

Campbell AM, Thomas BW (2008) Probabilistic traveling salesman problem with deadlines. *Transportation Science* 42(1):1–21.

Charitopoulos VM, Dua V, Papageorgiou LG (2017) Traveling salesman problem-based integration of planning, scheduling, and optimal control for continuous processes. *Industrial & Engineering Chemistry Research* 56(39):11186–11205.

Florio AM, Feillet D, Hartl RF (2018) The delivery problem: Optimizing hit rates in e-commerce deliveries. *Transportation Research Part B: Methodological* 117:455–472.

Gendreau M, Jabali O, Rei W (2014) Chapter 8: Stochastic vehicle routing problems. Toth P, Vigo D, eds., *Vehicle Routing: Problems, Methods, and Applications*, 213–239, MOS-SIAM Series on Optimization (Philadelphia, PA: Society for Industrial and Applied Mathematics), second edition.

Gendreau M, Jabali O, Rei W (2016) 50th anniversary invited article—future research directions in stochastic vehicle routing. *Transportation Science* 50(4):1163–1173.

Gendreau M, Laporte G, Séguin R (1996) Stochastic vehicle routing. *European Journal of Operational Research* 88(1):3–12.

Jaillet P (1985) *Probabilistic Traveling Salesman Problems*. Ph.D. thesis, Massachusetts Institute of Technology.

Jaillet P (1988) A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Operations Research* 36(6):929–936.

Kendall MG (1938) A new measure of rank correlation. *Biometrika* 30(1-2):81–93.

Kovacs AA, Golden BL, Hartl RF, Parragh SN (2014) Vehicle routing problems in which consistency considerations are important: A survey. *Networks* 64(3):192–213.

Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* 13(3):133–142.

Laporte G, Louveaux FV, Mercure H (1994) A priori optimization of the probabilistic traveling salesman problem. *Operations Research* 42(3):543–549.

Liu YH (2007) A hybrid scatter search for the probabilistic traveling salesman problem. *Computers & Operations Research* 34(10):2949–2963.

Liu YH (2008) Diversified local search strategy under scatter search framework for the probabilistic traveling salesman problem. *European Journal of Operational Research* 191(2):332–346.

Löffler M, Boysen N, Schneider M (2022) Picker routing in AGV-assisted order picking systems. *INFORMS Journal on Computing* 34(1):440–462.

Natarajan K, Song M, Teo CP (2009) Persistency model and its applications in choice modeling. *Management Science* 55(3):453–469.

Nelsen RB (2006) *An Introduction to Copulas* (New York: Springer), second edition.

Oyola J, Arntzen H, Woodruff DL (2017) The stochastic vehicle routing problem, a literature review, Part II: Solution methods. *EURO Journal on Transportation and Logistics* 6(4):349–388.

Oyola J, Arntzen H, Woodruff DL (2018) The stochastic vehicle routing problem, a literature review, Part I: Models. *EURO Journal on Transportation and Logistics* 7(3):193–221.

Özarık SS, Veelenturf LP, Van Woensel T, Laporte G (2021) Optimizing e-commerce last-mile vehicle routing and scheduling under uncertain customer presence. *Transportation Research Part E: Logistics and Transportation Review* 148:102263.

Padberg M, Rinaldi G (1990) An efficient algorithm for the minimum capacity cut problem. *Mathematical Programming* 47(1-3):19–36.

Padberg M, Sung TY (1991) An analytical comparison of different formulations of the travelling salesman problem. *Mathematical Programming* 52(1-3):315–357.

Panagiotelis A, Czado C, Joe H (2012) Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association* 107(499):1063–1072.

Shapiro A, Dentcheva D, Ruszczynski A (2021) *Lectures on Stochastic Programming: Modeling and Theory* (Society for Industrial and Applied Mathematics), third edition.

Shen S, Smith JC, Ahmed S (2010) Expectation and chance-constrained models and algorithms for insuring critical paths. *Management Science* 56(10):1794–1814.

Shmoys D, Talwar K (2008) A constant approximation algorithm for the a priori traveling salesman problem. Lodi A, Panconesi A, Rinaldi G, eds., *Integer Programming and Combinatorial Optimization*, volume 5035, 331–343 (Berlin, Heidelberg: Springer Berlin Heidelberg).

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8:229–231.

Song Y, Ulmer MW, Thomas BW, Wallace SW (2020) Building trust in home services—stochastic team-orienteering with consistency constraints. *Transportation Science* 54(3):823–838.

Stoer M, Wagner F (1997) A simple min-cut algorithm. *Journal of the ACM* 44(4):585–591.

Stroh AM, Erera AL, Toriello A (2022) Tactical design of same-day delivery systems. *Management Science* 68(5):3444–3463.

Tang H, Miller-Hooks E (2007) Solving a generalized traveling salesperson problem with stochastic customers. *Computers & Operations Research* 34(7):1963–1987.

Voccia SA, Campbell AM, Thomas BW (2013) The probabilistic traveling salesman problem with time windows. *EURO Journal on Transportation and Logistics* 2(1):89–107.

Weiler C, Biesinger B, Hu B, Raidl GR (2015) Heuristic approaches for the probabilistic traveling salesman problem. Moreno-Díaz R, Pichler F, Quesada-Arencibia A, eds., *Computer Aided Systems Theory*, 342–349 (Cham: Springer International Publishing).

Weyland D, Montemanni R, Gambardella LM (2012) Hardness results for the probabilistic traveling salesman problem with deadlines. Mahjoub AR, Markakis V, Milis I, Paschos VT, eds., *Combinatorial Optimization*, 392–403 (Berlin, Heidelberg: Springer Berlin Heidelberg).

Wissink P (2019) *The Probabilistic Travelling Salesman Off the Beaten Track: The Impact of Clustering, Construction, and Correlation*. Ph.D. thesis, University of Edinburgh.

Wissink P (2023) The traveling salesman problem with stochastic and correlated customers. *Transportation Science* 57(5):1321–1339.

Online Appendix for the Paper
"An exact algorithm for the probabilistic TSP via a new tractable convex representation of recourse"

A Summary of notation

Table A.1 summarizes the main notation used throughout the paper.

Table A.1: Summary of notation.

Symbol	Description
Notation for the underlying graph and tours	
$G = (V, E)$	Complete undirected graph with vertex/customer set V and edge set E.
n	Number of customers, i.e., $n = \|V\|$.
c_{ij}	Non-negative symmetric traveling cost associated with edge $\{i, j\} \in E$, i.e., $c_{ij} = c_{ji} \geq 0$.
Π	Set of all permutations of V.
$\sigma = (\sigma(1), \ldots, \sigma(n))$	A permutation of V.
$\sigma(k)$	Customer visited at position k in tour σ.
$\pi_i = \sigma^{-1}(i)$	Position (visiting order) of customer i in tour σ.
x	Decision variables defining a Hamiltonian tour, where $x_{ij} = 1$ if edge $\{i, j\}$ is selected in the a priori tour and $x_{ij} = 0$ otherwise.
$\mathcal{T}$	Set of feasible vectors x corresponding to Hamiltonian tours.
$\delta(i)$	Set of edges incident to customer i.
$E(S)$	Set of edges with both endpoints in S.
σ^x	Permutation of the vertices in V corresponding to an undirected Hamiltonian cycle induced by x.
S^x_{ij}, T^x_{ij}	Sets of intermediate customers on the path from i to j and the path from j to i along tour σ^x, excluding i and j.
$\mathcal{P}_{ij}$	Set of 2-partitions $\{S, T\}$ of $V \setminus \{i, j\}$.
Notation for random realizations and edge traversals	
$\xi \in \{0, 1\}^n$	Bernoulli random vector indicating which customers are present.
$\mathbb{P}(\xi)$	Probability of realization ξ.
$\mathbb{P}(\Xi)$	Probability of an event $\Xi \subseteq \{0, 1\}^n$, defined as $\mathbb{P}(\Xi) = \sum_{\xi \in \Xi} \mathbb{P}(\xi)$.
Γ^I_{ij}	Set of realizations where customers i and j are present and all customers in I are absent.
Ξ^x_{ij} (or Ξ^x_{ji})	Set of realizations where customers i and j are present and all customers in S^x_{ij} (or T^x_{ij}) are absent, i.e., $\Xi^x_{ij} = \Gamma^{S^x_{ij}}_{ij}$ and $\Xi^x_{ji} = \Gamma^{T^x_{ij}}_{ij}$.
$\Psi^\xi_{ij}(x)$	Traversal count of edge $\{i, j\}$ in the a posteriori tour defined by x under realization ξ.
$\Phi^\xi(\sigma)$	Traveling cost of the a posteriori tour induced by the a priori tour σ under realization ξ.
$\Phi^\xi(x)$ (or $\Phi^\xi(\sigma^x)$)	Traveling cost of the a posteriori tour defined by x under realization ξ.
$\mathbb{E}_\xi[\Psi^\xi_{ij}(x)]$	ETC of edge $\{i, j\}$ in the a posteriori tour induced by x.

Table A.1: Summary of notation (continued)

Symbol	Description
$\mathbb{E}_\xi[\Phi^\xi(x)]$ (or $\phi(x)$)	Expected traveling cost of the a posteriori tour induced by x.

B Computation of $\mathbb{P}(\Gamma_{ij}^I)$ in the correlated PTSP of Wissink (2023)

In this section, we provide the details of the computation of $\mathbb{P}(\Gamma_{ij}^I)$ in Archimedean or D-vine elliptical copula-based models (Wissink 2023). Note that these results were presented in Wissink (2023) and we include them for completeness.

To proceed, for $r \in V$, we denote by F_r the marginal cumulative distribution function (CDF) of the Bernoulli random variable ξ_r, i.e., $F_r(t) = \mathbb{P}(\xi_r \le t)$. Note that as ξ_r is a Bernoulli random variable, it follows $F_r(0) = \mathbb{P}(\xi_r = 0) = 1 - p_r$ (where $p_r := \mathbb{P}(\xi_r = 1)$), and $F_r(1) = \mathbb{P}(\xi_r \le 1) = 1$. For $i, j \in V$ and $I \subseteq V \backslash \{i, j\}$, to compute $\mathbb{P}(\Gamma_{ij}^I)$, we first rewrite it as:

$$
\begin{aligned}
\mathbb{P}(\Gamma_{ij}^I) &= \mathbb{P}(\xi_i = \xi_j = 1,\ \xi_k = 0,\ \forall\, k \in I) \\
&\overset{\text{(a)}}{=} \sum_{v_i \in \{0,1\}} \sum_{v_j \in \{0,1\}} \sum_{\{v_k\}_{k\in I} \in \{0,1\}^{|I|}} (-1)^{v_i + v_j + \sum_{k\in I} v_k}\, \mathbb{P}(\xi_i \le 1 - v_i,\ \xi_j \le 1 - v_j,\ \xi_k \le -v_k,\ k \in I) \\
&\overset{\text{(b)}}{=} \mathbb{P}(\xi_i \le 1,\ \xi_j \le 1,\ \xi_k \le 0,\ \forall\, k \in I) - \mathbb{P}(\xi_i \le 0,\ \xi_j \le 1,\ \xi_k \le 0,\ \forall\, k \in I) \\
&\quad - \mathbb{P}(\xi_i \le 1,\ \xi_j \le 0,\ \xi_k \le 0,\ \forall\, k \in I) + \mathbb{P}(\xi_i \le 0,\ \xi_j \le 0,\ \xi_k \le 0,\ \forall\, k \in I),
\end{aligned}
$$

where (a) comes from the finite-difference arguments (see, e.g., Panagiotelis et al. (2012)), and (b) holds because $\mathbb{P}(\xi_i \le 1 - v_i,\ \xi_j \le 1 - v_j,\ \xi_k \le -v_k,\ k \in I) = 0$ if $v_k = 1$ holds for some $k \in I$. From Sklar's theorem (Sklar 1959), any joint distribution function can be expressed as the composition of marginal CDFs and a copula function, yielding

$$
\begin{aligned}
\mathbb{P}(\Gamma_{ij}^I) &= C(F_i(1), F_j(1), \{F_k(0)\}_{k\in I}) - C(F_i(0), F_j(1), \{F_k(0)\}_{k\in I}) \\
&\quad - C(F_i(1), F_j(0), \{F_k(0)\}_{k\in I}) + C(F_i(0), F_j(0), \{F_k(0)\}_{k\in I}) \\
&\overset{\text{(a)}}{=} C(1, 1, \{1 - p_k\}_{k\in I}) - C(1 - p_i, 1, \{1 - p_k\}_{k\in I}) \\
&\quad - C(1, 1 - p_j, \{1 - p_k\}_{k\in I}) + C(1 - p_i, 1 - p_j, \{1 - p_k\}_{k\in I}),
\end{aligned} \tag{B.1}
$$

where $C : [0,1]^d \to [0,1]$ with $d = |I| + 2$, is the copula (function) linking the marginals to the joint distribution. When C is an Archimedean copula, it admits the closed-form representation

$$
C(u_1, \dots, u_d) = \psi^{-1}\left(\sum_{i=1}^{d} \psi(u_i)\right), \tag{B.2}
$$

which is generated by a one-dimensional function $\psi : [0,1] \to [0, +\infty)$. In this case, evaluations of $\mathbb{P}(\Gamma_{ij}^I)$ can be performed in $O(n)$ time. Typical examples of Archimedean copulas include Clayton and Gumbel copulas:

$$\psi^{\text{Clayton}}(t) = \frac{1}{\theta}(t^{-\theta} - 1) \text{ and } C^{\text{Clayton}}(u_1, \dots, u_d) = \left[1 + \sum_{i=1}^{d}\left(u_i^{-\theta} - 1\right)\right]^{-1/\theta}, \quad \theta > 0, \tag{B.3}$$

$$\psi^{\text{Gumbel}}(t) = (-\ln(t))^{\theta} \text{ and } C^{\text{Gumbel}}(u_1, \dots, u_d) = \exp\left[-\left(\sum_{i=1}^{d}(-\ln u_i)^{\theta}\right)^{1/\theta}\right], \quad \theta \geq 1. \tag{B.4}$$

When C is an elliptical copula, it does not admit a closed-form expression and evaluating it requires high-dimensional integrals. A typical example of this is the Gaussian copula:

$$\begin{aligned} C^{\text{Gauss}}(u_1, \dots, u_d) &= \Phi_{\Sigma}\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2} z^{\top}\Sigma^{-1} z\right) dz, \end{aligned} \tag{B.5}$$

where $z = (z_1, \dots, z_d)^{\top}$, Φ is the standard normal CDF, and Φ_{Σ} is the joint CDF of a multivariate normal distribution with correlation matrix Σ, and $\det(\Sigma)$ is the determinant of Σ.

Directly computing (B.5) requires evaluating high-dimensional Gaussian rectangle probabilities, which could be computationally demanding. To tackle this, the discrete-vine (D-vine) pair Gaussian copula can be used (Panagiotelis et al. 2012). It computes joint probabilities of the form $\mathbb{P}(\xi_S = v_S)$ (where $S \subseteq V$, ξ_S is the subvector of ξ restricted to S, and $v_S \in \{0, 1\}^{|S|}$) using bivariate Gaussian copulas and can be used as an approximation to the Gaussian copula. In particular, letting $S = \{i_1, \dots, i_s\}$ and applying Bayes' theorem, we obtain

$$\mathbb{P}(\xi_S = v_S) = \left(\prod_{\tau=1}^{s-1} \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{i_{\tau+1}} = v_{i_{\tau+1}}, \dots, \xi_{i_s} = v_{i_s})\right) \cdot \mathbb{P}(\xi_{i_s} = v_{i_s}). \tag{B.6}$$

Define

$$D_{\tau,\ell} := \begin{cases} \{i_\tau, \dots, i_\ell\} & \text{if } \ell \geq \tau; \\ \varnothing & \text{if } \ell \leq \tau - 1, \end{cases} \quad \forall\, \tau, \ell \in [s].$$

To compute (B.6), for any $1 \leq \tau < \ell \leq s$, we evaluate the following two families of conditional probabilities

$$\begin{aligned} \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell}} = v_{D_{\tau+1,\ell}}) &= \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}},\ \xi_{i_\ell} = v_{i_\ell}) \\ &= \frac{\mathbb{P}(\xi_{i_\tau} = v_{i_\tau},\ \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})}{\mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})}, \end{aligned} \tag{B.7}$$

$$\begin{aligned} \mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau,\ell-1}} = v_{D_{\tau,\ell-1}}) &= \mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{i_\tau} = v_{i_\tau},\ \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}) \\ &= \frac{\mathbb{P}(\xi_{i_\tau} = v_{i_\tau},\ \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})}{\mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})}. \end{aligned} \tag{B.8}$$

Similarly, Sklar's theorem implies that there exists a bivariate copula, denoted by $C_{i_\tau, i_\ell | D_{\tau+1,\ell-1}}(\cdot, \cdot \mid v_{D_{\tau+1,\ell-1}})$, such that

$$\begin{aligned} &\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}) \\ &\quad = \sum_{\delta_\tau \in \{0,1\}} \sum_{\delta_\ell \in \{0,1\}} (-1)^{\delta_\tau + \delta_\ell}\, \mathbb{P}\left(\xi_{i_\tau} \leq v_{i_\tau} - \delta_\tau,\ \xi_{i_\ell} \leq v_{i_\ell} - \delta_\ell \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}\right) \\ &\quad = \sum_{\delta_\tau \in \{0,1\}} \sum_{\delta_\ell \in \{0,1\}} (-1)^{\delta_\tau + \delta_\ell} C_{i_\tau, i_\ell | D_{\tau+1,\ell-1}}\left(F_{i_\tau | D_{\tau+1,\ell-1}}(v_{i_\tau} - \delta_\tau \mid v_{D_{\tau+1,\ell-1}}), F_{i_\ell | D_{\tau+1,\ell-1}}(v_{i_\ell} - \delta_\ell \mid v_{D_{\tau+1,\ell-1}}) \mid v_{D_{\tau+1,\ell-1}}\right), \end{aligned} \tag{B.9}$$

where $F_{i_\tau|D_{\tau+1,\ell-1}}(v_{i_\tau} - \delta_\tau \mid v_{D_{\tau+1,\ell-1}}) = \mathbb{P}(\xi_{i_\tau} \le v_{i_\tau} - \delta_\tau \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})$ and $F_{i_\ell|D_{\tau+1,\ell-1}}(v_{i_\ell} - \delta_\ell \mid v_{D_{\tau+1,\ell-1}}) = \mathbb{P}(\xi_{i_\ell} \le v_{i_\ell} - \delta_\ell \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})$. Under mild conditions, the conditional copula can be taken to be invariant with respect to the realization of the conditioning variables (Panagiotelis et al. 2012), that is,

$$C_{i_\tau,i_\ell|D_{\tau+1,\ell-1}}(u, v \mid v_{D_{\tau+1,\ell-1}}) = C_{i_\tau,i_\ell|D_{\tau+1,\ell-1}}(u, v), \quad \forall\, (u, v) \in [0, 1]^2,\ \forall\, v_{D_{\tau+1,\ell-1}} \in \{0, 1\}^{\ell-\tau-1}.$$

Therefore,

$$\begin{aligned} &\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}) \\ &= \sum_{\delta_\tau\in\{0,1\}} \sum_{\delta_\ell\in\{0,1\}} (-1)^{\delta_\tau+\delta_\ell} C_{i_\tau,i_\ell|D_{\tau+1,\ell-1}} \left(F_{i_\tau|D_{\tau+1,\ell-1}}(v_{i_\tau} - \delta_\tau \mid v_{D_{\tau+1,\ell-1}}), F_{i_\ell|D_{\tau+1,\ell-1}}(v_{i_\ell} - \delta_\ell \mid v_{D_{\tau+1,\ell-1}})\right). \end{aligned} \tag{B.10}$$

Since ξ_{i_τ} (or ξ_{i_ℓ}) is a Bernoulli random variable, it follows that

$$F_{i_\tau|D_{\tau+1,\ell-1}}(x \mid v_{D_{\tau+1,\ell-1}}) = \begin{cases} 0, & \text{if } x < 0; \\ \mathbb{P}(\xi_{i_\tau} = 0 \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}), & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1, \end{cases} \tag{B.11}$$

$$F_{i_\ell|D_{\tau+1,\ell-1}}(x \mid v_{D_{\tau+1,\ell-1}}) = \begin{cases} 0, & \text{if } x < 0; \\ \mathbb{P}(\xi_{i_\ell} = 0 \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}), & \text{if } 0 \le x < 1; \\ 1, & \text{if } x \ge 1. \end{cases} \tag{B.12}$$

In the D-vine pair Gaussian copula, the copula functions $C_{i_\tau,i_\ell|D_{\tau+1,\ell-1}}(\cdot,\cdot)$ in (B.10) are all bivariate Gaussian copulas (with possibly different 2×2 correlation matrices). To compute the conditional probabilities in (B.7) and (B.8), we first evaluate the pairwise probabilities $\{\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell}) : 1 \le \tau < \ell \le s\}$, and then recursively compute $\mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell}})$ and $\mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau,\ell-1}})$ in (B.7) and (B.8) for increasing values of $\ell - \tau$, starting from $\ell = \tau + 1$ and continuing until $\ell = s$. With these conditional probabilities, we can then evaluate $\mathbb{P}(\xi_S = v_S)$ as in (B.6). The overall computational complexity is $O(s^2) \le O(n^2)$ (Wissink 2023).

Wissink (2023) also considered a truncated version of the D-vine Gaussian copula, in which dependence is truncated beyond a prescribed threshold. In the truncated version with parameter L, if $|D_{\tau+1,\ell-1}| \ge L$, variables ξ_{i_τ} and ξ_{i_ℓ} are assumed to be conditionally independent:

$$\begin{aligned} &\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}) \\ &\qquad = \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})\, \mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}). \end{aligned} \tag{B.13}$$

Substituting (B.13) into (B.7) and (B.8) yields

$$\mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}, \xi_{i_\ell} = v_{i_\ell}) = \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}), \tag{B.14}$$

$$\mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{i_\tau} = v_{i_\tau},\ \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}) = \mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}}). \tag{B.15}$$

Therefore, (B.6) reduces to

$$\mathbb{P}(\xi_S = v_S) = \left(\prod_{\tau=1}^{s-1} \mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{i_{\tau+1}} = v_{i_{\tau+1}}, \dots, \xi_{i_{\min\{s,\tau+L\}}} = v_{i_{\min\{s,\tau+L\}}}) \right) \cdot \mathbb{P}(\xi_{i_s} = v_{i_s}), \tag{B.16}$$

where the last retained index is $i_{\min\{s,\tau+L\}}$ because the newly added variable ξ_{i_ℓ} is truncated only when $|D_{\tau+1,\ell-1}| = \ell - \tau - 1 \ge L$. To evaluate $\mathbb{P}(\xi_S = v_S)$, it suffices to first compute the pairwise probabilities $\{\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell}) \mid 1 \le \tau < \ell \le s,\ \ell \le \tau + L\}$ and then compute $\mathbb{P}(\xi_{i_\tau} = v_{i_\tau} \mid \xi_{D_{\tau+1,\ell}})$ and $\mathbb{P}(\xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau,\ell-1}})$ recursively for increasing values of $\ell - \tau$, starting from $\ell = \tau + 1$ and continuing until $\ell = \min\{s, \tau + L\}$. The overall computational complexity is $O(Ls) \le O(Ln)$. For more details, we refer to Wissink (2023).

C A counterexample to the validity of inequality (30)

In this section, we provide a counterexample demonstrating that inequality (30) is not valid for formulation (MILP) (or equivalently, for formulation (E-MILP)). Consider a PTSP instance with $V = \{1, 2, 3\}$ and $c_{ij} = 1$ for all $\{i, j\} \in E$. Suppose that the customer presences are independent, with $\mathbb{P}(\xi_1 = 1) = 1$ and $\mathbb{P}(\xi_2 = 1) = \mathbb{P}(\xi_3 = 1) = 1/2$. Then, from (31), it follows that $\hat{b}_{12} = \hat{b}_{13} = 1/2 + 1/4 = 3/4$ and $\hat{b}_{23} = 1/4 + 1/4 = 1/2$. Let x denote the a priori tour defined by $x_{12} = x_{13} = x_{23} = 1$. For this tour, (8) yields

$$\mathbb{E}_{\xi}[\Psi^{\xi}_{12}(x)] = \mathbb{E}_{\xi}[\Psi^{\xi}_{13}(x)] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}, \qquad \mathbb{E}_{\xi}[\Psi^{\xi}_{23}(x)] = \frac{1}{4}.$$

Define $\theta_{ij} := \mathbb{E}_{\xi}[\Psi^{\xi}_{ij}(x)]$ for all $\{i, j\} \in E$. Then, (17) implies that (x, θ) is feasible for formulation (MILP). However,

$$\sum_{\{i,j\}\in E} c_{ij}\theta_{ij} = \frac{7}{4} < 2 = \sum_{\{i,j\}\in E} \hat{b}_{ij}x_{ij}.$$

Thus, (x, θ) violates inequality (30), proving that the inequality is not valid for formulation (MILP). By contrast, (29) gives $b_{12} = b_{13} = 5/8$ and $b_{23} = 1/2$, and hence

$$\sum_{\{i,j\}\in E} c_{ij}\theta_{ij} = \frac{7}{4} = \sum_{\{i,j\}\in E} b_{ij}x_{ij}.$$

Therefore, inequality (28) holds with equality at (x, θ).

D Details for the distributions of ξ in the constructed PTSP instances

Following Wissink (2023), we use Kendall's τ (Kendall 1938) as the dependence measure for the Clayton, Gumbel, and Gaussian copulas, where dependence increases monotonically with Kendall's τ, although the implied dependence structure differs across copula families (Nelsen 2006). For the Clayton copula, $\tau = \frac{\theta}{\theta+2}$, so $\tau = 0$ corresponds to independence, and larger values of τ imply stronger lower-tail dependence. For the Gumbel copula, $\tau = 1 - \frac{1}{\theta}$, so $\tau = 0$ corresponds to independence, and larger values of τ imply stronger upper-tail dependence. In our computational experiments, we refer to $\tau = 0.3$ as low dependence and $\tau = 0.7$ as high dependence for the Clayton and Gumbel copulas. For the truncated version of the D-vine Gaussian copula, we follow Wissink (2023) to set $L = 4$. Thus, if $|D_{\tau+1,\ell-1}| \geq 4$, the variables ξ_{i_τ} and ξ_{i_ℓ} associated with $\mathbb{P}(\xi_{i_\tau} = v_{i_\tau}, \xi_{i_\ell} = v_{i_\ell} \mid \xi_{D_{\tau+1,\ell-1}} = v_{D_{\tau+1,\ell-1}})$ are taken to be conditionally independent. For the remaining cases with $|D_{\tau+1,\ell-1}| \leq 3$, we let all pair-copulas with the same conditioning-set size share the same bivariate Gaussian copula. Specifically, for each $D_{\tau+1,\ell-1}$ with $|D_{\tau+1,\ell-1}| = k$, we set

$$C_{i_\tau,i_\ell|D_{\tau+1,\ell-1}}(u, v) := C_k(u, v) = \Phi_{\Sigma_k}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right),$$

where

$$\Sigma_k = \begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix}, \quad \forall\, k = 0, 1, 2, 3.$$

Here, ρ_k is determined from Kendall's tau through $\tau_k = \frac{2}{\pi}\arcsin(\rho_k)$, $k = 0, 1, 2, 3$. In our experiments, $(\tau_0, \tau_1, \tau_2, \tau_3)$ is set to (0.3, 0.2, 0.1, 0.05) and (0.7, 0.4, 0.3, 0.2) for the low- and high-dependence settings, respectively.

Table E.1: Performance of MILP and E-MILP on PTSP instances in testset T1 with $p_i \in [0.95, 1.0]$.

	#inst	MILP					E-MILP				
		#S	T[s]	#N	G[%]	rG[%]	#S	T[s]	#N	G[%]	rG[%]
burma14	7	7	<0.1	9	0.0	0.3	7	<0.1	2	0.0	<0.1
ulysses16	7	7	0.1	56	0.0	1.1	7	0.1	17	0.0	0.8
ulysses22	7	7	0.3	114	0.0	0.9	7	0.2	12	0.0	0.4
fri26	7	7	0.3	25	0.0	0.5	7	0.2	6	0.0	0.1
bayg29	7	7	0.4	138	0.0	0.9	7	0.3	19	0.0	0.4
swiss42	7	7	3.0	42	0.0	0.4	7	0.4	6	0.0	<0.1
att48	7	7	5.0	978	0.0	1.5	7	1.1	68	0.0	0.9
eil51	7	7	6.3	1515	0.0	1.1	7	1.6	145	0.0	0.5
berlin52	7	7	1.0	53	0.0	0.2	7	0.4	2	0.0	<0.1
st70	7	7	59.4	18939	0.0	1.7	7	5.4	616	0.0	1.3
eil76	7	7	14.0	2500	0.0	0.9	7	3.1	125	0.0	0.4
pr76	7	2	5818.1	923895	0.3	3.0	7	708.0	97913	0.0	2.6
gr96	7	7	175.8	27768	0.0	2.1	7	22.3	1852	0.0	1.8
rat99	7	7	22.6	2300	0.0	1.2	7	6.3	148	0.0	0.7
kroA100	7	7	110.3	33348	0.0	1.6	7	15.6	1661	0.0	1.1
kroB100	7	7	372.1	103109	0.0	2.0	7	36.4	3011	0.0	1.6
kroC100	7	7	53.6	9054	0.0	1.5	7	7.7	240	0.0	1.0
kroD100	7	7	55.7	11153	0.0	1.5	7	11.1	796	0.0	1.0
kroE100	7	7	1066.5	265293	0.0	2.2	7	57.3	7649	0.0	1.6
rd100	7	7	23.3	2185	0.0	1.1	7	6.4	104	0.0	0.5
Independent	20	20	221.8	92137	0.0	1.5	20	8.4	3561	0.0	0.9
Clayton-high	20	20	112.6	41992	0.0	0.9	20	13.1	2799	0.0	0.7
Clayton-low	20	19	483.3	42713	<0.1	1.2	20	93.2	5178	0.0	0.9
Gumbel-high	20	19	609.6	76138	<0.1	1.2	20	100.7	8022	0.0	0.9
Gumbel-low	20	19	449.8	83198	<0.1	1.5	20	18.8	7032	0.0	0.8
Gaussian-high	20	19	443.5	85938	<0.1	1.4	20	58.3	10024	0.0	0.8
Gaussian-low	20	19	405.1	68751	<0.1	1.5	20	16.8	3421	0.0	0.8
Ave.	140	135	389.4	70124	<0.1	1.3	140	44.2	5720	0.0	0.8

E Detailed computational results on instances in testset T1 under MILP and E-MILP

Tables E.1 and E.2 present the detailed computational results on instances in testset T1 under MILP and E-MILP.

Table E.2: Performance of `MILP` and `E-MILP` on PTSP instances in testset T1 with $p_i \in [0.90, 1.0]$.

	# inst	MILP					E-MILP				
		# S	T [s]	# N	G [%]	rG [%]	# S	T [s]	# N	G [%]	rG [%]
burma14	7	7	0.1	54	0.0	1.8	7	2.5	10	0.0	0.7
ulysses16	7	7	0.5	432	0.0	2.9	7	0.2	75	0.0	2.0
ulysses22	7	7	2.6	1818	0.0	2.8	7	0.4	121	0.0	1.7
fri26	7	7	0.7	324	0.0	1.9	7	0.3	30	0.0	0.9
bayg29	7	7	4.3	2622	0.0	2.3	7	0.8	228	0.0	1.8
swiss42	7	7	3.4	1700	0.0	2.2	7	0.8	50	0.0	0.9
att48	7	7	47.4	14049	0.0	3.2	7	4.2	577	0.0	2.4
eil51	7	7	136.6	28306	0.0	2.4	7	23.0	2482	0.0	1.8
berlin52	7	7	5.1	830	0.0	1.3	7	1.0	43	0.0	0.5
st70	7	1	6191.2	671386	0.6	3.6	7	147.1	13395	0.0	2.8
eil76	7	7	1241.6	98579	0.0	2.5	7	86.3	3392	0.0	1.6
pr76	7	0	7200.0	818898	2.5	5.1	0	7200.0	582756	0.9	4.3
gr96	7	1	6302.7	664679	0.8	4.0	7	388.6	21853	0.0	3.2
rat99	7	7	1272.7	258012	0.0	2.7	7	64.2	4176	0.0	1.8
kroA100	7	1	6206.9	748680	1.2	3.6	7	309.8	26599	0.0	2.7
kroB100	7	1	6571.1	754251	1.4	4.1	7	1203.6	79181	0.0	3.3
kroC100	7	1	6179.8	744148	0.5	3.1	7	62.0	3880	0.0	2.0
kroD100	7	1	6213.1	738589	0.9	3.5	7	271.6	20555	0.0	2.5
kroE100	7	1	6741.7	779492	1.8	4.3	6	1833.9	121758	<0.1	3.2
rd100	7	5	4885.7	631246	0.1	3.3	7	108.8	6783	0.0	1.9
Independent	20	12	3208.2	472763	0.6	3.5	19	438.4	71640	<0.1	2.3
Clayton-high	20	19	814.6	135382	<0.1	1.9	19	390.5	35055	<0.1	1.5
Clayton-low	20	12	3347.9	170728	0.6	2.8	18	1132.0	17950	0.1	2.0
Gumbel-high	20	12	3523.0	219534	0.5	2.8	19	758.1	14157	<0.1	2.0
Gumbel-low	20	12	3114.2	476308	0.5	3.4	19	482.5	72231	<0.1	2.2
Gaussian-high	20	11	3413.1	483694	0.5	3.3	19	454.2	30415	<0.1	2.3
Gaussian-low	20	11	3301.5	476923	0.6	3.5	19	442.5	69334	<0.1	2.3
Ave.	140	89	2960.4	347905	0.5	3.0	132	585.5	44397	<0.1	2.1