ZIB

ZUSE INSTITUTE BERLIN

Zuse Institute Berlin

Berenike Masing[1] Niels Lindner[2],
Ralf Borndörfer[3]

The Price of Symmetric Line Plans in the Parametric City

[1] 0000-0001-7201-2412
[2] 0000-0002-8337-4387
[3] 0000-0001-7223-9174

The Price of Symmetric Line Plans in the Parametric City

Berenike Masing[*†]   Niels Lindner[*]   Ralf Borndörfer[*]

`masing@zib.de`   `lindner@zib.de`   `borndoerfer@zib.de`

May 10, 2022

Abstract

We consider the line planning problem in public transport in the Parametric City, an idealized model that captures typical scenarios by a (small) number of parameters. The Parametric City is rotation symmetric, but optimal line plans are not always symmetric. This raises the question to quantify the symmetry gap between the best symmetric and the overall best solution. For our analysis, we formulate the line planning problem as a mixed integer linear program, that can be solved in polynomial time if the solutions are forced to be symmetric. The symmetry gap is provably small when a specific Parametric City parameter is fixed, and we give an approximation algorithm for line planning in the Parametric City in this case. While the symmetry gap can be arbitrarily large in general, we show that symmetric line plans are a good choice in most practical situations.

1   Introduction

The overall goal of public transport planning is to satisfy the demand providing a user-friendly service, while limiting the operator's cost, see Ceder and Wilson (1986), Assad (1980), and Bussieck et al. (1997). An overview tailored to line planning in particular is given by Schöbel (2011). Karbstein (2013) as well as Schmidt (2014) consider passenger-oriented models, which focus on reducing transfers. The basic transit network design problem, which can include line planning, is surveyed by Kepaptsoglou and Karlaftis (2009) and Lopez (2014). Most recent approaches focus on representing a specific city as realistically as possible by tailor made "irregular networks". This produces the best possible results for the moment of consideration, but it makes generalization, benchmarking, and extrapolation hard, as network structures do not easily carry over to other cities, and not even to the same city twenty years in the future. An alternative approach is to consider *generic city structures*, that reflect the most prevalent economical and spatial aspects, but remain simple enough to be analyzed and understood. Well organized prototypcial (or even standardized) transportation networks could then be used in similar cities. Two examples are Manhattan type grids that were considered by Holroyd (1967), and ring-radial models as studied by Byrne (1975) and more recently by Badia (2020). There have also been hybrid approaches, e.g., Daganzo (2010) combines a grid-like model at the center with a hub-and-spoke structure surrounding it. A versatile model, and the one that we consider in this paper, is the so-called Parametric City by Fielbaum et al. (2016b). It provides an idealized representation of a city that can be adapted to different situations by a flexible choice of parameters.
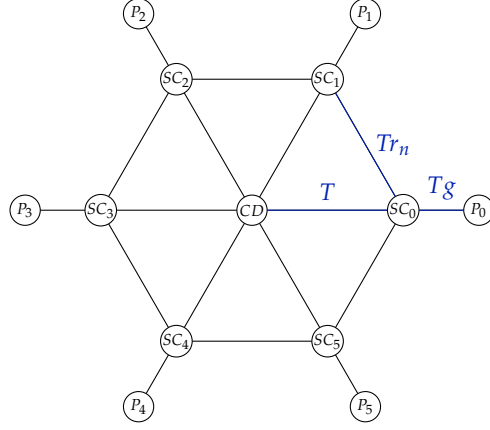

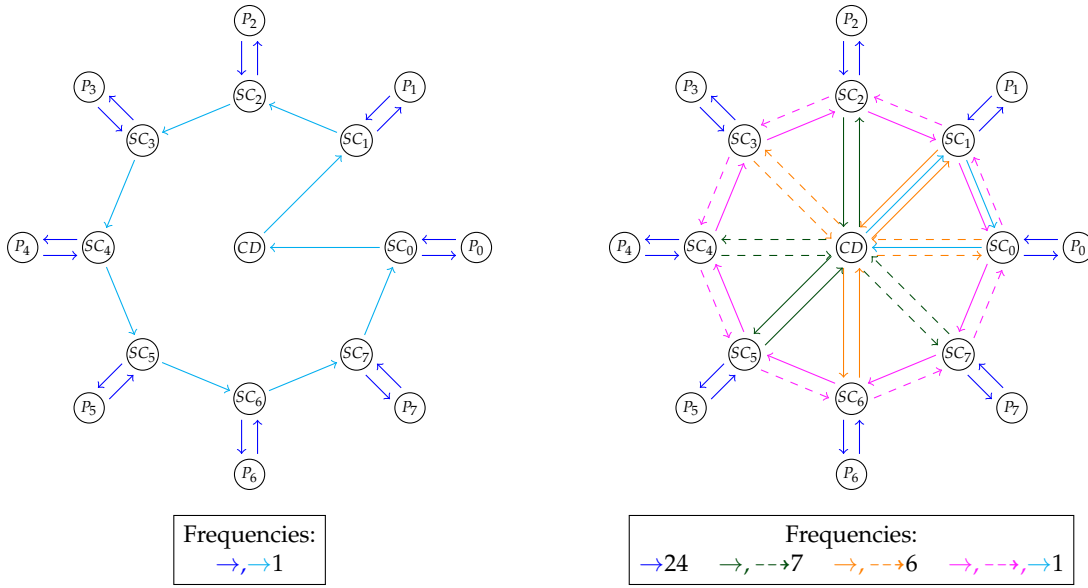Figure 1: The infrastructure graph of the Parametric City for $n = 6$

Figure 2: Examplary Asymmetric Optimal Line Plans

The main degree of freedom in the Parametric City is the design of the line system, i.e., the selection of the routes and their frequencies of operation, and one usually aims at combining the minimization of operator costs and the maximization of user comfort (e.g., short travel times, few changeovers, low waiting times). As the Parametric City is rotation symmetric in both its geometry and demand pattern, see an exemplary graph in Figure 1, one would expect that an optimal public transport system is equally symmetric, and, in particular, that the line system is symmetric. This is desirable because a symmetric system has a simple structure that makes it easy to operate and memorize. While, as it will turn out, this is true in most cases, there are instances in which it is possible to find asymmetric line plans that have better objective values than any symmetric line plan. Indeed, Figure 2 presents two exemplary optimal line plans for two instances of the Parametric City. Each line is depicted as a sequence of colored arrows, indicating its direction and a corresponding frequency. The left instance is highly artificial: Here we set the capacity of the vehicles to the total patronage, which means that all passengers fit into a single vehicle. Thus, we just need to ensure that there is a line visiting all stations in order to pick up all passengers, which results in the illustrated – clearly asym-

2

metric – optimal line plan. However, there are also examples with more realistic parameter choices which also result in asymmetric line plans, such as the one depicted on the right: Here, there are two orange lines of frequency 6, two green lines of frequency 7 and even a circular line traveling clockwise between SC_1, SC_0 and CD with frequency 1. Instead of, for example, using only central lines of frequency 7, it is beneficial to force a few of the passengers from SC_3 with destination CD to take a detour via the neighboring subcenters SC_2 or SC_4. By that, we can set the frequency to 6 on the orange line (which then runs at full capacity), while the green lines with frequency 7 can still fit the surplus of passengers from their neighbors. As neither the passenger paths, nor the line plans are rotated copies of each other, this line plan is asymmetric.

We investigate in this paper the properties of symmetric line plans in comparison to asymmetric ones within the Parametric City. When do symmetric input data result in optimal symmetric line plans? Under which circumstances are symmetric line plans a particularly bad choice? Our aim is to provide planners with a guideline: When are they justified in assuming symmetry? When can they use a symmetric solution as a reasonable approximation? In which cases is it detrimental to assume symmetry? To answer these questions, we model the line planning problem in the Parametric City as a mixed integer program, taking all possible lines into account (including unidirectional ring lines). It turns out that the properties of the Parametric City allow to reformulate this model in an arc-based way, which is computationally efficient. With this model, we can quantify the *Symmetry Gap* between optimal symmetric and asymmetric line plans. It turns out to be small in numerical experiments, and it is also possible to derive analytic bounds. In particular, if one of the parameters of the Parametric City is fixed, namely, the distance factor between the subcenters and their peripheries g, a $\left(1 + \frac{1+\sqrt{2}}{g}\right)$-factor approximation algorithm (with respect to the symmetry gap) can be derived. These results justify the use of symmetric line plans in practice. If the parameter g is not restricted, instances with arbitrary symmetry gaps exist.

2   Line Planning in the Parametric City

The Parametric City is described by three topological parameters, the number $n \in \mathbb{N}$ of radial connections, a radius $T > 0$, and an offset factor $g > 0$, and five demand parameters, the patronage $Y > 0$, and shares $0 < a, \alpha, \gamma, \tilde{\alpha}, \tilde{\gamma} < 1$. Line planning involves three further parameters, an objective weighing factor $\mu \in [0, 1]$, a vehicle capacity $K > 0$, and a frequency bound $\Lambda > 0$.

2.1   The Parametric City

The Parametric City (Fielbaum et al., 2016b) consists of an infrastructure graph and an associated origin-destination matrix. The graph is a planar embedding of a *helm graph* (Gallian, 2000) $\mathcal{G} = (V, A)$ with $2n + 1$ vertices and $3n$ pairs of anti-parallel arcs, where $n \geq 4$ is some natural number (see Figure 1 for $n = 6$). The central node CD at the origin is the *central business district*. It is connected to a cycle of n surrounding nodes $SC_0, SC_1, \ldots SC_{n-1}$, called *subcenters*, each of which is connected to a *periphery* node $P_0, P_1, \ldots P_{n-1}$. The subcenters and peripheries are placed around CD at distances $T > 0$ and $T(1 + g), g > 0$, at coordinates

$$SC_j = T[\sin(j\,2\pi/n), \cos(j\,2\pi/n)]^\top \qquad\qquad \text{for } j \in \mathbb{Z}/n\mathbb{Z},$$
$$P_j = T(1 + g)[\sin(j\,2\pi/n), \cos(j\,2\pi/n)]^\top \qquad\qquad \text{for } j \in \mathbb{Z}/n\mathbb{Z}.$$

The choice of the parameters n, T, and g completely determines the size and shape of the graph underlying the Parametric City. In particular, periphery P_i is placed on the continuation of the central axis through CD and SC_i, and the distance between two adjacent subcenters is $r_n T = 2 \sin(\pi/n) T$. In general, we denote by τ_a the length of an arc $a \in A$.

s, t	P_i	$P_j, j \neq i$	SC_i	$SC_j, j \neq i$	CD
P_i	0	0	$\frac{aY}{n}\beta$	$\frac{aY}{n(n-1)}\gamma$	$\frac{aY}{n}\alpha$
SC_i	0	0	0	$\frac{(1-a)Y}{n(n-1)}\tilde{\gamma}$	$\frac{(1-a)Y}{n}\tilde{\alpha}$
CD	0	0	0	0	0

Table 1: Demand $d_{s,t}$ of the Parametric City

The demand pattern is described by the origin-destination matrix $OD = (d_{s,t})_{(s,t) \in V \times V}$ according to Table 1 for $i, j \in \mathbb{Z}/n\mathbb{Z}$. It is supposed to model the morning rush hour: The peripheries are considered to be pure trip generators, CD is a pure trip attractor, and the subcenters are mixed districts, both generating and attracting trips. Parameter $Y > 0$ is the patronage. Parameter $a \in]0, 1[$ controls the share of travelers originating from the peripheries, so aY passengers come from a periphery, while $(1-a)Y$ start in a subcenter and none start at CD. Parameters $\alpha, \beta, \gamma \in]0, 1[$ are the percentages of trips from the peripheries to CD, their own subcenter, and the other subcenters, respectively. Similarly, $\tilde{\alpha} \in]0, 1[$ are the percentages of trips from the subcenters to CD and $\tilde{\gamma} \in]0, 1[$ the percentage of trips from the subcenters to other subcenters. It holds $\frac{\alpha}{\tilde{\alpha}} = \frac{\gamma}{\tilde{\gamma}}$ or equivalently $\tilde{\alpha} = \frac{\alpha}{\alpha + \gamma}, \tilde{\gamma}, = \frac{\gamma}{\alpha + \gamma}$ since $\alpha + \beta + \gamma = 1$ and $\tilde{\alpha} + \tilde{\gamma} = 1$ because it is assumed that CD is as attractive from a periphery as from a subcenter. Note that – while the OD-matrix is not symmetric in the classical sense, i.e., we have $d_{P_0, CD} \neq d_{CD, P_0}$ – the demand itself is rotation symmetric: For example, the demand from a periphery to the central business district is the same as that of any other periphery to CD.

A city like Berlin, which has no dominating center, but many (roughly) equally important subcenters, can be represented by choosing relatively large values for n, T, g to mimic the geometry, and a fairly large value for γ to model the dispersion of trips within the city.

2.2 The Line Planning Problem

The line planning problem (LPP) in the Parametric City can be formulated as a mixed integer program using two types of variables: $y_p \in \mathbb{R}$ for the passenger flow on path $p \in P$, and $f_l \in \mathbb{N}$ for the frequency of line $l \in L$, see, e.g., Borndörfer et al. (2007). Here, P is the set of all simple paths, while L is the line pool consisting of all simple directed cycles in $\mathcal{G}$ (a typical line will either be an unidirectional "ring line" or a path that is traversed forth and back). We refer to $f = (f_l)_{l \in L}$ as the line plan and say that a line l is part of the line plan if and only if $f_l > 0$; the same holds for the passenger flow. We define the sets P_a and L_a as the set of paths and lines which use arc $a \in A$, respectively. Further, $P_{s \to t}$ is the set of all s-t-paths. The resulting model is as follows.

Definition 2.1 (Line Planning Problem)**.**

$$(LPP)(K, \Lambda, \mu) \quad \min_{f, y} \quad \mu \sum_{l \in L} \tau_l f_l + (1 - \mu) \sum_{p \in P} \tau_p y_p =: cost(f, y)$$

$$\text{s.t.} \quad \sum_{p \in P_{s \to t}} y_p = d_{st} \qquad \forall (s, t) \in D \tag{1a}$$

$$\sum_{p \in P_a} y_p - \sum_{l \in L_a} f_l K \leq 0 \quad \forall a \in A \tag{1b}$$

$$\sum_{l \in L_a} f_l \leq \Lambda \qquad \forall a \in A \tag{1c}$$

$$f_l \in \mathbb{N} \qquad \forall l \in L \tag{1d}$$

$$y_p \geq 0 \qquad \forall p \in P \tag{1e}$$

The *passenger flow conditions* (1a) in combination with the non-negativity constraints (1e) ensure that passengers get routed and demand is met. The vehicle capacity is denoted by the constant $K > 0$. The *capacity constraints* (1b) guarantee that there is sufficient vehicle space to transport all passengers on arc $a \in A$. Finally, the *street capacity constraints* (1c) prevent overcrowded streets by ensuring that the accumulated frequency of an arc is not larger than a parameter $\Lambda > 0$. The objective is a combination of operator and user costs that are weighed by parameter $\mu \in [0, 1]$; this is standard in most line planning approaches. The running and travel times are equated with the total length of a line or path, i.e., $\tau_l = \sum_{a \in l} \tau_a$ and $\tau_p = \sum_{a \in p} \tau_a$, where τ_a is the length of arc $a \in A$ with respect to the Parametric City model, i.e.,

$$\tau_a = \begin{cases} T & a \in \{(CD, SC_i), (SC_i, CD) \mid i \in \mathbb{Z}/n\mathbb{Z}\}, \\ gT & a \in \{(P_i, SC_i), (SC_i, P_i) \mid i \in \mathbb{Z}/n\mathbb{Z}\}, \\ r_n T & a \in \{(SC_i, SC_{i\pm 1}) \mid i \in \mathbb{Z}/n\mathbb{Z}\}; \end{cases} \tag{2}$$

here and elsewhere, we write $i \pm q$ for $i, q \in \mathbb{Z}/n\mathbb{Z}$ instead of $(i \pm q) \bmod n$.

2.3 An Arc-Based Model

The main computational difficulty with the line planning model LPP is the large number of integer line frequency variables, which is exponential if the line pool L is completely unrestricted (and still quadratic in the number of nodes if L is restricted to simple cycles). However, the line variables are always aggregated over arcs $F_a := \sum_{l \in L_a} f_l$; in particular, for the objective holds

$$\sum_{l \in L} \tau_l f_l = \sum_{l \in L} \sum_{a \in l} \tau_a f_l = \sum_{a \in A} \sum_{l \in L_a} \tau_a f_l = \sum_{a \in A} \tau_a F_a. \tag{3}$$

The aggregation F induces a circulation, and conversely, any integer circulation can be decomposed into an equivalent set of lines. The resulting arc-based line planning model $ALPP$ has only $6n$ integer variables and reads as follows:

Definition 2.2 (Arc-Based Line Planning Problem)**.**

$$(ALPP)(K, \Lambda, \mu) \quad \min_{F, y} \quad \mu \sum_{a \in A} \tau_a F_a + (1 - \mu) \sum_{p \in P} \tau_p y_p =: cost_A(F, y)$$

$$\text{s.t.} \quad \sum_{p \in P_{s \to t}} y_p = d_{s,t} \qquad \forall (s, t) \in V \times V \tag{4a}$$

$$\sum_{p \in P_a} y_p - F_a K \leq 0 \qquad \forall a \in A \tag{4b}$$

$$F_a \leq \Lambda \qquad\qquad \forall a \in A \qquad\qquad (4c)$$

$$\sum_{a \in \delta^+(v)} F_a - \sum_{a \in \delta^-(v)} F_a = 0 \quad \forall v \in V \qquad\qquad (4d)$$

$$F_a \in \mathbb{N} \qquad\qquad \forall a \in A \qquad\qquad (4e)$$

$$y_p \geq 0 \qquad\qquad \forall p \in P \qquad\qquad (4f)$$

The arc-based formulation is not only smaller, it will also turn out that it is easier to analyze than the line-based one. We note the following properties of feasible solutions of *ALPP* :

Observation 2.3. *If (F, y) is feasible for ALPP, then*

1. $F_{P_i, SC_i} = F_{SC_i, P_i} \geq \left\lceil \frac{Y_a}{nK} \right\rceil$ *for all $i \in \mathbb{Z}/n\mathbb{Z}$, with equality if (F, y) is optimal and $\mu \neq 0$.*

2. $F_{(SC_k, CD)} > 0$ *for some $k \in \mathbb{Z}/n\mathbb{Z}$.*

3. $F_{(CD, SC_k)} > 0$ *for some $k \in \mathbb{Z}/n\mathbb{Z}$.*

3 Symmetry

While the graph $G = (V, A)$ of the Parametric City is clearly rotation symmetric, it is not so clear what a symmetric demand, passenger flow, or line plan is, and in fact, different notions can be considered.

Definition 3.1 (Rotation). We identify a vertex $v \in V$ with its coordinates in the plane and define its rotation ρ_z by the angle $z \cdot 2\pi/n$ around the origin, $z \in \mathbb{Z}/n\mathbb{Z}$, as

$$\rho_z : V \to V, \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \mapsto R_z v = \begin{bmatrix} cos(\frac{2\pi}{n} z) & -sin(\frac{2\pi}{n} z) \\ sin(\frac{2\pi}{n} z) & cos(\frac{2\pi}{n} z) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

Rotations can be extended to arcs and arbitrary vertex tuples componentwise:

$$\rho_z : V^m \to V^m, \quad (v_0, v_1, \dots, v_{m-1}) \mapsto (R_z v_0, R_z v_1, \dots, R_z v_{m-1}).$$

Rotating the vertices of the Parametric City by ρ_z results in

$$\rho_z(CD) = R_z \begin{bmatrix} 0 \\ 0 \end{bmatrix} = CD,$$

$$\rho_z(SC_j) = R_z SC_j = T \begin{bmatrix} cos(\frac{(j+z)2\pi}{n}) \\ sin(\frac{(j+z)2\pi}{n}) \end{bmatrix} = SC_{j+z},$$

$$\rho_z(P_j) = R_z P_j = T(1+g) \begin{bmatrix} cos(\frac{(j+z)2\pi}{n}) \\ sin(\frac{(j+z)2\pi}{n}) \end{bmatrix} = P_{j+z}.$$

Consequently, subcenters get rotated onto subcenters, peripheries onto peripheries, and the central business district remains fixed. We regard a property of the Parametric City to be symmetric if it has the same value for all rotations ρ_z. The demand is then rotation symmetric in this sense. In the same manner, we define symmetric solutions:

Definition 3.2 (Symmetric Solution). Consider a solution (f, y) to *LPP* and the equivalent solution (F, y) to *ALPP*.

1. The line plan f is *line-symmetric* if for all $l \in L :$ $f_l = f_{\rho_z(l)}$ for all $z \in \mathbb{Z}/n\mathbb{Z}$.

2. The passenger flow y is *path-symmetric* if for all $p \in P :$ $y_p = y_{\rho_z(p)}$ for all $z \in \mathbb{Z}/n\mathbb{Z}$.

3. The line plan f is *arc-symmetric* if for all $a \in A$: $\sum_{l \in L_a} f_l = \sum_{l \in L_{\rho_z(a)}} f_l$ for all $z \in \mathbb{Z}/n\mathbb{Z}$.

4. The frequency plan F is *arc-symmetric* if $F_a = F_{\rho_z(a)}$ for all $z \in \mathbb{Z}/n\mathbb{Z}$.

The solution (f, y) is called *symmetric* if Conditions 1 and 2 hold, while (F, y) is *symmetric* if Conditions 2 and 4 hold.

These definitions give rise to the following sequence of results.

Lemma 3.3. *Any symmetric line plan f is is arc-symmetric.*

Lemma 3.4. *An arc-symmetric frequency plan of a feasible solution (F, y) has the following properties:*

- $F_{(P_0, SC_0)} = F_{(P_z, SC_z)} = F_{(SC_z, P_z)} \in \mathbb{N}_{>0}$ *for all $z \in \mathbb{Z}/n\mathbb{Z}$,*

- $F_{(SC_0, SC_1)} = F_{(SC_z, SC_{z+1})}$ *for all $z \in \mathbb{Z}/n\mathbb{Z}$,*

- $F_{(SC_0, SC_{n-1})} = F_{(SC_z, SC_{z-1})}$ *for all $z \in \mathbb{Z}/n\mathbb{Z}$,*

- $F_{(CD, SC_0)} = F_{(CD, SC_z)} = F_{(SC_z, CD)} \in \mathbb{N}_{>0}$ *for all $z \in \mathbb{Z}/n\mathbb{Z}$.*

We denote the corresponding frequency values by F_P, F_{S+}, F_{S-}, and F_C, i.e., $F_P := F_{(P_0, SC_0)}$, $F_{S+} := F_{(SC_0, SC_1)}$, $F_{S-} := F_{(SC_1, SC_0)}$ and $F_C := F_{(SC_0, CD)}$, see Figure 3.

Lemma 3.5. *For any arc-symmetric solution (F, y) to ALPP, there exists a line-symmetric solution (f, y) to LPP with the same objective value.*

Proof. As the previous lemma states, an arc-symmetric solution has at most four different frequencies (F_P on the peripheral arcs, F_C on the axes incident to CD, and F_{S-} and F_{S+} on the arcs between subcenters), see again Figure 3. These frequencies induce a canonical symmetric line plan. □
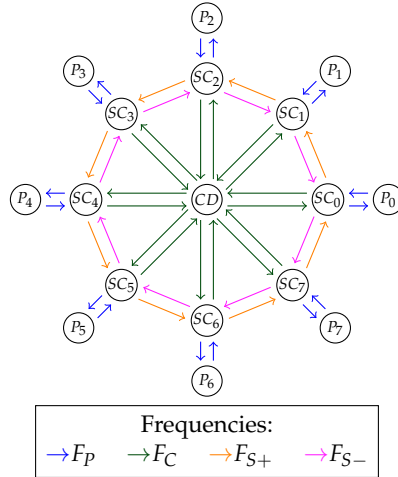
Figure 3: Symmetric Frequency/Line Plan

Lemma 3.6 (Construction of a Symmetric Solution). *For any feasible solution (F, y) of ALPP, a symmetric solution (F^s, y^s) can be constructed as follows:*

$$F_a^s := \left\lceil \frac{1}{n} \sum_{z=0}^{n-1} F_{\rho_z(a)} \right\rceil \quad \text{for } a \in A, \qquad y_p^s := \frac{1}{n} \sum_{z=0}^{n-1} y_{\rho_z(p)} \quad \text{for } p \in P.$$

The user costs $\sum_{p \in P} \tau_p y_p$ remain constant under this symmetrization.

Proof. For a feasible solution (F, y) of $ALPP(K, \Lambda, \mu)$, let $\rho_z(F)$ and $\rho_z(y)$ denote the rotated frequency plan and passenger flow, respectively. Clearly, if (F, y) is feasible, then any rotation $(\rho_z(F), \rho_z(y))$ is feasible as well. Consequently $\sum_{z=0}^{n-1} (\rho_z(F), \rho_z(y))/n$ is feasible for the LP relaxation of $ALPP(K, \Lambda, \mu)$.

Rounding up $\sum_{z=0}^{n-1} \rho_z(F)/n$ to F^s is no problem in the capacity constraints $\sum_{p \in P_a} y_p - F_a K \leq 0$ (4b). This is also true for the street capacity constraints $F_a \leq \Lambda$ (4c). Indeed, for any arc a, they hold in particular for $F_{a_{max}} := \max\{F_{a'} \mid a' = \rho_z(a), z \in \mathbb{Z}/n\mathbb{Z}\} \geq \frac{1}{n} \sum_{z=0}^{n-1} F_{\rho_z(a)}$, and as $F_{a_{max}}$ is integer, they hold for F_a^s. The flow conservation constraints (4d) are also fulfilled: At node CD we have

$$\sum_{a \in \delta^+(CD)} F_a = \sum_{a \in \delta^-(CD)} F_a.$$

This implies that $F_a^s = F_{\bar{a}}^s$ for all $a \in \cup_{z=0}^{n-1}(SC_z, CD), \bar{a} \in \cup_{z=0}^{n-1}(CD, SC_z)$. The same holds at each node $P_j, z \in \mathbb{Z}/n\mathbb{Z}$. Because F^s is symmetric, we have $F_{SC_0, SC_1}^s = F_{\rho_z(SC_0, SC_1)}^s$, in particular $F_{SC_j, SC_{j+1}}^s = F_{SC_{j-1}, SC_j}^s$; as well as $F_{SC_{j+1}, SC_j}^s = F_{SC_j, SC_{j-1}}^s$. Consequently, flow is preserved in each node SC_j as well. Finally, the user costs remain the same, since they depend only on path lengths, which are identical for every rotation. $\qquad \square$

Corollary 3.7. *ALPP is feasible if and only if a symmetric solution exists.*

Lemma 3.8. *If (F, y) is a feasible arc-symmetric solution to ALPP, there exists a symmetric passenger flow $\tilde{y}$ such that the solution $(F, \tilde{y})$ is feasible and $cost_A(F, y) = cost_A(F, \tilde{y})$.*
The same holds for a feasible symmetric solution (f, y) of LPP.

Proof. Construct $(F, \tilde{y}) := (F^s, y^s)$ according to Lemma 3.6; as F is arc-symmetric, $F^s = F$ and the operator costs do not change. For a solution (f, y) of LPP proceed analogously considering aggregated frequency values $F_a := \sum_{l \in L_a} f_l$ for all $a \in A$. $\qquad \square$

Lemma 3.9. *If (F, y) is an optimal solution of ALPP with a symmetric passenger flow, then there exists a symmetric frequency plan $\tilde{F}$ such that $(\tilde{F}, y)$ is feasible and optimal as well.*
The same holds for an optimal path-symmetric solution (f, y) of LPP.

Proof. Denote the aggregated passenger flow per arc by $Y_a = \sum_{p \in P_a} y_p$ for every $a \in A$. As y is symmetric, $Y_a = Y_{\rho_z(a)}$ for any arc $a \in A$ and any $z \in \mathbb{Z}/n\mathbb{Z}$. The integer and capacity requirements further imply $F_a \geq \lceil Y_a/K \rceil$ for all $a \in A$, and by flow conservation at CD and P_i,

$$\sum_{i=0}^{n-1} F_{(SC_i, CD)} = \sum_{i=0}^{n-1} F_{(CD, SC_i)} \quad \text{and} \quad F_{(P_i, SC_i)} = F_{(SC_i, P_i)} \quad \text{for } i \in \mathbb{Z}/n\mathbb{Z}.$$

We define a symmetric frequency plan $\tilde{F}$ by giving the four values $\tilde{F}_P, \tilde{F}_{S+}, \tilde{F}_{S-}, \tilde{F}_C$ according to Lemma 3.4:

$$\tilde{F}_P := F_{(P_0, SC_0)}, \quad \tilde{F}_{S+} := \lceil Y_{(SC_0, SC_1)}/K \rceil, \quad \tilde{F}_{S-} := \lceil Y_{(SC_0, SC_{n-1})}/K \rceil, \quad \tilde{F}_C := \lceil Y_{(SC_0, CD)}/K \rceil.$$

Then $\tilde{F}$ satisfies all capacity constraints on periphery arcs and on arcs connecting subcenters, and also on central district arcs as no trips originate in CD and hence $Y_{(CD, SC_i)} \leq Y_{(SC_i, CD)}$. This shows that $(\tilde{F}, y)$ is feasible. For the optimality of $(\tilde{F}, y)$, we see that by definition of $\tilde{F}$, we have $\tilde{F}_a \leq F_a$ for all arcs except possibly for $a \in \{(CD, SC_i) \mid i \in \mathbb{Z}/n\mathbb{Z}\}$. However, since these arcs and their anti-parallel counterparts have the same lengths, flow conservation at CD implies

$$\sum_{i=0}^{n-1} \tau_{(CD, SC_i)} F_{(CD, SC_i)} = \sum_{i=0}^{n-1} \tau_{(SC_i, CD)} F_{(SC_i, CD)} \geq \sum_{i=0}^{n-1} \tau_{(SC_i, CD)} \tilde{F}_C.$$

Therefore, the cost of $(\tilde{F}, y)$ is at most the cost of (F, y), but the latter was already optimal.

For LPP, consider aggregated frequencies $F_a := \sum_{l \in L_a} f_l$ to obtain arc-frequencies, symmetrize, and construct a symmetric line plan as described in Lemma 3.5. $\qquad \square$

Our discussion can be summarized as follows.

Proposition 3.10 (Sufficient Condition for Symmetry). *An arc-symmetric, line-symmetric, or path-symmetric optimal solution is sufficient for the existence of a symmetric optimal solution.*

In other words, if a symmetric optimal solution exists, then it is enough to find an optimal solution with a symmetric line plan, *or* frequency plan, *or* passenger flow. Instances of the Parametric City can be analyzed with respect to the existence of optimal symmetric solutions by comparing the line optimization model $ALPP$ with its restriction $ALPP$ to arc-symmetric solutions

$$ALPP_{\mathcal{S}} = ALPP$$
$$s.t. \quad F_a = F_{\rho_z(a)} \quad \forall z \in \mathbb{Z}/n\mathbb{Z}, a \in A.$$

If the optimal objective values coincide, there is an arc-symmetric solution to $ALPP$, otherwise, there is a gap.

4 Symmetry Gap

For a feasible (mixed integer) linear program P, we denote the objective value of an optimal solution by $\mathrm{OptVal}(P)$.

Definition 4.1 (Symmetry Gap). For an instance of the line planning problem in the Parametric City, define the *absolute symmetry gap* as

$$\Gamma_{abs} := \mathrm{OptVal}(ALPP_{\mathcal{S}}) - \mathrm{OptVal}(ALPP)$$

and the *(relative) symmetry gap* as

$$\Gamma := \frac{\mathrm{OptVal}(ALPP_{\mathcal{S}}) - \mathrm{OptVal}(ALPP)}{\mathrm{OptVal}(ALPP)}$$

if $ALPP$ is feasible; $\Gamma_{abs} = 0$ and $\Gamma = 0$ otherwise.

The symmetry gap is well defined, since we require the patronage Y as well as demand and arc-lengths in the graph to be positive. Thus, there must exist an arc a with frequency $F_a > 0$ as well as a path p with $y_p > 0$, ensuring that $\mathrm{OptVal}(ALPP) > 0$. Note that $\Gamma_{abs} = 0$ and $\Gamma = 0$ hold if and only if there exists a symmetric optimal solution (F, y) or no solution exists.

4.1 Bounds

Lemma 4.2. *The absolute symmetry gap in the Parametric City is bounded by*

$$\Gamma_{abs} \leq 2\mu \left(\tau_{(SC_0, SC_1)} + \tau_{(CD, SC_0)} \right) (n-1) = 2\mu T (1 + r_n) (n-1).$$

Proof. Let (F, y) be the optimal solution to the unrestricted $ALPP$ and (F^s, y^s) the symmetrized solution as in the proof of Lemma 3.6. As user costs are invariant with respect to symmetrization, see Lemma 3.6, the absolute symmetry gap is bounded by the difference in operator costs:

$$\Gamma_{abs} \leq cost(F^s, y^s) - cost(F, y) = \mu \sum_{a \in A} \tau_a \left(F_a^s - F_a \right).$$

By construction of F^s, $\sum_{z=0}^{n-1} \left(F_{\rho_z(a)}^s - F_{\rho_z(a)} \right) \leq n - 1$ for all $a \in A$. On the peripheral arcs however, by Observation 2.3, the frequencies F and hence F^s are always the same. Therefore, by using the explicit arc costs τ_a (cf. Equation (2)), we obtain an upper bound on the absolute gap:

$$\Gamma_{abs} \leq \mu \sum_{a \in A} \tau_a \left(F_a^s - F_a \right)$$

$$\leq \mu \left(\sum_{z=0}^{n-1} T r_n (F_{\rho_z(SC_0, SC_1)}^s - F_{\rho_z(SC_0, SC_1)} + F_{\rho_z(SC_1, SC_0)}^s - F_{\rho_z(SC_1, SC_0)}) \right.$$

$$\left. + \sum_{z=0}^{n-1} T \left(F_{\rho_z(SC_0, CD)}^s - F_{\rho_z(SC_0, CD)} + F_{\rho_z(CD, SC_0)}^s - F_{\rho_z(CD, SC_0)} \right) \right)$$

$$\leq 2\mu \left(T r_n (n - 1) + T (n - 1) \right) = 2\mu T (1 + r_n) (n - 1). \qquad \square$$

Corollary 4.3. *If $\mu = 0$ and $ALPP$ is feasible, then there exists a symmetric, optimal solution to the line planning problem for any instance of the Parametric City.*

To find an upper bound on the relative symmetry gap, we first need to determine a lower bound on the objective of $ALPP$: Such a lower bound, actually to the LP relaxation of $ALPP$, is provided by a certain minimum-cost flow problem. Its optimal objective value can in turn be bounded from below by a term in parameters of the Parametric City that is independent from the number of zones and from the demand pattern.

Definition 4.4 (Uncapacitated Minimum-Cost Flow Problem).

$$(UMCFP) \quad \min_{y} \quad \sum_{p \in P} \bar{c}_p y_p = cost_{UMCFP}(y)$$

$$\text{s.t.} \quad \sum_{p \in P_{st}} y_p = d_{s,t} \quad \forall (s,t) \in D \qquad (5a)$$

$$y_p \geq 0 \qquad \forall p \in P \qquad (5b)$$

with $\bar{c}_p := \sum_{a \in p} \bar{c}_a$ and

$$\bar{c}_a := \begin{cases} (2\mu/K + (1 - \mu))T g & \text{for } a = (P_i, SC_i), \\ (1 - \mu)T g & \text{for } a = (SC_i, P_i), \\ (\mu/K + (1 - \mu))T r_n & \text{for } a = (SC_i, SC_{i\pm1}), \\ (2\mu/K + (1 - \mu))T & \text{for } a = (SC_i, CD), \\ (1 - \mu)T & \text{for } a = (CD, SC_i). \end{cases}$$

Proposition 4.5. *The optimal objective value of the minimum-cost flow problem UMCFP is given by*

$$\text{OptVal}(UMCFP) = TY\lambda(\alpha, \gamma),$$

(a) Shortest paths for one zone

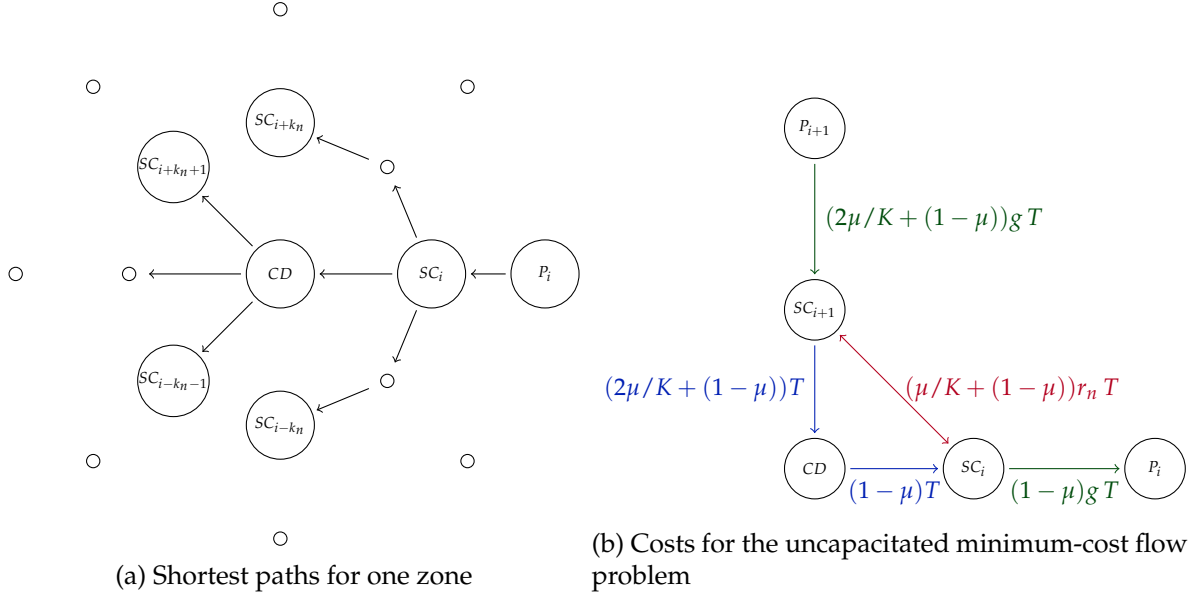(b) Costs for the uncapacitated minimum-cost flow problem

Figure 4: Shortest path-related parameters

where

$$\lambda(\alpha, \gamma) = \frac{k_n(k_n + 1)r_n + 2(n - 2k_n + 1)}{n - 1} \left(\frac{\mu}{K} + (1 - \mu) \right) (a\gamma + (1 - a)\tilde{\gamma})$$
$$+ \left(\frac{2\mu}{K} + (1 - \mu) \right) (a\alpha + (1 - a)\tilde{\alpha} + ga),$$

and $k_n := \left\lfloor \frac{2}{r_n} \right\rfloor$. Furthermore, independent of the number of zones and the demand pattern,

$$(1 + ga - a) \left(\left(2 - \frac{2}{\pi} \right) \frac{\mu}{K} + 1 - \mu \right) \ \leq \ \lambda(\alpha, \gamma) \ \leq \ 4(1 + ga) \left(\frac{2\mu}{K} + 1 - \mu \right).$$

Proof. We can solve $UMCFP$ by determining the shortest path for each origin-destination pair. For each zone, these are the paths as depicted in Figure 4a. The passenger flow is determined by the demand. The details can be found in the appendix. $\square$

Lemma 4.6. *For any instance of the Parametric City, the optimal objective value of the uncapacitated minimum-cost flow problem ($UMCFP$) provides a lower bound for the line planning problem, i.e.,*

$$\text{OptVal}(UMCFP) \leq \text{OptVal}(ALPP).$$

Proof. $UMCFP$ is equivalent to the following relaxation of $ALPP$: Relax the integrality constraints to non-negativity constraints, as well as the street-capacity constraints ($\Lambda \to \infty$). There are no travelers towards peripheries, the effective cost of a flow being directed via arc (P_i, SC_i) is consequently twice as high, to compensate for the "empty seats" in the opposite direction. Similarly at the central node: Since there are no travelers originating from the central node, any flow using an outward arc from CD must travelalong some arc (SC_i, CD), any traveler remaining in the center "blocks" a seat on the outgoing arcs. Consequently, one can assign twice the costs for the incoming arcs at CD and set costs along the outgoing arcs to zero. A detailed proof can be found in the appendix. $\square$

Corollary 4.7. *For any instance of the Parametric City with fixed parameters $K \in \mathbb{N}_{>0}$ and $\mu \in [0,1]$, and $\lambda(\alpha, \gamma)$ as defined in Proposition 4.5, the symmetry gap is at most*

$$\Gamma \leq \mu \frac{2(1+r_n)(n-1)}{Y\lambda(\alpha,\gamma)}.$$

Corollary 4.8. *The symmetry gap in the Parametric City is bounded by*

$$\Gamma \leq \mu \frac{2(1+r_n)(n-1)}{Y\left(\left(2-\frac{2}{\pi}\right)\frac{\mu}{K}+1-\mu\right)(1+ga-a)},$$

for fixed parameters $K \in \mathbb{N}_{>0}, \mu \in [0,1]$, and $\Lambda \geq 1$, independent of the demand parameters α and γ.

For a realistic choice of parameters, the two terms $\left(\left(2-\frac{2}{\pi}\right)\frac{\mu}{K}+1-\mu\right)$ and $(1+ga-a)$ carry little weight, and the symmetry gap is dominated by the ratio $(n-1)/Y$, which will be very small in real-life applications. In general, a better bound can be derived by establishing a lower bound on $\mathrm{OptVal}(MILP_A)$ which takes effect when the lower bound derived from the uncapacitated minimum-cost flow problem goes to zero.

Lemma 4.9 (Lower Bound on Operator Costs). *For any instance of the Parametric City, the operator costs $\mu \sum_{a \in A} \tau_a F_a$ are at least $\mu T(2ng + 2 + (n-1)r_n)$.*

Proof. Suppose F is a feasible solution. From Lemma 2.3 we know that $F_a = F_{a'} \geq 1$ for all arcs a, a' incident with some periphery P_i. This means that

$$\sum_{z=0}^{n-1} F_{\rho_z(SC_0, P_0)} + F_{\rho_z(P_0, SC_0)} \geq 2n. \tag{6}$$

Furthermore, at least one of the incoming arcs as well as one of the outgoing arcs at CD needs to be used and these numbers coincide by flow conservation, i.e.,

$$\sum_{z=0}^{n-1} F_{\rho_z(SC_0, CD)} = \sum_{z=0}^{n-1} F_{\rho_z(CD, SC_0)} \geq 1 \tag{7}$$

Now consider the subgraph $W = (\bar{V}, \bar{A})$, with $\bar{V} = \{CD, SC_0, SC_1, \ldots, SC_{n-1}\}$ and $\bar{A} = \{a \in G[\bar{V}] : F_a > 0\}$. Due to the positive demand and flow conservation for F, W must be a strongly connected and there must be at least $n+1$ arcs in W, so that

$$\sum_{z=0}^{n-1} \left(F_{\rho_z(SC_0, SC_1)} + F_{\rho_z(SC_1, SC_0)} + F_{\rho_z(SC_0, CD)} + F_{\rho_z(CD, SC_0)} \right) \geq n+1.$$

Equation (7) and the existence of an arc (SC_j, CD) with $F_{(SC_j, CD)} \geq 1$ imply

$$\sum_{z=0}^{n-1} \left(F_{\rho_z(SC_0, SC_1)} + F_{\rho_z(SC_1, SC_0)} \right) + 2 \sum_{\substack{z=0 \\ z \neq j}}^{n-1} F_{\rho_z(SC_0, CD)} \geq n-1.$$

The operator costs therefore satisfy

$$\mu \sum_{a \in A} \tau_a F_a = \mu \left(2 \sum_{z=0}^{n-1} Tg F_{\rho_z(SC_0, P_0)} + \sum_{z=0}^{n-1} Tr_n F_{\rho_z(SC_0, SC_1)} + \sum_{z=0}^{n-1} Tr_n F_{\rho_z(SC_1, SC_0)} \right.$$

$$\left. + 2 \sum_{\substack{z=0 \\ z \neq j}}^{n-1} T F_{\rho_z(SC_0, CD)} + 2TF_{(SC_j, CD)} \right)$$

$$\geq \mu(2nTg + (n-1)\min\{2T, Tr_n\} + 2T).$$

Since $r_n = 2\sin(\pi/n) < 2$ for all $n \geq 4$ the claim follows. $\qquad \square$

Corollary 4.10 (Upper Bound on Γ). *For any instance of the Parametric City, and independent on the choice of parameters, except for g, the symmetry gap can be bounded by the constant*

$$\Gamma \leq \frac{(1+\sqrt{2})}{g}.$$

Proof. Lemma 4.9 and the fact that the objective value of $ALPP$ is a sum of operator and user costs imply $\text{OptVal}(ALPP) \geq \mu \sum_{a \in A} \tau_a F_a$. Using this and the fact that $r_n \leq \sqrt{2} < 2$ for all $n \in \mathbb{N}_{\geq 4}$ gives rise to the bound

$$\Gamma \leq \mu \frac{2T(1+r_n)(n-1)}{\text{OptVal}(ALPP)} \leq \frac{2(1+r_n)}{2g+r_n} \leq \frac{(1+\sqrt{2})}{g}. \qquad \square$$

Proposition 4.11. *The relative symmetry gap Γ in the Parametric City is bounded as follows:*

$$\Gamma \leq C_n(\alpha, \gamma) \leq C_n \leq \frac{(1+\sqrt{2})}{g}$$

where

$$C_n(\alpha, \gamma) := \min \begin{cases} \frac{\mu(n-1)\,2\,(1+r_n)}{Y\rho(\alpha,\gamma)}, \\ \frac{2(1+r_n)}{2g+r_n}, \end{cases} \quad \text{and} \quad C_n := \min \begin{cases} \frac{\mu(n-1)\,2(1+r_n)}{Y\left(\left(2-\frac{2}{\pi}\right)\frac{\mu}{K}+1-\mu\right)(1+ga-a)}, \\ \frac{2(1+r_n)}{2g+r_n}. \end{cases}$$
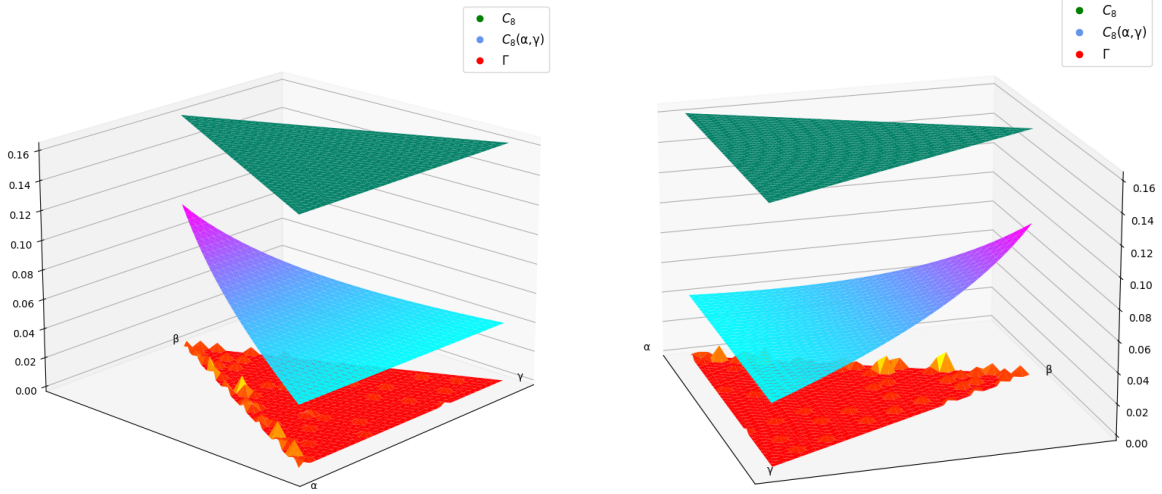

Figure 5: Theoretical symmetry gap for the Parametric City for $n = 8, Y = 24000, K = 100, g = 1/3, \mu = 1, a = 0.8$ in comparison to computational results (cf. Section 5)

For constant patronage Y, the bound C_n increases roughly linearly with the number n of zones, while for constant n, the influence of the term $1/Y$ leaves its mark for small Y. Figure 5 plots the bounds $C_n(\alpha, \gamma)$ and C_n for (realistic) parameter choices in comparison to the actual gap Γ obtained from computational experiments. It is evident that $C_n(\alpha, \gamma)$ is considerably closer to the symmetry gap than C_n. However, even C_n is smaller than 0.16 independent of the demand, less than 0.12 as a function of α, β, and γ, and less than 0.04 for small β. The real gap Γ is considerably smaller – in most cases, the symmetric solutions are optimal, and otherwise, the gap is below 1.22% (see Figure 10 and Table 2b for additional comparison). This is good news for city planners.

13

4.2 Approximation Algorithm

We will show in this subsection that the symmetrization of an optimal solution of the line planning problem in the Parametric City actually gives rise to an approximation algorithm. Using the observtion that symmetric solutions have at most four different arc-frequencies (Lemma 3.4) gives rise to an alternative formulation of $ALPP_S$ involving only three integer valued variables. Indeed, we can add the following restrictions to $ALPP_S$ w.l.o.g.:

$$F_a = F_{\bar{a}} \qquad \text{for all } a \in \bigcup_{z=0}^{n-1}\{(SC_z, CD)\} \text{ and } \bar{a} \in \bigcup_{z=0}^{n-1}\{(CD, SC_z)\},$$

$$F_a = F_{\bar{a}} = \left\lceil \frac{Ya}{nK} \right\rceil \qquad \text{for all } a \in \bigcup_{z=0}^{n-1}\{(SC_z, P_z)\} \text{ and } \bar{a} \in \bigcup_{z=0}^{n-1}\{(P_z, SC_z)\}.$$

The latter is a simple assignment constraint that fixes the frequencies along the peripheral arcs to optimal vaules. The resulting model is as follows.

Definition 4.12 (3-Integer Symmetric Line Planning Problem).

$$(ALPP3_S) \min_{F_{S+}, F_{S-}, F_C, y} \mu \left(\kappa_C F_C + \kappa_S F_{S+} + \kappa_S F_{S-} \right) + (1 - \mu) \sum_{p \in P} \tau_p y_p + \mu \, \kappa_P F_P$$

$$\text{s.t.} \quad \sum_{p \in P_{s \to t}} y_p = d_{s,t} \qquad \forall (s,t) \in D \tag{8a}$$

$$\sum_{p \in P_a} y_p - F_C K \leq 0 \qquad \forall a \in \{(CD, SC_z), (SC_z, CD), z \in \mathbb{Z}/n\mathbb{Z}\} \tag{8b}$$

$$\sum_{p \in P_a} y_p - F_{S+} K \leq 0 \qquad \forall a \in \{(SC_z, SC_{z+1}), z \in \mathbb{Z}/n\mathbb{Z}\} \tag{8c}$$

$$\sum_{p \in P_a} y_p - F_{S-} K \leq 0 \qquad \forall a \in \{(SC_{z+1}, SC_z), , z \in \mathbb{Z}/n\mathbb{Z}\} \tag{8d}$$

$$F_C, F_{S+}, F_{S-} \leq \Lambda \tag{8e}$$

$$F_C, F_{S+}, F_{S-} \in \mathbb{N} \tag{8f}$$

$$y_p \geq 0 \qquad \forall p \in P \tag{8g}$$

with the constant $F_P = \left\lceil \frac{Ya}{nK} \right\rceil \leq \Lambda$ and cost parameters

$$\kappa_C = 2n\tau_{(SC_0, CD)} = 2n \, T$$
$$\kappa_S = n\tau_{(SC_0, SC_1)} = n \, r_n T$$
$$\kappa_P = 2n\tau_{(SC_0, P_0)} = 2n \, Tg.$$

Proposition 4.13. *The symmetric line planning problem $ALPP_S$ is solvable in polynomial time.*

Proof. Mixed-integer programming problems with a fixed number of variables are solvable in polynomial time, as was proven by Lenstra (1983). $\qquad\square$

Definition 4.14 (Approximation Algorithm (LPA)).
Input: Instance of a Parametric City
Output: Best symmetric solution (f, y) to LPP if feasible, $\varnothing$ otherwise.
1. Solve $ALPP_S$. If feasible, retrieve (F, y), otherwise return $\varnothing$.
2. Use (F, y) to receive (f, y) as in Lemma 3.5. Return (f, y).

Theorem 4.15. *LPA is a κ-factor approximation algorithm for the line planning problem in the Parametric City for fixed g, where $\kappa = 1 + \frac{(1+\sqrt{2})}{g}$.*

Proof. LPA is a polynomial time algorithm, as $ALPP_{\mathcal{S}}$ can be solved in polynomial time, cf. Proposition 4.13, and we can assign frequencies to lines as described in Lemma 3.5 in $O(n)$. Let PC_g be the set of all instances of Parametric Cities for a given value of the parameter g. Due to the the upper bound on the symmetry gap from Proposition 4.11, the performance ratio $\frac{\text{OptVal}(ALPP_{\mathcal{S}})}{\text{OptVal}(ALPP)}$ is bounded by

$$\sup_{pc \in PC_g} \frac{\text{OptVal}(ALPP_{\mathcal{S}})}{\text{OptVal}(ALPP)} = \max_{pc \in PC_g} (1 + \Gamma) \leq 1 + \frac{1 + \sqrt{2}}{g}. \qquad \square$$

We now give a family of instances that shows that the symmetry gap is indeed unbounded in the worst case.

Proposition 4.16. *Let PC be the set of all instances of the Parametric City. Then the worst case performance ratio of the approximation algorithm LPA is unbounded, i.e.,*

$$\sup_{pc \in PC} \frac{\text{OptVal}(ALPP_{\mathcal{S}})}{\text{OptVal}(ALPP)} = \infty.$$

Proof. Consider a Parametric City with parameters Y, α, γ, a, T. Set the scalarization parameter to $\mu = 1$ such that only operator costs are considered. Further, choose any $n \in \mathbb{N}_{\geq 4}$ and a large K, e.g., $K = Y$, which means that the total patronage can fit into a single vehicle. Lastly, choose a small g, say $g = 1/n$.

As all passengers fit into one vehicle, the frequencies of an optimal solution of $ALPP$ and $ALPP_{\mathcal{S}}$ on any arc are either 0 or 1. Consider the following line plan: Set $F_{(SC_0,CD)} = 1$ and $F_{(CD,SC_1)} = 1$, all other arcs incident to CD are set to zero. Further, assign $F_{\rho_z(SC_0,SC_1)} = 1$ for all $z \in \mathbb{Z}/n\mathbb{Z}, z \neq 0$, i.e., use all inter-subcenter arcs in counter clockwise direction except for (SC_0, SC_1). Lastly, use all peripheral arcs, that is $F_a = 1$ for all $a \in \{(P_z, SC_z), (SC_z, P_z), z \in \mathbb{Z}/n\mathbb{Z}\}$. Any other arcs set to zero. This results in the representative line plan in Figure 6a. The cost function reduces to pure operator costs since $\mu = 1$. The objective value is therefore

$$cost(F, y) = 2ng + (n-1)r_n + 2$$

which is exactly the lower bound for operator costs of Lemma 4.9 and must consequently be optimal. Up to rotation and reflection, this solution is the same for all $n \geq 4$.

From previous considerations we already know that in a solution to $ALPP_{\mathcal{S}}$, all arcs incident to the center must have positive and, in particular, equal frequency. As a minimal requirement for a feasible solution we thus have $F_P \geq 1$ and $F_C \geq 1$. Thus, assign frequency 1 to all arcs incident to a periphery or the central business district, set all other arcs to zero. This line plan gives rise to a feasible solution – see Figure 6b for reference: Each node can be reached from any other node. Again, since the vehicle capacity is large enough to fit all passengers of the whole system, we can find admissible passenger paths. The associated objective value is

$$\text{OptVal}(ALPP_{\mathcal{S}}) = T(2ng + 2n).$$

Thus, for the chosen instance of the Parametric City, we have the ratio

$$\frac{\text{OptVal}(ALPP_{\mathcal{S}})}{\text{OptVal}(ALPP)} = \frac{T(2ng + 2n)}{T(2ng + (n-1)r_n + 2)} = \frac{2ng + 2n}{2ng + (n-1)r_n + 2}$$

15

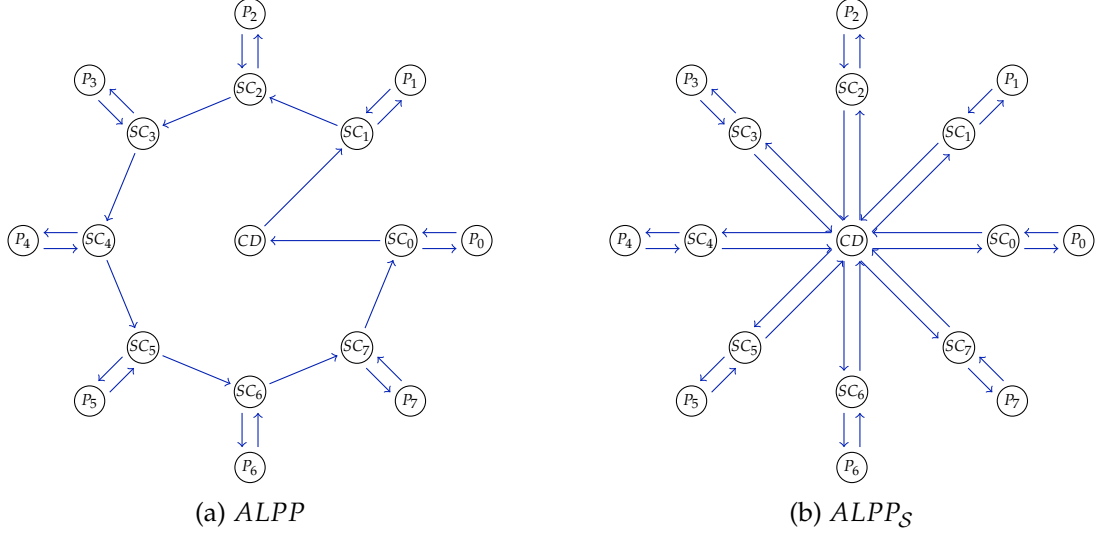(a) *ALPP*   (b) *ALPP$_\mathcal{S}$*

Figure 6: Optimal frequency plans for the extreme example: $n \geq 4, K = Y, g = 1/n$ for any choice of Y, a, α, γ, T; blue arcs indicate Frequency 1.

As we chose $g = 1/n$ and using the fact that $r_n(n-1) \leq 2\pi$, we find the lower bound

$$\frac{2ng + 2n}{2ng + (n-1)r_n + 2} = \frac{2 + 2n}{2 + (n-1)r_n + 2} \geq \frac{2 + 2n}{2 + 2\pi + 2} = \frac{1 + n}{2 + \pi}.$$

As n was chosen freely, this lower bound becomes arbitrarily large for increasing n. The maximal performance ratio over *all* instances of the Parametric City hence goes to infinity. $\square$

5   Computational Study

How important are asymmetric solutions and how much better are they really? We study this question in this section computationally by considering a large number of instances of *ALPP* with varying inputs. In all instances, we choose a total patronage of $Y = 24000$, from which $a = 0.8$ originates in the peripheries. The distance between the subcenters and the central business district is $T = 30$ and the peripheries are at distance gT from the subcenters, with $g = 1/3$. We vary the rest of the parameters: the vehicle capacity K, the number of zones n, and the scalarization parameter μ. We then compare the results in maps dependent on α, β, γ. We solve both *ALPP* and *ALPP$_\mathcal{S}$* to obtain the classification of each instance of the Parametric City for all $\alpha, \gamma \in [0.025, 0.95]$ such that $\alpha + \beta + \gamma = 1$ with a step size of 0.025. We choose the MIP solver Gurobi 9 (Gurobi Optimization, LLC, 2020) and consider a solution to be optimal, when the relative MIP optimality gap is below a tolerance of 10^{-4}.

5.1   Sensitivity Analysis

5.1.1   Influence of the Scalarization Parameter μ

The first set of computational experiments considers the influence of the parameter μ. We evaluated the Parametric City for $n = 8$ for four different choices of μ, namely for the two extreme cases of $\mu = 0$ and $\mu = 1$, as well as $\mu = 0.5$ and 0.75, see Figure 7.

The number of symmetric solutions greatly outweighs the number of asymmetric ones. For $\mu = 0$ there are none, for $\mu = 0.5$ there is only one, for $\mu = 0.75$ there 13, and for $\mu = 1$
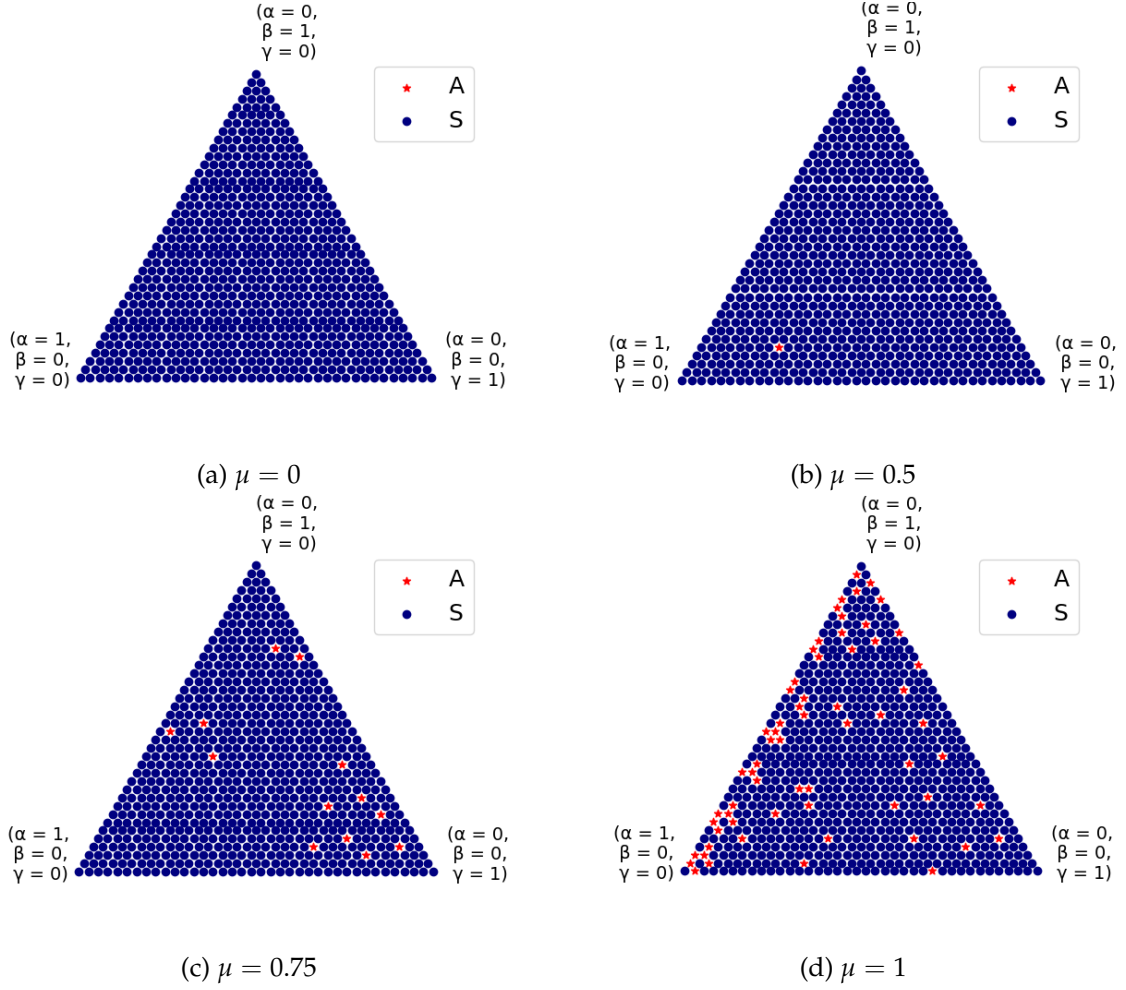
Figure 7: Optimal solutions (A/S: a-/symmetric) for $n = 8, K = 100$ with respect to μ

approximately 8.5% are asymmetric solutions. Asymmetric optimal line plans are hence rare, but not singular or exceptional. Except for $\mu = 0$, i.e., when operator costs are ignored, one cannot assume that optimal solutions are symmetric. Indeed, the number of asymmetric cases increases with μ. This is not particularly surprising: The larger μ, the more focus is on the operators cost and the less expensive is it to reroute people from their preferred, shortest path to some longer detour in order to decrease the frequency along some arcs.

For the following computations, we fix the scalarization parameter to $\mu = 1$, because this choice produces the most diverse results with respect to symmetry. This setting is also relevant from a practical point of view as well: More often than not, the main objective is the minimization of operating costs while providing enough service to cover demand.

5.1.2 Influence of the Number of Zones n

To examine how the number of zones affects the type of solutions, we compared them for $n = \{4, 5, 6, 7, 8\}$. The results can be seen in Figure 8 and Figure 7d for $n = 8$.

The results are somewhat surprising. There are asymmetric solutions for all choices of n, even for the "maximally symmetric" case $n = 6$, in which the distance between two subcenters is the same as between a subcenter and the central business district. The Parametric City with six zones is the only one in which the shortest path between two vertices is not unique, which
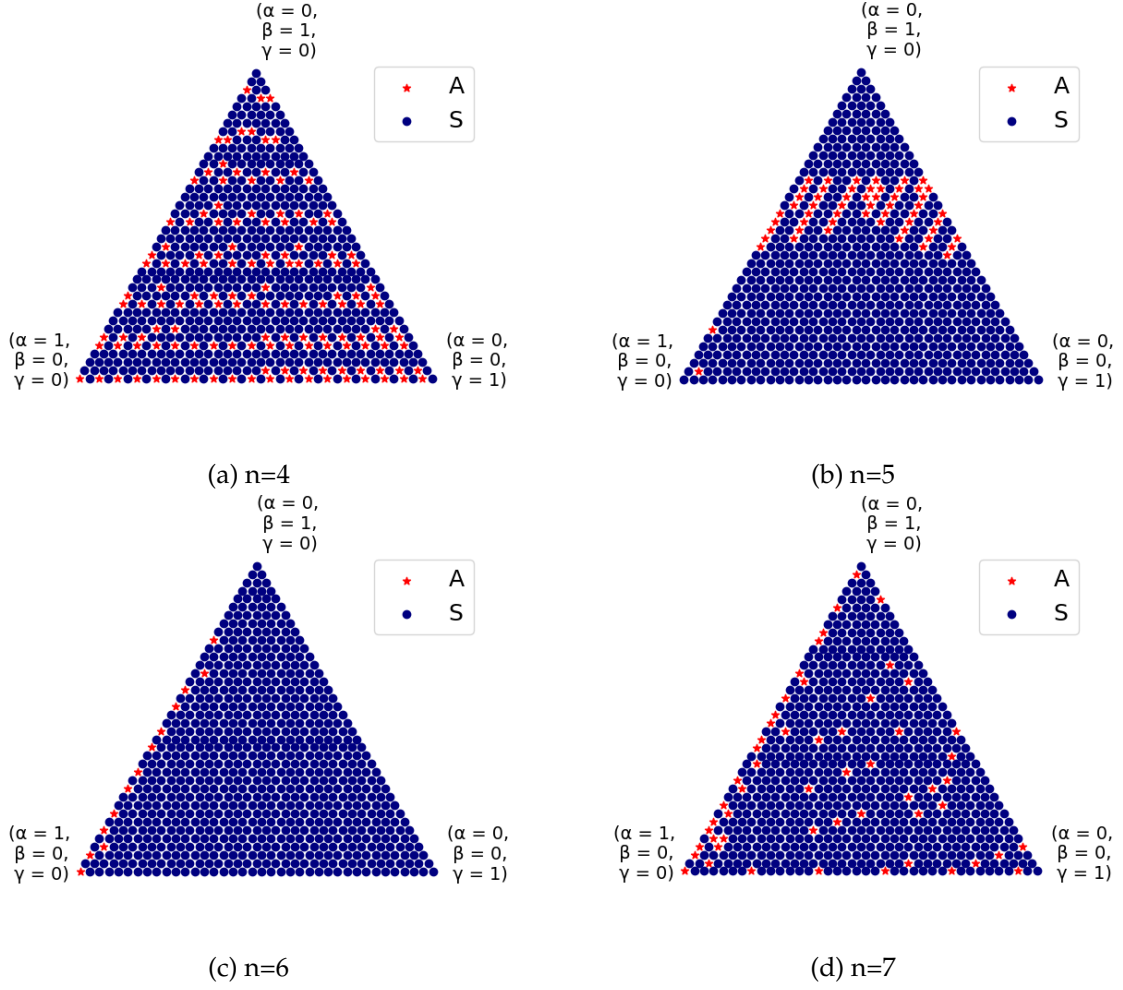
(a) n=4

(b) n=5

(c) n=6

(d) n=7

Figure 8: Classification of optimal solutions for $K = 100, \mu = 1$ in dependence of n

gives rise to more optimal solutions. One could expect more of them to be symmetric. And indeed, the number of asymmetric solutions is reduced in comparison to other choices of n, but they do exist.

Another noticeable aspect is the accumulation of asymmetric cases on the left-hand side of the parameter triangle. This corresponds to low values of γ., i.e., the fraction of total trips from a periphery to *all* other subcenters. The demand between a periphery and a different subcenter is scaled by $\frac{1}{n-1}$ (cf. Table 1), and is thus even smaller. As a consequence, we have more occurrences in which some of these few passengers can be rerouted to other arcs, which still have available capacity, such that the frequency along the direct path can be reduced. This correlates with an observation we will make when investigating the influence of the vehicle capacity K.

5.1.3 Influence of the Vehicle Capacity K

We now study the impact of the vehicle capacity K. For our sample size, it turns out that the absolute number of asymmetric solutions increases with K. For example, for $K = 50$ there are approximately 6.3% asymmetric solutions, while for $K = 100$ we find 8.5%, and for $K = 150$ even 11.2% are asymmetric. This can be explained in the following way: When the vehicle size is larger, it is more likely that there is still some capacity available to accommodate more

18

passengers by detouring them via non-direct paths and thus reduce the frequency elsewhere. This is a similar phenomenon as for small values of γ.
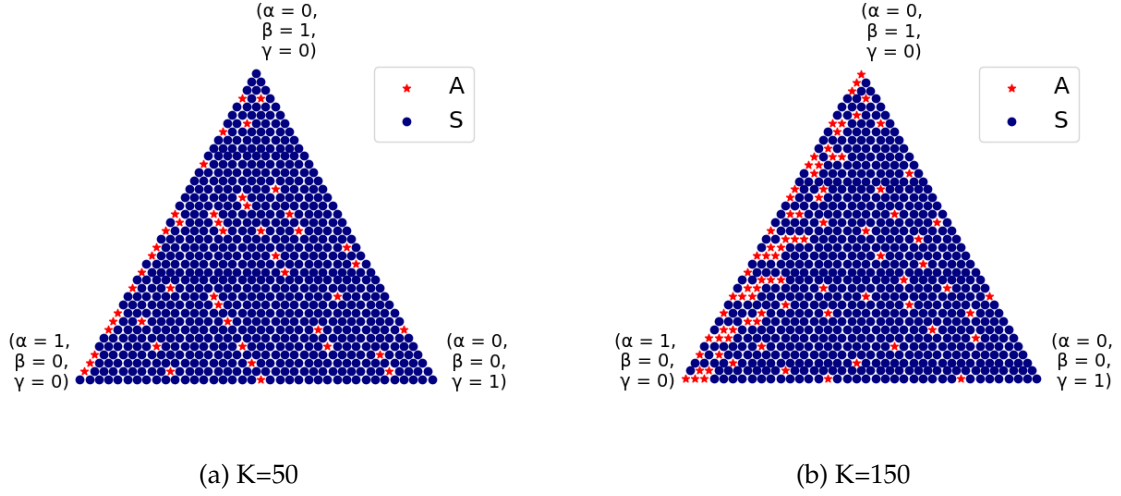

(a) K=50

(b) K=150

Figure 9: Classification of optimal solutions for $n = 8, \mu = 1$ for in dependence of parameter K

5.2 Qualitative Analysis

This section studies the influence of the solution type on the overall costs. We plot the optimal values of $ALPP_S$ and $ALPP$ in a 3D-plot depending on the demand parameters α, β and γ. The cost of $ALPP$ is marked with red dots, the cost of $ALPP_S$ with blue ones. If the costs are equal, the red and blue dots are merged into a purple one for better readability. We include the markers $\alpha = 1, \beta = 1$ and $\gamma = 1$ to indicate the coordinates: At $\alpha = 1$ we have the coordinates $(\alpha, \beta, \gamma) = (1, 0, 0)$ etc. It turns out the absolute symmetry gap is so small that the difference is hard to see with the bare eye. We therefore rather plot relative symmetry gaps as in Figure 10[1]; there, the relative gap is less than 0.012, i.e., symmetric solutions deviate from the optimal solutions by no more than 1.2%.

The relative symmetry gaps look very similar for all instances of the previous section, such that we discuss Figure 10 with $n = 8, K = 100, \mu = 1$ as a general representative. Varying one parameter and fixing all else, we make the following observations:

- Total costs are lowest in the right corner for all data sets, for large β, i.e., when most people travel from the periphery only to their own subcenter. This is an expected result, since the frequency on any arc (P_i, SC_i) is $F_{(P_i, SC_i)} = \lceil \frac{Y_a}{nK} \rceil$ by Lemma 2.3 – independent of α, β, γ. Consequently, if β is large, most of the passengers with origin in the peripheries will travel merely over one peripheral arc, while for smaller β they will use other arcs as well, contributing to a larger total number of frequencies.

- The symmetry gap seems to increase with K: Table 2a records the maximal gap over all tested α, β, γ for each Parametric City with $\mu = 1$ and for $n = 6$ and $n = 8$ for different choices of K.

- In comparison to n we observe maximal gap values according to Table 2b: They increase with n until $n = 6$ and decrease afterwards.

---

[1]The red markers in the graph are included just for orientation: They are put at 0 for depth perception such that the deviation of symmetric solutions from the optimum can be easier ascribed to the corresponding (α, β, γ) tuple.
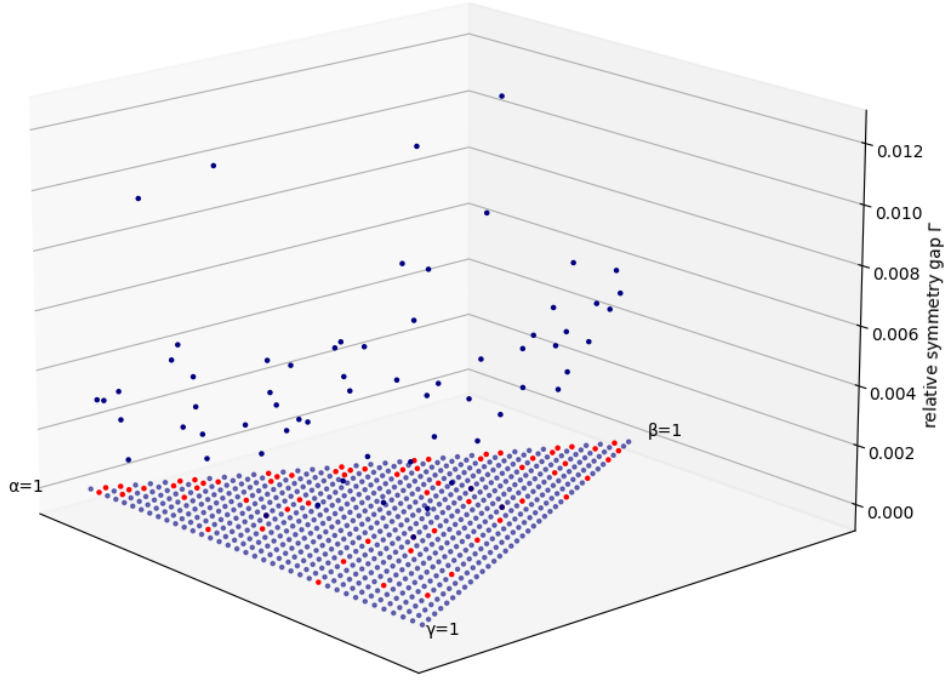
Figure 10: Relative symmetry gap of solutions for $n = 8, K = 100, \mu = 1$

- For increasing μ, the total costs decrease on average, as well as the minimum and maximum values. The latter correlates with the general behavior of the lower bound gained from the cost of the uncapacitated minimum-cost flow problem. In contrast, the maximum relative gap increases with μ in our experiments, as can be seen in Table 2c.

5.3 Comparison to Fielbaum et al.

The next batch of computations compares our results to those of Fielbaum et al. In multiple publications (Fielbaum et al., 2016a, 2018; Fielbaum Schnitzler, 2019), they evaluate the performance of a transportation network with respect to the Value of the Resources Consumed (VRC), a socio-economic cost function first introduced and analyzed by Jansson (1980) and later adjusted to include operator costs and vehicle sizes by Jara-Díaz and Gschwender (2009). The VRC accounts mainly for distance-based travel times from both an operator's and a user's perspective, but it also measures delay due to boarding or alighting, as well as waiting times at vehicle arrivals. For a single line, the VRC can be computed by a formula. The article by Fielbaum et al. (2016a) generalizes the VRC to a multiple-line model and applies it to the Parametric City by numerical minimization.

The transportation network is obtained by solving a variation of the model LPP with an objective that is supposed to approximate VCR. In solving the line planning problem, Fielbaum et al. use heuristc approaches and make the general assumption that one can take "advantage of the symmetry of the city" (Fielbaum Schnitzler, 2019, p. 12). They optimize over one zone only and replicate the optimal solution in the other zones.

The line planning problem that Fielbaum et al. consider differs slightly from ours. Firstly, the line pool differs from ours: They use only "lines that are shortest-path between their origins and destinations" (Fielbaum Schnitzler, 2019, p. 14) to reduce its size. As was described in

20

	$n = 6$			$n = 8$		
K	50	100	150	50	100	150
max. gap	0.82%	1.66%	2.38%	0.51%	1.22%	3.21%

(a) In dependence of K for fixed ($\mu = 1$)

n	4	5	6	7	8
max. gap	0.27%	0.58%	1.65%	1.25%	1.22%

(b) In dependence of n with ($K = 100, \mu = 1$)

	$\mu = 0$	$\mu = 0.5$	$\mu \approx 0.878$	$\mu = 1$
average cost	855 477.3	434 727.5	116 055.4	17 943.7
min. cost	409 138.0	208 296.3	56 167.2	6 974.8
max. cost	1 250 409.4	632 742.0	165 634.4	13 181.0
max. gap	0.0%	0.000611%	0.1%	1.22%

(c) In dependence of μ ($K = 100, n = 8$)

Table 2: Maximal gap of symmetric solutions of computational results dependent on different parameter choices

Section 2.2 our arc-based formulation of the line planning problem allows us to recreate any line corresponding to a directed cycle in the graph. We therefore consider a larger line pool. Secondly, they use fixed values to balance operator costs $c_o = 10.65\$/h$ against user costs $p_v = 1.48\$/h$. In their notation the objective is

$$ \text{“} \sum_{R \in U} p_v\, t_R\, y_R + \sum_{l \in L} c_0\, B_l \text{”} \quad \text{(Fielbaum Schnitzler, 2019, p. 87).} $$

This objective can be directly translated into our notation: U is the set of paths, so it corresponds to our set P, L is the set of lines, t_R is the time spent traveling along path $R \in U$, which means that it corresponds to our parameter τ_p for $p \in P$. The parameter B_l is the fleet size of line l. As they define frequency as "total fleet size divided by cycle time" (Fielbaum Schnitzler, 2019, p. 6), this corresponds to $B_l = c_l f_l$. Therefore, Fielbaum's cost function in our notation can be written as

$$ (c_0 + p_v) \left(\frac{p_v}{c_0 + p_v} \sum_{p \in P} \tau_p y_p + \frac{c_0}{c_0 + p_v} \sum_{l \in L} \tau_l f_l \right). $$

By choosing $\mu = \frac{c_0}{c_0 + p_v}$ this objective corresponds exactly to the one of LPP up to the constant factor $c_0 + p_v$. A constant factor in a linear objective function has no influence on the optimal solution (f, y) of the problem. Therefore, by solving $ALPP$ with $\mu = \frac{c_0}{c_0 + p_v} \approx 0.878$ we consider a comparable objective function.

For this choice of the parameter μ, asymmetric solutions occur relatively frequently, see, e.g., Figure 11a which contains approximately 8.5% of asymmetric optimal line plans. However, the relative symmetry gap is extremely small, the largest gap being less than 0.11%. We conclude that even though symmetry should not be assumed by default, in this particular case, symmetric solutions are close to optimal, and the symmetry assumption of Fielbaum et al. is, in this sense, justified.
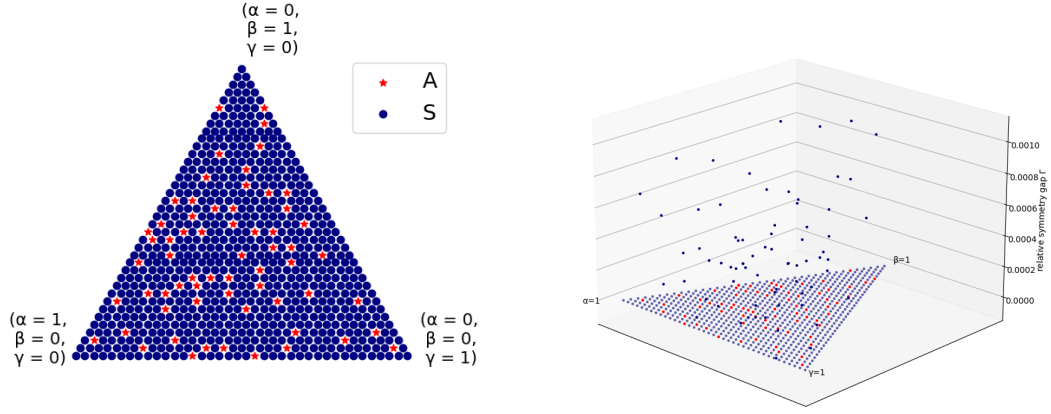
(a) Classification of optimal solutions for $n = 8, Y = 24000, g = 1/3, a = 0.8$ and $\mu \approx 0.878$ corresponding to Fielbaum's results

(b) Relative gap for $n = 8, Y = 24000, g = 1/3, a = 0.8$ and $\mu \approx 0.878$ corresponding to Fielbaum's results

5.4 Computation Time

The computation time needed to solve one instance of the Parametric City varies, depending on the input and the model. Table 3 reports the average, minimal, and maximal time needed to compute one of the batches presented above, namely, for the 741 instances corresponding to Figure 7d. All computations were performed on a 3.50GHz Intel(R) Xeon(R) CPU E3-1245 v5 machine running Linux. All instances could be solved fairly quickly: On average, $ALPP$ was solved in approximately four seconds, but it could take up to 6.25 minutes. However, this was significantly larger than the time needed to compute the symmetric model. $ALPP_{\mathcal{S}}$ terminated for all instances in less than half a second. Even if these computation times are small, the differences are striking: On average, solving $ALPP$ took approximately 145 times longer than solving $ALPP_{\mathcal{S}}$.

	$ALPP$	$ALPP_{\mathcal{S}}$
Average	4.022	0.028
Minimum	0.027	0.022
Maximum	374.686	0.039

Table 3: Computation times in seconds: symmetric model $ALPP_{\mathcal{S}}$ in comparison to the general model $ALPP$ for 741 instances of the Parametric City (cf. Figure 7d)

6 Conclusions

We investigated the line planning problem in the Parametric City, which combines a versatile generic city planning paradigm with a flexible optimization model allowing the construction of arbitrary lines. Although the Parametric City is inherently rotation symmetric with respect to both shape and demand, it turns out that optimal line plans are not necessarily symmetric. However, the differences between symmetric and asymmetric optimal solutions are rather small in practice, while in theory an approximation algorithm can derived (fixing one of the parameters), building upon the polynomial-time solvability of the line planning problem in Parametric City when line frequencies have to obey rotation symmetry. From a planners' perspective, restricting to symmetric line plans is hence a preeminent starting point to design a transportation network.

References

Assad, A. A. (1980). Modelling of rail networks: Toward a routing/makeup model. *Transportation Research Part B: Methodological*, 14(1):101 – 114.

Badia, H. (2020). Comparison of bus network structures in face of urban dispersion for a ring-radial city. *Networks and Spatial Economics*, 20:233–271.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132.

Bussieck, M., Winter, T., and Zimmermann, U. (1997). Discrete optimization in public rail transport. *Mathematical Programming*, 79:415–444.

Byrne, B. F. (1975). Public transportation line positions and headways for minimum user and system cost in a radial case. *Transportation Research*, 9(2):97–102.

Ceder, A. and Wilson, N. H. (1986). Bus network design. *Transportation Research Part B: Methodological*, 20(4):331 – 344.

Daganzo, C. F. (2010). Structure of competitive transit networks. *Transportation Research Part B: Methodological*, 44(4):434–446.

Fielbaum, A., Jara-Díaz, S., and Gschwender, A. (2016a). Optimal public transport networks for a general urban structure. *Transportation Research Part B: Methodological*, 94:298 – 313.

Fielbaum, A., Jara-Díaz, S., and Gschwender, A. (2016b). A parametric description of cities for the normative analysis of transport systems. *Networks and Spatial Economics*, 17.

Fielbaum, A., Jara-Díaz, S., and Gschwender, A. (2018). Transit line structures in a general parametric city: The role of heuristics. *Transportation Science*, 52.

Fielbaum Schnitzler, A. S. (2019). *Effects of the Introduction of Spatial and Temporal Complexity on the Optimal Design, Economies of Scale and Pricing of Public Transport*. PhD thesis, Universidad de Chile.

Gallian, J. (2000). A dynamic survey of graph labeling. *Electron J Combin DS6*, 19.

Gurobi Optimization, LLC (2020). Gurobi optimizer reference manual.

Holroyd, E. (1967). The optimum bus service: A theoretical model for a large uniform urban area. *Proceedings of the 3rd International Symposium on the Theory of Road Traffic Flow*, pages 308–328.

Jansson, J. O. (1980). A simple bus line model for optimisation of service frequency and bus size. *Journal of Transport Economics and Policy*, 14(1):53–80.

Jara-Díaz, S. and Gschwender, A. (2009). The effect of financial constraints on the optimal design of public transport services. *Transportation*, 36:65–75.

Karbstein, M. (2013). *Line Planning and Connectivity*. PhD thesis, Technische Universität Berlin.

Kepaptsoglou, K. and Karlaftis, M. (2009). Transit route network design problem: Review. *Journal of Transportation Engineering*, 135(8):491–505.

Lenstra, H. W. (1983). Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):538–548.

Lopez, F. (2014). Integrating network design and frequency setting in public transportation networks: A survey. *SORT (Statistics and Operations Research Transactions)*, 38:181–214.

Schmidt, M. E. (2014). *Integrating routing decisions in public transportation problems*. Springer.

Schöbel, A. (2011). Line planning in public transportation: Models and methods. *OR Spektrum*, 34:1–20.

A Appendix

Proof of Proposition 4.5. The cost values correspond to those of the uncapacitated minimum-cost flow problem corresponding to Figure 4b. For better readability, let us summarize arc costs of the same type:

$$\bar{c}_{ps} := \bar{c}_{(P_i, SC_i)} = (2\mu/K + (1-\mu))T g,$$
$$\bar{c}_{sp} := \bar{c}_{(SC_i, P_i)} = (1-\mu)T g,$$
$$\bar{c}_{ss} := \bar{c}_{(SC_i, SC_{i\pm 1})} = (\mu/K + (1-\mu))T r_n,$$
$$\bar{c}_{sc} := \bar{c}_{(SC_i, CD)} = (2\mu/K + (1-\mu))T,$$
$$\bar{c}_{cs} := \bar{c}_{(CD, SC_i)} = (1-\mu)T$$

for all $i \in \mathbb{Z}/n\mathbb{Z}$.

It is obvious that for an OD pair $(SC_i, SC_{i\pm j})$ the shortest path is either via the central business district (SC_i, CD, SC_j), which means that it contributes $2T(\mu/K + (1-\mu))$ to the objective value, or along the subcenter ring – $(SC_i, SC_{i+1}, \ldots, SC_{i\pm j})$ – with cost $jTr_n(\mu/K + (1-\mu))$. Thus, for $j \in \{1, \ldots, k_n\}$ with $k_n := \left\lfloor \frac{2}{r_n} \right\rfloor$ the shortest path from SC_i to SC_{i+j} leads over neighboring subcenters, while the rest pass through CD. The same holds for the OD-pairs $(P_i, SC_{i\pm j})$. An overview of the shortest paths is depicted in Figure 4a, with corresponding costs $\bar{c}_{s\to t}$ in Table 4.

s	t	shortest path	cost $\bar{c}_{s\to t}$
P_i	SC_i	(P_i, SC_i)	$\bar{c}_{ps}$
P_i	$SC_{i\pm j}, j \in \{1, \ldots, k_n\}$	$(P_i, SC_i, SC_{i\pm 1}, \ldots, SC_{i\pm j})$	$\bar{c}_{ps} + j\bar{c}_{ss}$
P_i	$SC_{i\pm j}, j \in \{k_n, \ldots, \lfloor\frac{n}{2}\rfloor\}$	$(P_i, SC_i, CD, SC_{i\pm j})$	$\bar{c}_{ps} + \bar{c}_{cs} + \bar{c}_{ss}$
P_i	CD	(P_i, SC_i, CD)	$\bar{c}_{ps} + \bar{c}_{cs}$
SC_i	$SC_{i\pm j}, j \in \{1, \ldots, k_n\}$	$(SC_i, SC_{i\pm 1}, \ldots, SC_{i\pm j})$	$j\bar{c}_{ss}$
SC_i	$SC_{i\pm j}, j \in \{k_n, \ldots, \lfloor\frac{n}{2}\rfloor\}$	$(SC_i, CD, SC_{i\pm j})$	$\bar{c}_{sc} + \bar{c}_{cs}$
SC_i	CD	(SC_i, CD)	$\bar{c}_{cs}$

Table 4: Shortest paths for each origin-destination pair $(s, t) \in D$

First, consider only the cost of passengers traveling from SC_0 to any of the other subcenters:

$$\sum_{j=1}^{n} \bar{c}_{SC_0 \to SC_j} \, d_{SC_0,SC_j} = 2 \sum_{j=1}^{k_n} j \, \bar{c}_{ss} \, d_{SC_0,SC_j} + \sum_{j=k_n+1}^{n-(k_n+1)} (\bar{c}_{sc} + \bar{c}_{cs}) \, d_{SC_0,SC_j}$$

$$= 2 \frac{k_n(k_n+1)}{2} \bar{c}_{ss} \, d_{SC_0,SC_j} + (n - 2k_n + 1)(\bar{c}_{sc} + \bar{c}_{cs}) \, d_{SC_0,SC_1}$$

$$= \left(k_n(k_n+1)\bar{c}_{ss} + (n - 2k_n + 1)(\bar{c}_{sc} + \bar{c}_{cs}) \right) d_{SC_0,SC_1}.$$

The shortest path costs for passengers starting at one periphery to all foreign subcenters can be calculated similarly. Since each path has to take arc (P_0, SC_0), the term $\bar{c}_{ps}$ needs to be added for each of the $n-1$ paths. Thus we get

$$\sum_{j=1}^{n} \bar{c}_{P_0 \to SC_j} \, d_{P_0,SC_j} = \left(k_n(k_n+1)\bar{c}_{ss} + (n - 2k_n + 1)(\bar{c}_{sc} + \bar{c}_{cs}) + \bar{c}_{ps}(n-1) \right) d_{P_0,SC_1}.$$

Passengers only start in either a periphery or a subcenter, none originate at the center. Thus one can calculate the explicit costs of a single zone simply by adding these two terms and inserting the values for the demand d :

$$\sum_{v \in V} d_{P_0,v} \bar{c}_{P_0 \to v} + \sum_{v \in V} d_{S_0,v} \bar{c}_{SC_0 \to v}$$

$$= T \frac{Y}{n} \left(\left(k_n(k_n+1)r_n + 2(n - 2k_n + 1) \right) \left(\frac{\mu}{K} + (1-\mu) \right) \left(\frac{a\gamma + (1-a)\tilde{\gamma}}{(n-1)} \right) \right.$$

$$\left. + \left(\frac{2\mu}{K} + (1-\mu) \right) (a\alpha + (1-a)\tilde{\alpha}) + \left(\frac{2\mu}{K} + (1-\mu) \right) g\, a \right)$$

$$= T \frac{Y}{n} \lambda(\alpha, \gamma).$$

This implies that the cost for zone 0 depends linearly on Y/n – the total amount of people from a single zone. As the costs are the same for each zone, the explicit optimal value of the shortest path problem must be n times the cost of a single zone, where each passenger takes the shortest path:

$$cost_{UMCFP}(y) = \sum_{p \in P} \bar{c}_p y_p = T Y \lambda(\alpha, \gamma). \qquad \square$$

Proof of Lemma 4.6. Let $RLPP$ denote the described relaxation of $ALPP$, i.e.,

$$(RLPP) \quad \min_{F,y} \quad \sum_{a \in A} \tau_a F_a + \sum_{p \in P} \tau_p y_p = cost_{RLPP}(F, y)$$

$$\text{s.t.} \quad \sum_{p \in P_{st}} y_p = d_{s,t} \qquad \forall (s,t) \in D \tag{9a}$$

$$\sum_{p \in P_a} y_p - K F_a \geq 0 \quad \forall a \in A \tag{9b}$$

$$y_p \geq 0 \qquad \forall p \in P \tag{9c}$$

$$F_a \geq 0 \qquad \forall a \in A \tag{9d}$$

Suppose we have a solution (F,y) to $RLPP$. Clearly, y is feasible for $UMCFP$ as well. What is left to show is that $\text{OptVal}(UMCFP) \leq \text{OptVal}(RLPP)$. We know that $F_{(P_i,SC_i)} = F_{(SC_i,P_i)}$; with Condition 4c we get

$$\sum_{p \in P_{(P_i,SC_i)}} \frac{y_p}{K} \leq F_{(P_i,SC_i)} = F_{(SC_i,P_i)} \tag{10}$$

for all $i \in \mathbb{Z}/n\mathbb{Z}$. Similarly, $\sum_{i=0}^{n-1} F_{(SC_i,CD)} = \sum_{i=0}^{n-1} F_{(CD,SC_i)}$ has to hold. This results in

$$\max \left\{ \sum_{i=0}^{n-1} \sum_{p \in P_{(SC_i,CD)}} \frac{y_p}{K}, \sum_{i=0}^{n-1} \sum_{p \in P_{(CD,SC_i)}} \frac{y_p}{K} \right\} \leq \sum_{i=0}^{n-1} F_{(SC_i,CD)} = \sum_{i=0}^{n-1} F_{(CD,SC_i)} \tag{11}$$

in order to fulfill the capacity constraints.

For the subcenter arc-frequencies, we obtain analogously:

$$\sum_{p \in P_{(SC_i,SC_{i\pm1})}} \frac{y_p}{K} \leq F_{(SC_i,SC_{i\pm1})}. \tag{12}$$

Note, however, that both directions need to be considered separately.

By applying inequalities (10), (11) and (12) to each of the corresponding arc-frequencies, as well as inserting the values of the arc-costs, we can estimate $\sum_{a \in A} \tau_a F_a$ as

$$\sum_{a \in A} \tau_a F_a \geq 2 T g \sum_{i=0}^{n-1} \sum_{p \in P_{(P_i,SC_i)}} \frac{y_p}{K} + 2 T \sum_{i=0}^{n-1} \sum_{p \in P_{(SC_i,CD)}} \frac{y_p}{K}$$

$$+ T r_n \sum_{i=0}^{n-1} \sum_{p \in P_{(SC_{(i+1)},SC_i)}} \frac{y_p}{K} + T r_n \sum_{i=0}^{n-1} \sum_{p \in P_{(SC_{(i-1)},SC_i)}} \frac{y_p}{K}.$$

By plugging this into the cost function of *RLPP* and some simple rearrangement, we have

$$cost_{RLPP}(F,y) = \mu \sum_{a \in A} \tau_a F_a + (1 - \mu) \sum_{p \in P} \tau_p y_p = \sum_{a \in A} \sum_{p \in P_a} \bar{c}_a y_p = \sum_{p \in P} \bar{c}_p y_p = cost_{UMCFP}(y),$$

from which the claim follows. $\qquad\square$

Proof of Corollary 4.8. Consider the formula for the optimal value of the uncapacitated minimum-cost flow problem $\mathrm{OptVal}(UMCFP) = TY\lambda(\alpha, \gamma)$ from Proposition 4.5. We will analyze the term

$$\frac{k_n(k_n + 1)r_n + 2(n - 2k_n + 1)}{(n - 1)}$$

and show that $\frac{k_n(k_n+1)r_n}{(n-1)} \in [\frac{2}{\pi}, 2]$ and $\frac{n-2k_n+1}{n-1} \in [1 - \frac{2}{\pi}, 1]$ for all $n \geq 4$: As $r_n T$ is the distance of two subcenters and $k_n = \left\lfloor \frac{2}{r_n} \right\rfloor$, i.e., the largest number such that the length of k_n subcenter-arcs is shorter than two central arcs, we have $k_n r_n \leq 2$ and $(k_n + 1)r_n > 2$. Further, we have the bound $r_n n \leq 2\pi$, because r_n is the side length of an n-gon with vertices on the unit circle, thus the perimeter is less than 2π. By a similar argument we can inscribe a circle of radius $\cos\left(\frac{\pi}{n}\right)$ within that regular n-gon such that the sides of the polygon touch the circle tangentially. The perimeter of the n-gon must therefore be larger than the circumference of the inscribed circle. As the radius of the inscribed circle increases with n, its minimum is at $n = 4$, thus we have

$$nr_n > 2\pi \cos\left(\frac{\pi}{n}\right) \geq 2\pi \cos\left(\frac{\pi}{4}\right) \geq \sqrt{2}\pi.$$

Further, we use the fact that $n - 2k_n + 1 \leq n - 1$ and $k_n + 1 \leq n - 1$. Therefore, for $n \geq 4$ we have the three estimates

$$\frac{k_n(k_n + 1)r_n}{n - 1} \leq \frac{2(k_n + 1)}{n - 1} \leq 2,$$

$$\frac{k_n(k_n + 1)r_n}{n - 1} \geq \frac{2k_n}{n - 1} \geq \frac{2k_n}{2\pi/r_{n-1}} \geq \frac{r_{n-1}k_n}{\pi} \geq \frac{r_{n-1}k_{n-1}}{\pi} \geq \frac{2}{\pi},$$

$$\frac{n - 2k_n + 1}{n - 1} \leq 1.$$

For the lower bound on $\frac{n-2k_n+1}{n-1}$ we need to examine the term in more detail. We can derive the following inequality by making use of the property $nr_n > 2\pi \cos\left(\frac{\pi}{n}\right)$:

$$\frac{n-2k_n+1}{n-1} \geq \frac{n-2k_n+1}{n} = \frac{n-2\left\lfloor\frac{2}{r_n}\right\rfloor+1}{n} \geq 1 - \frac{4}{r_n n} + \frac{1}{n} \geq 1 - \frac{2}{\pi\cos\left(\frac{\pi}{n}\right)} + \frac{1}{n}.$$

For the last term, define the help function $h : [4,\infty[\to \mathbb{R}$ with $h(x) = 1 - \frac{2}{\pi\cos\left(\frac{\pi}{x}\right)} + \frac{1}{x}$. It is possible to show that h is monotonically decreasing on the interval $[8,\infty[$: The function is differentiable with the first derivative

$$\frac{d}{dx}h(x) = \frac{1}{x^2}\left(\frac{2\sin\left(\frac{\pi}{x}\right)}{\cos^2\left(\frac{\pi}{x}\right)} - 1\right).$$

On the interval $[8,\infty[$, both $\sin\left(\frac{\pi}{x}\right)$ and $\cos\left(\frac{\pi}{x}\right)$ are positive; the former is monotonically decreasing, while the latter (and thus $\cos^2\left(\frac{\pi}{x}\right)$ as well) is monotonically increasing. Hence, the following holds true

$$\frac{2\sin\left(\frac{\pi}{x}\right)}{\cos^2\left(\frac{\pi}{x}\right)} < \frac{2\sin\left(\frac{\pi}{8}\right)}{\cos^2\left(\frac{\pi}{8}\right)} = \frac{2\sqrt{2-\sqrt{2}}}{2+\sqrt{2}} \approx 0.448 < 1$$

for all $x \geq 8$. This implies that on the interval $[8,\infty[$, h is monotonically decreasing since its derivative is negative. It is clear that the limit exists with

$$\lim_{x\to\infty} h(x) = 1 - \frac{2}{\pi}.$$

Thus we have $\frac{n-2k_n+1}{n-1} \geq 1 - \frac{2}{\pi}$ for all $n \in \mathbb{N}_{\geq 8}$. For $n \in \{4,5,6,7\}$ the bound can be verified easily by plugging in the values $k_4 = k_5 = 1$ and $k_6 = k_7 = 2$. Together, the following inequality holds for all $n \in \mathbb{N}_{\geq 4}$:

$$2 - \frac{2}{\pi} \leq \frac{k_n(k_n+1)r_n + 2(n-2k_n+1)}{n-1} \leq 4.$$

Finally, we can estimate $\lambda(\alpha,\gamma)$ by making use of the properties $\alpha + \beta + \gamma = 1$ and $\tilde{\alpha} + \tilde{\gamma} = 1$:

$$\lambda(\alpha,\gamma) \leq 4\left(\frac{\mu}{K} + (1-\mu)\right)(a\gamma + (1-a)\tilde{\gamma})$$
$$+ \left(\frac{2\mu}{K} + (1-\mu)\right)(a\alpha + (1-a)\tilde{\alpha}) + \left(\frac{2\mu}{K} + (1-\mu)\right)ga$$
$$\leq 4\left(\frac{2\mu}{K} + (1-\mu)\right)(1+ga)$$

and

$$\lambda(\alpha,\gamma) \geq \left(2 - \frac{2}{\pi}\right)\left(\frac{\mu}{K} + (1-\mu)\right)(a\gamma + (1-a)\tilde{\gamma})$$
$$+ \left(\frac{2\mu}{K} + (1-\mu)\right)(a\alpha + (1-a)\tilde{\alpha}) + \left(\frac{2\mu}{K} + (1-\mu)\right)ga$$
$$\geq \left(\left(2 - \frac{2}{\pi}\right)\frac{\mu}{K} + (1-\mu)\right)(1+ga-a),$$

from which the claim follows. $\qquad\square$

B List of Symbols

Symbol	Description	Page
A	set of arcs in the Parametric City	3
a	fraction of people starting in the peripheries	4
$ALPP$	arc-based line planning problem	5
CD	central business district	3
τ_a, τ_l, τ_p	cost of/time to travel along arc a, line l, or path p	4,5,5
$\bar{c}_a, \bar{c}_p$	cost parameters of $UMCFP$	10
$d_{s,t}$	demand from node s to t	4
F	frequency variables of arcs of $ALPP$	5
F_C	frequency variable of $ALPP_S$ corresp. to central arcs	7
F_P	frequency variable of $ALPP_S$ corresp. to peripheral arcs	7
F_{S+}	frequency variable of $ALPP_S$ corresp. to arcs (SC_j, SC_{j+1})	7
F_{S-}	frequency variable of $ALPP_S$ corresp. to arcs (SC_j, SC_{j-1})	7
f	frequency variables of lines of LPP	4
$\mathcal{G} = (V, A)$	graph of the Parametric City	3
g	factor for distance between SC_j and P_j	3
K	vehicle capacity for number of passengers	5
k_n	largest value s.t. the shortest SC_0-SC_{k_n}-path is along subcenters	10
L	set of lines corresponding to directed circuits	4
L_a	set of lines using arc a	4
LPP	line-based problem	5
n	number of zones in the Parametric City	3
$\mathrm{OptVal}(P)$	optimal value of problem P	9
P	set of (passenger) paths corresponding to directed simple paths	4
P_a	set of paths using arc a	4
$P_{s \to t}$	set of paths with origin s and destination t	4
P_j	periphery of zone j	3
r_n	factor for distance between two subcenters	4
SC_j	subcenter of zone j	3
T	distance between CD and a subcenter	3
$UMCFP$	uncapacitated minimum-cost flow problem as relaxation of $ALPP$	10
V	set of nodes in the Parametric City	3
Y	total patronage	4
y	passenger variables of $LPP, ALPP, ALPP_S, UMCFP$	4
α	fraction of people traveling from P_j to CD	4
$\tilde{\alpha}$	fraction of people traveling from SC_j to CD	4
β	fraction of people traveling from P_j to SC_j	4
γ	fraction of people traveling from P_j to other subcenters	4
$\tilde{\gamma}$	fraction of people traveling from SC_j to other subcenters	4
$\delta^+(v), \delta^-(v)$	set of outgoing, incoming arcs at node v	6
Λ	frequency capacity for arcs	5
$\lambda(\alpha, \gamma)$	scaled optimal value of $UMCFP$	10
μ	factor to weigh operator against user costs	5
	$\mu = 0$ means no operator costs, $\mu = 1$ no user costs	
$\rho_z(x)$	rotation of vertex tuple x by $z\,2\pi/n$	6
$a \in p, a \in l$	path p, line l uses arc a	5