Worst-case demand distributions in vehicle routing

John Gunnar Carlsson*and Mehdi Behroozi†

March 29, 2016

Abstract

A recent focal point in research on the vehicle routing problem (VRP) is the issue of robustness in which customer demand is uncertain. In this paper, we conduct a theoretical analysis of the demand distributions whose induced workloads are as undesirable as possible. We study two common variations of VRP in a continuous approximation setting: the first is the VRP with time windows, and the second is the capacitated VRP, in which regular returns to the vehicle's point of origin are required.

1 Introduction

Since its original formulation in 1959, two of the primary features that have distinguished the *vehicle routing problem* (VRP) from the *travelling salesman problem* (TSP) have been the introduction of *capacities* on vehicles that originate from a central depot [27] and the imposition of *time windows* that constrain the times when customers can be visited [45]. Not surprisingly, the imposition of such constraints presents a major obstacle in obtaining solutions to a problem instance, both in terms of the added computational burden of finding solutions and in the overall quality of the solution itself. As identified in [15], one useful feature of the capacitated VRP is that one can actually describe the additional cost somewhat concretely:

> Any solution for the capacitated VRP has two cost components; the first component is proportional to the total "radial" cost between the depot and the customers. The second component is proportional to the "circular" cost; the cost of traveling between customers. This cost is related to the cost of the optimal traveling salesman tour. It is well known [11] that, for large N, the cost of the optimal traveling salesman tour grows like $\sqrt{N}$, while the total radial cost between the depot and the customers grows

like N.... Therefore, it is intuitive that when the number of customers is large enough the first cost component will dominate the optimal solution value.

The additional cost due to time windows is more difficult to quantify, although insights can be made under certain assumptions [25]:

Imagine an extreme case, where only a tiny fraction of the customers have very stringent...time window constraints. Because, as we shall see, the distance travelled increases with [the square root of the number of time windows], the total system cost may be large because of the requirements of very few customers.

This paper addresses VRP from the perspective of a *continuous approximation model*: we assume that a fleet of vehicles must provide service to a contiguous planar geographic region, and our goal is to quantify precisely the role that vehicle capacities and time windows play in the worst-case workloads of the vehicles. We assume that customer demands are independently sampled from a (possibly unknown) demand distribution, and study the asymptotic behavior of the worst-case distributions as the number of customers becomes large. In this sense, our paper is philosophically similar to (for example) [18], which analytically determines trade-offs between transportation and inventory costs, [36], which shows how to route emergency relief vehicles to beneficiaries in a time-sensitive manner, and [38], which describes a simple geometric model for determining the optimal mixture of a fleet of vehicles that perform distribution. The basic premise of the continuous approximation paradigm is that one replaces combinatorial quantities that are difficult to compute with simpler mathematical formulas, which (under certain conditions) provide accurate estimations of the desired quantity[20, 30, 44, 46]. Such approximations exist for many combinatorial problems, such as the travelling salesman problem [11, 28], facility location [33, 35, 47], and any *subadditive Euclidean functional* such as a minimum spanning tree, Steiner tree, or matching [48, 54, 55]. In our computational districting experiment, an approximation of this kind is used as the first level of an optimization problem in which we design service zones that are associated with different vehicles. for example. Our study of the VRP with time windows adopts similar assumptions to those of [25, 26], namely, that the service period is divided into a collection of pre-specified intervals of equal duration. One might contrast this model with other approaches like [29], which assumes that time windows are independently drawn from an arbitrary probability measure. Our study of the capacitated VRP makes extensive use of upper and lower bounds derived in [16, 24, 34], as well as seminal results on the TSP that can be found in [11, 48, 54].

A more recent focal point in research on VRP and its variants is the isssue of *robustness* in which one seeks a policy that performs as well as possible against all possible realizations of demand that are compatible with some set of observations or initial conditions. Robust methodologies for the capacitated VRP were first introduced in the

paper [57], which adapts the methodology of [13] to solve problems in which customer demands and travel times are uncertain; the goal is to find vehicle routes that meet all feasibility requirements in the worst-case scenario, which occurs precisely when all customer demands and travel times attain their worst-case realizations simultaneously. In most models of the robust VRP, one has a pre-defined ambiguity region and seeks a set of routes that is as good as possible with respect to all of the outcomes; this ambiguity region is usually described as a finite collection of scenarios or a polyhedral set [5, 10, 32, 41, 53, 57], although the recent paper [6] adopts a "robust mean-variance" approach that minimizes a weighted sum of the average cost and the variance of a route when sampled over many scenarios. In our problem, we are concerned with robustness in the *distributional* sense [19]: we seek the distribution of demand for which the expected cost of a tour is as high as possible, while remaining consistent with some observed data samples or some parameters derived thereof. The most closely related result to our paper is [21], which determines the worst-case spatial demand distribution for the TSP when the first and second moments are fixed. Our paper can be seen as a generalization of these principles to the cases where vehicles have capacities and time window constraints.

Our present work uses the notion of robustness to study the negative consequences of fluctuation in demand for delivery services, in either a spatial or temporal sense. Demand fluctuation is of particular concern for emerging delivery services such as Good Eggs, DoorDash, BiteSquad, and Caviar [1, 2, 3, 4], which face extremely high volatility in demand due to seasonality and the time-sensitive nature of the requests they satisfy [37, 42]:

> Our business is highly dependent on diner behavior patterns that we have observed over time. In our metropolitan markets, we generally experience a relative increase in diner activity from September to April and a relative decrease in diner activity from May to August. In addition, we benefit from increased order volume in our campus markets when school is in session and experience a decrease in order volume when school is not in session, during summer breaks and other vacation periods. Diner activity can also be impacted by colder or more inclement weather, which typically increases order volume, and warmer or sunny weather, which typically decreases order volume. Seasonality will likely cause fluctuations in our financial results on a quarterly basis. In addition, other seasonality trends may develop and the existing seasonality and diner behavior that we experience may change or become more extreme.

In total, this paper makes the following contributions: Section 3 analyzes the vehicle routing problem with time windows, characterizing the worst-case distributions that can arise when demand varies over a specified time horizon. Section 4 deals with the capacitated VRP, and Section 5 extends this analysis to more sophisticated models in which we have information about the mean or covariance of the demand distribution and describes some computational

experiments.

2 Preliminaries

We make the following notational conventions in this paper: given a point set X, the *star network* of X is written $\mathsf{SN}(X)$ and consists of the network in which each point in X is connected to some central "depot" point (the location of this central point will be made clear from context). Vehicle capacities are either denoted by the letter c, indicating that a vehicle can visit c destinations before returning to its depot, or by the capacity coefficient t, which satisfies the relationship $c = t\sqrt{|X|}$; this is a standard and useful representation, as can be seen in Section 4.2 of [23] or the paper [24]. A TSP tour of a set of points will be denoted by $\mathsf{TSP}(X)$. A capacitated VRP tour of a set of points is written $\mathsf{VRP}(X)$, where we suppress the capacity in the interest of notational brevity. Finally, we say that $f(x) \in o(g(x))$ if $\lim_{x \to \infty} f(x)/g(x) = 0$, we say that $f(x) \sim g(x)$ if $\lim_{x \to \infty} f(x)/g(x) = 1$, we say that $f(x) \in \mathcal{O}(g(x))$ if $f(x) \le \alpha g(x)$ for some $\alpha > 0$ and all x exceeding some threshold x_0, and we say that $f(x) \in \Omega(g(x))$ if $f(x) \ge \alpha g(x)$ for all x exceeding some threshold x_0.

To approximate the length of a TSP tour of a collection of points, we will use the well-known *BHH Theorem* [11], which says that the length of an optimal TSP tour of a set of points follows a law of large numbers:

Theorem 1. *Suppose that $X = \{X_1, X_2, \dots\}$ is a sequence of random points i.i.d. according to a probability density function $f(\cdot)$ defined on a compact planar region $\mathcal{R}$. Then with probability one, the length of $\mathsf{TSP}(X)$ satisfies*

$$\lim_{N \to \infty} \frac{\text{length}(\mathsf{TSP}(X))}{\sqrt{N}} = \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)} \, dA$$

where β is a constant and $\bar{f}(\cdot)$ represents the absolutely continuous part of $f(\cdot)$.

It is additionally known that $0.6250 \le \beta \le 0.9204$ and estimated that $\beta \approx 0.7124$; see [8, 11]. Theorem 1 can also be expressed deterministically, removing any assumptions about the distribution of the points X_i; see for example [31, 39]:

Theorem 2. *There exists a constant α satisfying the following: if $X = \{X_1, X_2, \dots\}$ is any sequence of points contained in a compact planar region $\mathcal{R}$ with area 1, then*

$$\limsup_{N \to \infty} \frac{\mathsf{TSP}(X_1, \dots, X_N)}{\sqrt{N}} \le \alpha \, .$$

Furthermore, it is also true that $(4/3)^{1/4} \le \alpha < 1.392$.

4

The paper [52] actually proves a stronger result that these worst-case point sets must be uniformly distributed in an asymptotic sense.

The following result from [34] gives upper and lower bounds for a capacitated VRP tour:

Theorem 3. *For any set X of demand points serviced by a fleet of vehicles with capacity c that originate from a single depot, we have*

$$\max\left\{\frac{2}{c}\operatorname{length}(\mathsf{SN}(X)), \operatorname{length}(\mathsf{TSP}(X))\right\} \leq \operatorname{length}(\mathsf{VRP}(X)) \leq 2\left\lceil\frac{|X|}{c}\right\rceil \cdot \frac{\operatorname{length}(\mathsf{SN}(X))}{|X|} + (1 - 1/c)\operatorname{length}(\mathsf{TSP}(X)).$$
$$(1)$$

3 Worst-case temporal demand distributions

In this section, we will study the worst-case *temporal* distributions for a particular version of the vehicle routing problem with time windows. In order to make this problem tractable, we assume as in [25] that the service period is divided into m equally long time periods and that each customer is placed into one of these time periods, so that each period i has n_i customers. In order to isolate the temporal aspects of our problem, we will assume that all demand is uniformly distributed throughout the service region, which has area 1.

Since each period has a finite length, there is a maximum distance ℓ that each vehicle can travel within each time window (equal to the speed of the vehicle multiplied by the duration of the time window), which thereby induces a limit on the number of customers that can be visited in that time period. Typically, delivery services charge a flat fee for each delivery [1, 2], and thus the revenue from providing service to customers is simply linearly proportional to the number of customers. In order to simplify our analysis, we will assume that one vehicle, capable of traversing a distance ℓ in each time period, provides service to the n_i customers (or as many of the n_i customers as possible) in each time period i. We assume that there is a fixed price p that the company charges for each order and that there is a price q that the company incurs per unit distance of travel. Thus, if the length of the TSP tour of these points $X_1, \ldots, X_{n_i}$ is less than ℓ, then the company receives a net profit of

$$pn_i - q\operatorname{length}(\mathsf{TSP}(X_1, \ldots, X_{n_i}))$$

in that time period. By applying Theorem 1, we can approximate $\operatorname{length}(\mathsf{TSP}(X_1, \ldots, X_{n_i})) \approx \beta\sqrt{n_i}$. If the length of the TSP tour exceeds ℓ – or equivalently, if $\beta\sqrt{n_i} > \ell$ – then the company's goal is to make as much money as possible (i.e. visit as many customers as possible), without violating this distance constraint. In order to study this scenario, the following theorem is useful:

5

Theorem 4. *Suppose that $\{X_1, \ldots, X_n\}$ are uniformly distributed in the unit square and let $\ell > 0$. Let N_ℓ denote the maximum number of points X_i that can be visited with a path of length ℓ. Then $\mathbf{E}(N_\ell) \in \Omega(\ell\sqrt{n})$ and $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ as $n \to \infty$. In particular, we have $\mathbf{E}(N_\ell) \geq \alpha^{-1}\ell\sqrt{n}$ and $\mathbf{E}(N_\ell) \leq 3.41\ell\sqrt{n}$ for sufficiently large n, where α is the constant from Theorem 2.*

Before proving this result, the following lemma is helpful:

Lemma 5. *Suppose that $\{X_1, \ldots, X_n\}$ are uniformly distributed in the unit square and let $\ell > 0$. Then*

$$\Pr(\mathrm{length}(\mathsf{TSP}(X_1, \ldots, X_n)) \leq \ell) \leq \frac{n!}{(2n-2)!}(2\pi\ell^2)^{n-1}$$

as $n \to \infty$.

Proof. See Section A of the online supplement. $\qquad\square$

The proof of Theorem 4 follows:

Proof of Theorem 4. The fact that $\mathbf{E}(N_\ell) \in \Omega(\ell\sqrt{n})$ is straightforward: simply take a TSP tour of all n points, whose length L satisfies $L \leq \alpha\sqrt{n}$ as in Theorem 2. If we divide this tour into segments of length ℓ, we are then left with at most $\lceil \alpha\sqrt{N}/\ell \rceil$ segments which cumulatively touch all n points. If we select one of these segments uniformly at random, the expected number of points in the segment is at least $n/\lceil \alpha\sqrt{n}/\ell \rceil \sim \alpha^{-1}\ell\sqrt{n}$, as desired.

The proof that $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ is trickier: using Lemma 5, and letting $c = 3.41 > e\sqrt{\pi/2}$, we can write

$$
\begin{aligned}
\mathbf{E}(N_\ell) \quad &\leq \quad c\ell\sqrt{n} + n\Pr(N_\ell \geq c\ell\sqrt{n}) \\
&= \quad c\ell\sqrt{n} + n\Pr(\text{there exists a path through } \lceil c\ell\sqrt{n} \rceil \text{ of the } n \text{ points with length} \leq \ell) \\
&\leq \quad c\ell\sqrt{n} + np\binom{n}{\lceil c\ell\sqrt{n} \rceil}
\end{aligned}
$$

by the union bound, where p is the probability that a uniformly distributed set of $\lceil c\ell\sqrt{n} \rceil$ points $\{X_1, \ldots, X_{\lceil c\ell\sqrt{n} \rceil}\}$ has a TSP tour whose length does not exceed ℓ. It will suffice to confirm that $np\binom{n}{\lceil c\ell\sqrt{n} \rceil} \to 0$ as $n \to \infty$. By
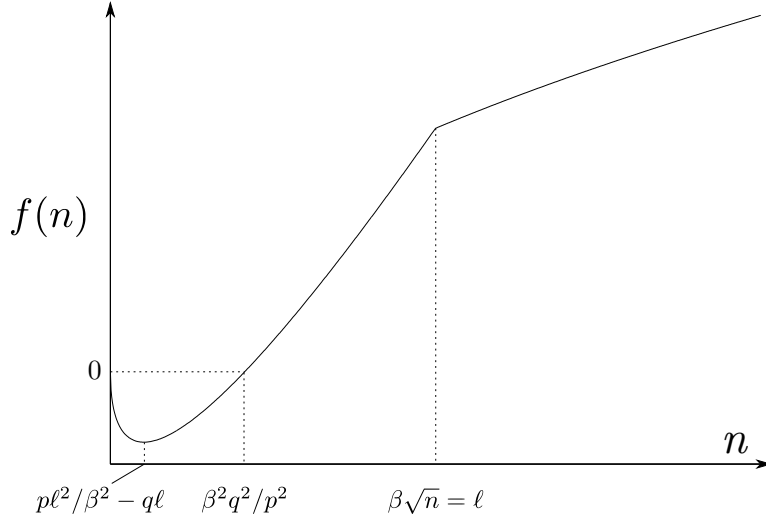
Figure 1: The profit function $f(n)$ as defined in (2).

Lemma 5, we can substitute for p, giving

$$
\begin{aligned}
np\binom{n}{\lceil c\ell\sqrt{n}\rceil} &\leq \frac{\lceil c\ell\sqrt{n}\rceil!}{(2\lceil c\ell\sqrt{n}\rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n}\rceil - 1}\cdot n\cdot\binom{n}{\lceil c\ell\sqrt{n}\rceil} \\
&\leq \frac{\lceil c\ell\sqrt{n}\rceil!}{(2\lceil c\ell\sqrt{n}\rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n}\rceil - 1}\cdot\frac{n^{\lceil c\ell\sqrt{n}\rceil + 1}}{\lceil c\ell\sqrt{n}\rceil!} \\
&\leq \frac{n^{\lceil c\ell\sqrt{n}\rceil + 1}}{(2\lceil c\ell\sqrt{n}\rceil - 2)!}(2\pi\ell^2)^{\lceil c\ell\sqrt{n}\rceil - 1} \\
&\leq \frac{n^2(2c\ell\sqrt{n} - 2)^{3/2}}{\sqrt{2\pi}}\cdot\left[\frac{2\pi e^2\ell^2 n}{(2c\ell\sqrt{n} - 2)^2}\right]^{c\ell\sqrt{n}},
\end{aligned}
$$

where we have used Stirling's approximation [51] in the last inequality. Since $c > e\sqrt{\pi/2}$, the bracketed quantity is eventually less than 1 for sufficiently large n, at which point the above quantity approaches 0 since the left term is polynomially increasing but the right term is super-polynomially decreasing. Thus, we conclude that $\mathbf{E}(N_\ell) \in \mathcal{O}(\ell\sqrt{n})$ as desired, which completes the proof. $\qquad\square$

The preceding theorem establishes that, if n points are sampled in a region with area 1 and $\ell < \beta\sqrt{n}$, then the maximum number of points that can be visited using a path of length ℓ is proportional to $\ell\sqrt{n}$. In order to impose continuity between the domains $\ell < \beta\sqrt{n}$ and $\ell \geq \beta\sqrt{n}$, we approximate this maximum number of points as $\ell\sqrt{n}/\beta$. Thus, we propose the following profit function when n points are sampled in a region with area 1 with p and q:

$$
f(n) := \begin{cases} pn - q\beta\sqrt{n} & \text{if } \beta\sqrt{n} \leq \ell \\ p\ell\sqrt{n}/\beta - q\ell & \text{otherwise .} \end{cases} \tag{2}
$$

7

The profit function $f(n)$ is shown in Figure 1, and is convex until $n = (\ell/\beta)^2$ (and concave thereafter), with a minimizer at $n = p\ell^2/\beta^2 - q\ell$. We will assume that $\frac{\ell p}{\beta^2 q} \geq 1$, which implies that $f(n) \geq 0$ when $\beta\sqrt{n} = \ell$ (in other words, the company is profitable when demand matches the limit ℓ exactly). If we assume that the aggregate demand over m time periods is a constant N, then the worst-case demand distribution over m time periods is then obtained by solving the optimization problem

$$\underset{n_1,\ldots,n_m}{\text{minimize}} \sum_{i=1}^{m} f(n_i) \qquad s.t. \tag{3}$$
$$\sum_{i=1}^{n} n_i = N$$

for all $n_i \geq 0$. The following theorem characterizes these worst-case distributions; we either have a large amount of demand concentrated in a single time period, or uniform demand throughout all time periods:

Theorem 6. *The optimal solution $n_1^*, \ldots, n_m^*$ to problem (3) takes one of two forms:*

1. *Either $n_i^* = N/m$ for all n_i^*, or*

2. *A single largest entry (say n_1^*) satisfies*
$$n_1^* = (z^*)^2 > \ell^2/\beta^2 \,,$$

   where z^* is a solution to the quartic equation

$$4\beta^2 p^2 z^4 - 4\beta\ell p^2 z^3 + (\ell^2 p^2 + \beta^4 m q^2 - \beta^4 q^2 - 4N\beta^2 p^2)z^2 + 4N\beta\ell p^2 z - N\ell^2 p^2 = 0 \,,$$

   and all other entries n_i^* satisfy
$$\frac{\beta^2 q^2}{4p^2} < n_i^* = \frac{N - n_1^*}{m-1} < \frac{\beta^2 q^2}{p^2} \,.$$

Furthermore, if
$$N \leq \frac{1}{4}\left(\frac{\ell}{\beta} + \frac{q\beta\sqrt{m}}{p}\right)^2 \,,$$

then the optimal solution takes the first form, and if

$$N \geq \left(\frac{\ell}{\beta} + \frac{q\beta\sqrt{m}}{p}\right)^2 \,,$$

the optimal solution takes the second form.

Proof. It will suffice to consider the equivalent problem

$$\underset{x_1,\ldots,x_m}{\text{minimize}}\, g(x_i) \qquad s.t.$$

$$\sum_{i=1}^{n} x_i \;=\; c$$

for all $x_i \geq 0$, where we define

$$g(x) := a \cdot \begin{cases} \frac{1}{b}\left(x - \sqrt{x}\right) & \text{if } x \leq b^2 \\[2mm] \sqrt{x} - 1 & \text{otherwise ;} \end{cases}$$

this is evident by applying the transformation $a \mapsto q\ell$, $b \mapsto \frac{p\ell}{q\beta^2}$, $c \mapsto \frac{p^2 N}{q^2\beta^2}$ and $x_i \mapsto \frac{p^2 n_i}{q^2\beta^2}$. The two cases described above are equivalent to

1a. Either $x_i^* = c/m$ for all x_i^*, or

2a. A single largest entry (say x_1^*) satisfies

$$x_1^* = (z^*)^2 > b^2 \,,$$

where z^* is a solution to the quartic equation

$$4z^4 - 4bz^3 + (b^2 + m - 4c - 1)z^2 + 4bcz - b^2 c = 0 \,,$$

and all other entries x_i^* satisfy

$$1/4 < x_i^* = (c - x_1^*)/(n-1) < 1.$$

Our assumption that $\frac{\ell p}{\beta^2 q} \geq 1$ is equivalent to an assumption that $b \geq 1$. First of all, it is obvious that there are no minimizers other than cases 1a and 2a; this is an immediate consequence of the fact that the function $g(\cdot)$ is convex on its first component and concave on its second. The quartic function in case 2a simply arises from equating derivatives in the first and second components, and the fact that $1/4 < x_i^* < 1$ (also in case 2a) is also implied by this (the derivative of the second component can only take values between 0 and $a/2b$). Our proof will be complete if we can show that case 1a is optimal when $c \leq \frac{1}{4}(b + \sqrt{m})^2$ and that case 2a is optimal when $c \geq (b + \sqrt{m})^2$. We will assume without loss of generality that $a = 1$.

If $c \leq \frac{1}{4}(b + \sqrt{m})^2$, then we will verify that case 1a is optimal by taking a lower bound of case 2a. Since case 2a must have $g(x_1^*) > b - 1$ and $g(x_i^*) > -1/4$ for all other i, a valid lower bound is simply $b - 1 + (m-1)(-1/4)$.

9

We also have $1/4 < (c - x_1^*)/(m-1) < (c - b^2)/(m-1)$, which we can rearrange for c to obtain

$$b^2 + n/4 - 1/4 < c \le \frac{1}{4}(b + \sqrt{n})^2 \,.$$

We now want to show that case 1a is optimal, i.e. that

$$[b - 1 + (n-1)(-1/4)] - \frac{1}{b}\left(c - \sqrt{cn}\right) \ge 0 \,.$$

The left-hand side of the above is concave in c and is therefore minimized at the endpoints $c = b^2 + m/4 - 1/4$ and $c = 1/4(b + \sqrt{m})^2$; it is routine to verify that the desired inequality holds by using the fact that $b > 0$ and $m \ge 2$.

If $c \ge (b + \sqrt{m})^2$, it will suffice to verify that case 1a is sub-optimal; this is immediate because we can simply set $x_1 = c$ and $x_i = 0$ otherwise, which would incur a cost of $\sqrt{c} - 1$, as opposed to a cost of $1/b(c - \sqrt{cm})$. We then want to show that

$$\frac{1}{b}(c - \sqrt{cm}) - (\sqrt{c} - 1) = 1 - (1 + \sqrt{m}/b)\sqrt{c} + c/b \ge 0 \,;$$

the left-hand side of the above is decreasing in m and is therefore minimized when m is as large as possible, namely when $c = (b + \sqrt{m})^2$, in which case the above expression is precisely equal to 1, which completes the proof. $\qquad\square$

Theorem 6 distinguishes between two worst-case demand distributions: either the workloads are uniform across time windows, or there is one time window that contains the vast majority of the work, with all other time windows being non-profitable. One can determine which of these distributions is worst via the quantity $N(\ell/\beta + q\beta\sqrt{m}/p)^{-2}$.

4 Worst-case scenarios for the capacitated VRP

In this section, we assume that a fleet of vehicles must originate from a single depot in a service region and visit a set of demand points X. These vehicles are *capacitated*, in the sense that they must return to the depot after visiting a specified number of destinations c (in other words, we assume that the "loads", so to speak, at each of the points $x_i \in X$ are the same). Our goal is to describe the *spatial* distributions of demand that maximize the workloads of the vehicles that provide service.

To begin this section, we re-phrase Theorem 3 in a probabilistic sense as follows: suppose that demand points $\{X_1, \ldots, X_N\}$ are independent samples from a probability distribution $f(\cdot)$ defined on a compact subset $\mathcal{R}$ of the plane. Assume without loss of generality that the single depot point is the origin $(0,0) \in \mathbb{R}^2$. It is then immediately clear that the average length of the star network $\mathsf{SN}(X)$ is simply $\mathbf{E}\operatorname{length}(\mathsf{SN}(X)) = N \iint_{\mathcal{R}} \|x\| f(x)\, dA$. We

assume, as explained in the introduction, that our capacity constraints take the form $c = t\sqrt{N}$. We will consider the expected length of the VRP tour of points that are sampled from $f(\cdot)$. By exchanging the expectation and $\max\{\cdot,\cdot\}$ operators, we can express the bound (1) as

$$\max\left\{\frac{2\sqrt{N}}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA,\ \beta\sqrt{N}\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right\} + o(\sqrt{N})$$
$$\leq\ \mathbf{E}\,\mathrm{length}(\mathsf{VRP}(X))$$
$$\leq\ 2\left\lceil\frac{\sqrt{N}}{t}\right\rceil\iint_{\mathcal{R}}\|x\|f(x)\,dA + \left(1-\frac{1}{\sqrt{N}\cdot t}\right)\beta\sqrt{N}\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA + o(\sqrt{N})\,.$$

Note that $\left\lceil\sqrt{N}/t\right\rceil$ is simply the number of vehicles needed to provide service. Since we are interested in the limiting behavior as $N\to\infty$, we can therefore write

$$\sqrt{N}\cdot\max\left\{\frac{2}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA,\ \beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right\} \lesssim \mathbf{E}\,\mathrm{length}(\mathsf{VRP}(X)) \lesssim \sqrt{N}\cdot\left(\frac{2}{t}\iint_{\mathcal{R}}\|x\|f(x)\,dA + \beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA\right) \tag{4}$$

where we have adopted the notation "$\lesssim$" to denote the "approximate" inequality, neglecting the lower-order terms (large values of N imply that $\left\lceil\sqrt{N}/t\right\rceil \approx \sqrt{N}/t$ and that $1/(\sqrt{N}\cdot t)\approx 0$). We are free to neglect the $o(\sqrt{N})$ terms since we are concerned with the worst-case demand distribution $f^*(\cdot)$ and we are simply assuming that N is large. Note that the two terms in the $\max\{\cdot\}$ operator on the left-hand side of the above represent the two sources of cost as identified in the introduction. If we search for distributions that maximize either of these expressions in isolation, the trade-off between these costs becomes clear: if we consider only the worst-case "radial" cost by maximizing $^2\!/_t\iint_{\mathcal{R}}\|x\|f(x)\,dA$, then the resulting distribution will simply be an atomic mass located at the farthest point from the depot, $\arg\max_{x\in\mathcal{R}}\|x\|$. Of course, since such a distribution has no continuous component whatsoever, the corresponding "circular" cost is zero. Similarly, if we consider only the worst-case "circular" cost by maximizing $\beta\iint_{\mathcal{R}}\sqrt{\bar{f}(x)}\,dA$, it is not hard to verify (by a simple application of Jensen's inequality) that the resulting distribution is uniform on $\mathcal{R}$. The corresponding "radial" cost, while not equal to zero, is then $^2\!/_t\iint_{R}\|x\|\,dA$, substantially less than that obtained from the atomic distribution.

We are now interested in determining the distribution $f^*(\cdot)$ that maximizes $\mathbf{E}\,\mathrm{length}(\mathsf{VRP}(X))$, which we will bound from above by maximizing the right-hand side of (4). The concave maximization problem given by

$$\underset{f(\cdot)}{\text{maximize}}\ \iint_{\mathcal{R}}\frac{2}{t}\|x\|f(x) + \beta\sqrt{f(x)}\,dA \qquad s.t. \tag{5}$$
$$\iint_{\mathcal{R}}f(x)\,dA\ =\ 1$$
$$f(x)\ \geq\ 0 \quad \forall x\in\mathcal{R}\,,$$

11

with $f(\cdot)$ belonging to (for example) the Banach space L^2 over $\mathcal{R}$, can readily be approached using standard techniques of vector space optimization [43], and admits a Lagrangian dual problem in a single variable given by

$$\underset{\nu}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu - \frac{2}{t}\|x\|} \, dA + \nu \qquad s.t. \tag{6}$$

$$\nu \geq \frac{2}{t}\|x\| \quad \forall x \in \mathcal{R} \,.$$

The optimality conditions for the dual problem, which say that

$$\iint_{\mathcal{R}} \frac{\beta^2}{4(\nu - \frac{2}{t}\|x\|)^2} \, dA = 1 \,,$$

immediately imply the appropriate distributional form for $f^*(\cdot)$:

Theorem 7. *The optimal solution $f^*(\cdot)$ that solves (5) is of the form*

$$f^*(x) = \frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \,,$$

where $\nu^* \geq \frac{2}{t}\max_{x \in \mathcal{R}}\|x\|$ is the unique scalar such that $f^*(\cdot)$ integrates to one over the region.

Proof. Thanks to weak duality (which says that any feasible solution to (5) has an objective value that is at most equal to that of a feasible solution to (6)), it will suffice to verify that the primal and dual objective values are the same when $f^*(\cdot)$ and ν^* are as defined above:

$$
\begin{aligned}
\iint_{\mathcal{R}} \frac{2}{t}\|x\| f^*(x) + \beta\sqrt{f^*(x)} \, dA &= \iint_{\mathcal{R}} \left(\frac{2}{t}\|x\| \left[\frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \right] + \beta\sqrt{\frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2}} \right) dA \\
&= \iint_{\mathcal{R}} \frac{\beta^2}{2(\nu^* - \frac{2}{t}\|x\|)^2} \cdot \left(\nu^* - \frac{\|x\|}{t} \right) dA \\
&= \iint_{\mathcal{R}} \left(\frac{1}{4} \cdot \frac{\beta^2}{\left(\nu^* - \frac{2}{t}\|x\|\right)^2} \left(\nu^* - \frac{2}{t}\|x\| \right) + \frac{\nu^*\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \right) dA \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\left(\nu^* - \frac{2}{t}\|x\|\right)^2} \left(\nu^* - \frac{2}{t}\|x\| \right) dA + \nu^* \underbrace{\iint_{\mathcal{R}} \frac{\beta^2}{4(\nu^* - \frac{2}{t}\|x\|)^2} \, dA}_{=1} \\
&= \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu^* - \frac{2}{t}\|x\|} \, dA + \nu^*
\end{aligned}
$$

as desired. Figure 2 shows a collection of surface plots of the worst-case distribution $f^*(\cdot)$ on the unit square for varying values of t.

$\square$

(a) $t = 0.9$; $\nu^* = 2.5$ (b) $t = 0.5$; $\nu^* = 4.5$ (c) $t = 0.3$; $\nu^* = 7.5$ (d) $t = 0.1$; $\nu^* = 22.5$
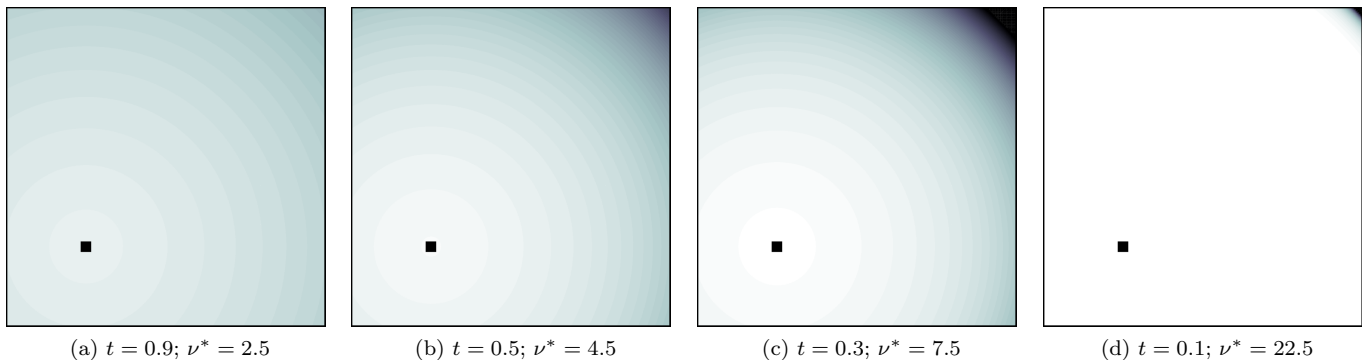
Figure 2: Surface plots of $f^*(\cdot)$ with decreasing values of t from left to right. The depot is indicated by the black square and darker color values correspond to higher densities.

One of the salient attributes of the worst-case distribution $f^*(\cdot)$ as written above is that the presence of the square root in the objective of (5) establishes an inverse proportionality between the optimal solution $f^*(\cdot)$ and the *square* of the distance to the depot (with some additional additive and multiplicative weights from the dual variable ν^* and the parameter t). This same inverse proportionality is shared by the classical *geographic gravity model* [7, 50, 58], which is "the most common formulation of the spatial interaction method" [50] and has historically been used to model a wide variety of demographic phenomena such as population migration [56], spatial utility for retail stores [49], and trip distributions between cities [60]. This would appear to lend credibility to our expression $f^*(\cdot)$, inasmuch as it takes a form that closely matches that of distributions for related problems.

Note that in proving Theorem 7, we maximized the right-hand side of (4), which is an upper bound of $\mathbf{E} \, \text{length}(\mathsf{VRP}(X))$. It is natural to wonder how much error we are incurring by maximizing something that is itself already an upper bound of the quantity of interest. To this end we offer the following:

Theorem 8. *For any convex planar region $\mathcal{R}$ that contains the origin and any scalar t, we have*

$$\frac{\max_{f(\cdot)} \frac{2}{t} \iint_{\mathcal{R}} \|x\| f(x) \, dA + \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)} \, dA}{\max_{f(\cdot)} \max \left\{ \frac{2}{t} \iint_{\mathcal{R}} \|x\| f(x) \, dA, \beta \iint_{\mathcal{R}} \sqrt{\bar{f}(x)} \, dA \right\}} \leq 3/2 \,, \tag{7}$$

where the maximization "$\max_{f(\cdot)}$" is taken over all probability distributions $f(\cdot)$ defined on $\mathcal{R}$.

Proof. See Section B of the Online Supplement. Note that it is obvious that the above ratio is bounded above by 2 because the numerator is the sum of the two terms in the $\max\{\cdot, \cdot\}$ expression in the denominator. □

13

5 Worst-case distributions with moment information

Our recent article [21] studies the uncapacitated version of (5) (that is, the TSP), where we are presented with additional constraints on the first and second moments of the distribution $f(\cdot)$; we require that $\iint_{\mathcal{R}} x f(x) \, dA = \mu$ for some given $\mu \in \mathcal{R}$ and that $\iint_{\mathcal{R}} xx^T f(x) \, dA \preceq \Sigma + \mu\mu^T$ where $\Sigma \succ \mathbf{0}$ is a given covariance matrix. In practice, these values of μ and Σ would be obtained by observing historical data. In this section, we study the problem (5) in which these moment constraints are present, so that the new problem of interest is written as

$$\underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{R}} \frac{2}{t} \|x\| f(x) + \beta\sqrt{f(x)} \, dA \qquad s.t. \tag{8}$$

$$\iint_{\mathcal{R}} x f(x) \, dA \;=\; \mu$$

$$\iint_{\mathcal{R}} xx^T f(x) \, dA \;\preceq\; \Sigma + \mu\mu^T$$

$$\iint_{\mathcal{R}} f(x) \, dA \;=\; 1$$

$$f(x) \;\geq\; 0 \quad \forall x \in \mathcal{R} \,,$$

with $f(\cdot)$ belonging to (for example) the Banach space L^2 over $\mathcal{R}$, whose dual is

$$\underset{\nu \in \mathbb{R}, \lambda \in \mathbb{R}^2, Q \in \mathbb{R}^{2\times 2}}{\text{minimize}} \iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{\nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\|} \, dA + \nu + \lambda^T \mu + (\Sigma + \mu\mu^T) \bullet Q \qquad s.t. \tag{9}$$

$$\nu + \lambda^T x + x^T Q x \;\geq\; \frac{2}{t}\|x\| \quad \forall x \in \mathcal{R}$$

$$Q \;\succeq\; \mathbf{0} \,.$$

We again find that the optimality conditions for (9), which say that

$$\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{(\nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\|)^2} \, dA \;=\; 1$$

$$\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x}{(\nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\|)^2} \, dA \;=\; \mu$$

$$\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 xx^T}{(\nu + \lambda^T x + x^T Q x - \frac{2}{t}\|x\|)^2} \, dA \;\preceq\; \Sigma + \mu\mu^T \,,$$

describe the optimal form of $f^*(\cdot)$:

Proposition 9. *The optimal solution $f^*(\cdot)$ that solves (8) is of the form*

$$f^*(x) = \frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|\right]^2} \,, \tag{10}$$

where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, and $Q^* \succeq \mathbf{0}$ satisfy

$$\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\| \geq 0$$

everywhere on $\mathcal{R}$, and

$$
\begin{aligned}
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2} \, dA &= 1 \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2} \, dA &= \mu \\
\iint_{\mathcal{R}} \frac{1}{4} \cdot \frac{\beta^2 x x^T}{(\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t}\|x\|)^2} \, dA &\preceq \Sigma + \mu\mu^T .
\end{aligned}
$$

Proof. See Section C of the online supplement for a rigorous proof (this is almost the same as in our earlier paper [21]). Figure 3 shows a few plots of $f^*(\cdot)$ for varying values of μ, Σ, and t. $\qquad \square$

One of the salient attributes of the worst-case distribution $f^*(\cdot)$, as can readily be observed in Figure 3, is that small values of t (as in 3a) result in worst-case distributions that are tightly clustered around the centroid. In addition, as we vary the covariance matrix in 3e-3h, the worst-case distribution also becomes more spread out.

5.1 The impact of second-moment information

In this section we describe some results that shed some light on the usefulness of second-moment information: it is clear that distributions with high VRP costs are those that are as "spread out" as possible, either in the sense of having large amounts of demand located away from the depot (when capacities are low and thus the "radial cost" is dominant) or in the sense of being distributed as uniformly as possible (when capacities are high and thus the "circular cost" is dominant). One would therefore expect that, in determining the worst-case cost of an unknown distribution, the covariance bound $\iint_{\mathcal{R}} x x^T f(x) \, dA \preceq \Sigma + \mu\mu^T$ should be particularly relevant.

In order to derive a rough understanding of the effect of this covariance bound, we consider a simple model in which $\mathcal{R}$ is a disk $\mathcal{D}$ with area A, Σ is a diagonal matrix whose entries are both a scalar s, and both the depot and the mean of the distribution μ are located at the origin. Purely for the ease of exposition, we find it easier to consider the radial cost and the circular cost independently of one another. The problem of maximizing the radial
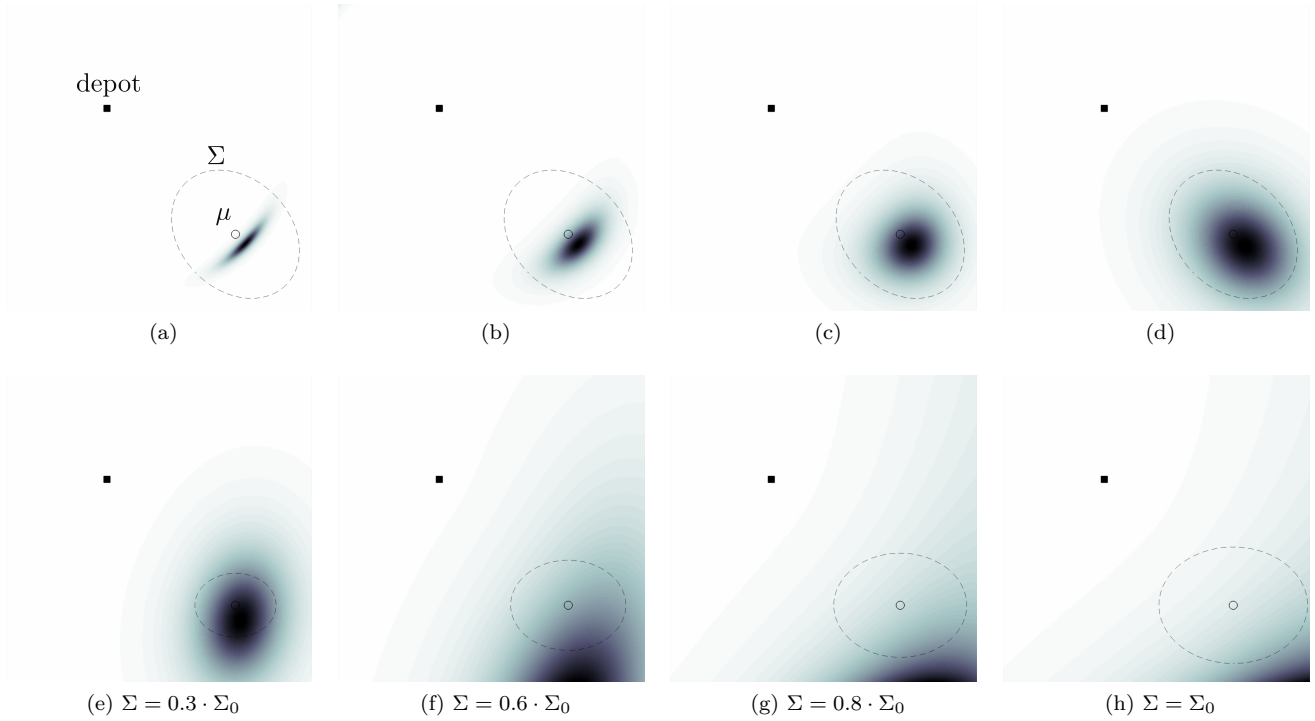
Figure 3: Surface plots of $f^*(\cdot)$ for $\mu = (0.75, 25)$, a depot located at $(0.33, 0.66)$, and varying values of Σ and t. In 3a-3d, we fix $\Sigma = \left(\begin{smallmatrix} 0.075 & -0.02 \\ -0.02 & 0.075 \end{smallmatrix}\right)$ and we have $t \in \{0.01, 0.1, 1, 10\}$; in 3e-3h, we fix $t = 1$ and let Σ be a scalar multiple of $\Sigma_0 = \left(\begin{smallmatrix} 0.06 & 0 \\ 0 & 0.095 \end{smallmatrix}\right)$.

cost is given by

$$\underset{f(\cdot)}{\operatorname{maximize}} \iint_{\mathcal{D}} \frac{2}{t} \|x\| f(x) \, dA \qquad s.t. \tag{11}$$

$$\iint_{\mathcal{D}} x f(x) \, dA = (0,0)^T$$

$$\iint_{\mathcal{D}} x x^T f(x) \, dA \preceq \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}$$

$$\iint_{\mathcal{D}} f(x) \, dA = 1$$

$$f(x) \geq 0 \quad \forall x \in \mathcal{D},$$

and the following result is entirely unsurprising:

Proposition 10. *An optimal solution to problem (11) occurs when $f(\cdot)$ is a uniform mixture of two atomic masses at the points $(\pm r, 0)^T$, where $r = \min\{\sqrt{s}, \sqrt{A/\pi}\}$.*

Proof. See Section D of the online supplement. Note that the size of the service region $\mathcal{D}$ has no effect on the solution to (11) provided that $s \leq A/\pi$. $\qquad \square$

A more interesting problem is that of maximizing the circular cost, which is instead given by

$$\underset{f(\cdot)}{\operatorname{maximize}} \iint_{\mathcal{R}} \beta \sqrt{f(x)} \, dA \qquad s.t. \tag{12}$$

$$\iint_{\mathcal{D}} x f(x) \, dA = (0,0)^T$$

$$\iint_{\mathcal{D}} x x^T f(x) \, dA \preceq \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}$$

$$\iint_{\mathcal{D}} f(x) \, dA = 1$$

$$f(x) \geq 0 \quad \forall x \in \mathcal{D}$$

whose dual problem is

$$\underset{\nu \in \mathbb{R}, \lambda \in \mathbb{R}^2, Q \in \mathbb{R}^{2 \times 2}}{\operatorname{minimize}} \iint_{\mathcal{D}} \frac{1}{4} \cdot \frac{\beta^2}{\nu + \lambda^T x + x^T Q x} \, dA + \nu + s(q_{11} + q_{22}) \qquad s.t. \tag{13}$$

$$\nu + \lambda^T x + x^T Q x \geq 0 \quad \forall x \in \mathcal{D}$$

$$Q \succeq \mathbf{0}.$$

By exploiting the symmetry of $\mathcal{D}$ and using convexity of our objective functions and feasible regions, it is not hard to show that the first-moment constraint $\iint_{\mathcal{R}} x f(x) \, dA = (0,0)^T$ in (12) is redundant, and thus we can remove the

17

variable λ from (13). It follows similarly that the optimal matrix Q must also be a diagonal matrix whose entries are both a scalar q, and we are therefore free to consider the simpler dual problem

$$\underset{\nu, q \in \mathbb{R}}{\text{minimize}} \iint_{\mathcal{D}} \frac{1}{4} \cdot \frac{\beta^2}{\nu + q\|x\|^2} \, dA + \nu + 2sq \qquad s.t. \tag{14}$$
$$\nu, q \geq 0.$$

The optimal solution to (12) can be described by way of (14) as follows:

Theorem 11. *For fixed values of s, the cost* OBJ *of (12) satisfies*

$$\mathsf{OBJ} \sim \beta \sqrt{2\pi s \log A}$$

as $A \to \infty$.

Proof. See Section E of the online supplement; the key idea is to set

$$q = \frac{\beta}{4} \cdot \sqrt{\frac{2\pi \log A}{s}}$$
$$\nu = \frac{\beta^2}{4} \cdot \sqrt{\frac{2\pi s}{\log A}}$$

and look at the resulting asymptotic series in A. $\qquad \square$

Figure 4 shows a plot of the cost OBJ as a function of the variance term s for $A = 1$. Notice that, when s is very small, the worst-case workload is close to 0 because all demand must then be tightly clustered around the depot. Moreover, for $s > {}^A/_{4\pi}$, the covariance constraint is not binding because the uniform distribution on the region $\mathcal{D}$ is already feasible for (12), and therefore OBJ is constant thereafter. The key insight that Theorem 11 offers us is a concise expression that relates the covariance of the demand distribution and the workload that vehicles incur. Applying Theorem 11, we can therefore conclude that a valid asymptotic bound for the worst-case VRP workload is therefore

$$\mathbf{E} \, \text{length}(\mathsf{VRP}(X)) \lesssim \sqrt{N} \cdot \left(\frac{2}{t} \min\{\sqrt{s}, \sqrt{A/\pi}\} + \beta \sqrt{2\pi s \log A} \right).$$

Remark. If we want to consider arbitrary covariance matrices Σ, one approach is to use an alternate support region
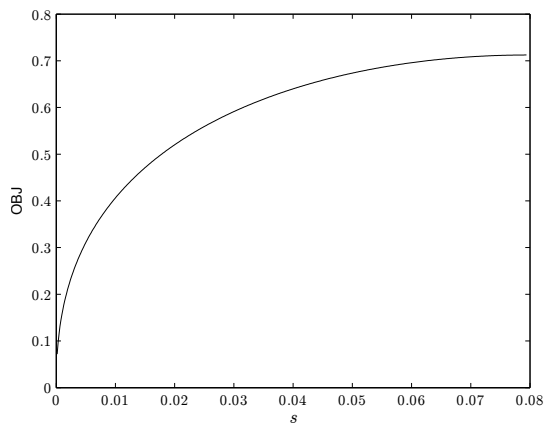
Figure 4: The worst-case cost OBJ to problem (12) for $s \in (0, A/4\pi]$ and $A = 1$. The curve above shows the relationship between the spatial covariance of demand (measured through the single diagonal element s) and the worst-case cost of problem (12).

for which the integral in the objective function is easier to calculate. In particular, one might consider the problem

$$
\begin{aligned}
\underset{f(\cdot)}{\text{maximize}} \iint_{\mathcal{E}} \beta \sqrt{f(x)} \, dA \qquad & s.t. \\
\iint_{\mathcal{E}} x f(x) \, dA &= (0,0)^T \\
\iint_{\mathcal{E}} x x^T f(x) \, dA &\preceq \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \\
\iint_{\mathcal{E}} f(x) \, dA &= 1 \\
f(x) &\geq 0 \quad \forall x \in \mathcal{E}
\end{aligned}
$$

where $\mathcal{E}$ is the ellipse that satisfies

$$
\mathcal{E} := \left\{ (x_1, x_2) \in \mathbb{R}^2 : \left(\frac{x_1}{s_1}\right)^2 + \left(\frac{x_2}{s_2}\right)^2 \leq \frac{A}{\pi s_1 s_2} \right\},
$$

which clearly has area A. The objective function of the above problem can obviously be converted to an integral over a disk with area A by applying the change of variables $x_1 \mapsto u_1 \sqrt{s_1/s_2}$ and $x_2 \mapsto u_2 \sqrt{s_2/s_1}$, so that the resulting integral is taken over a disk of area A in the (u_1, u_2) plane. The cost OBJ of the optimal solution (12) then satisfies

$$
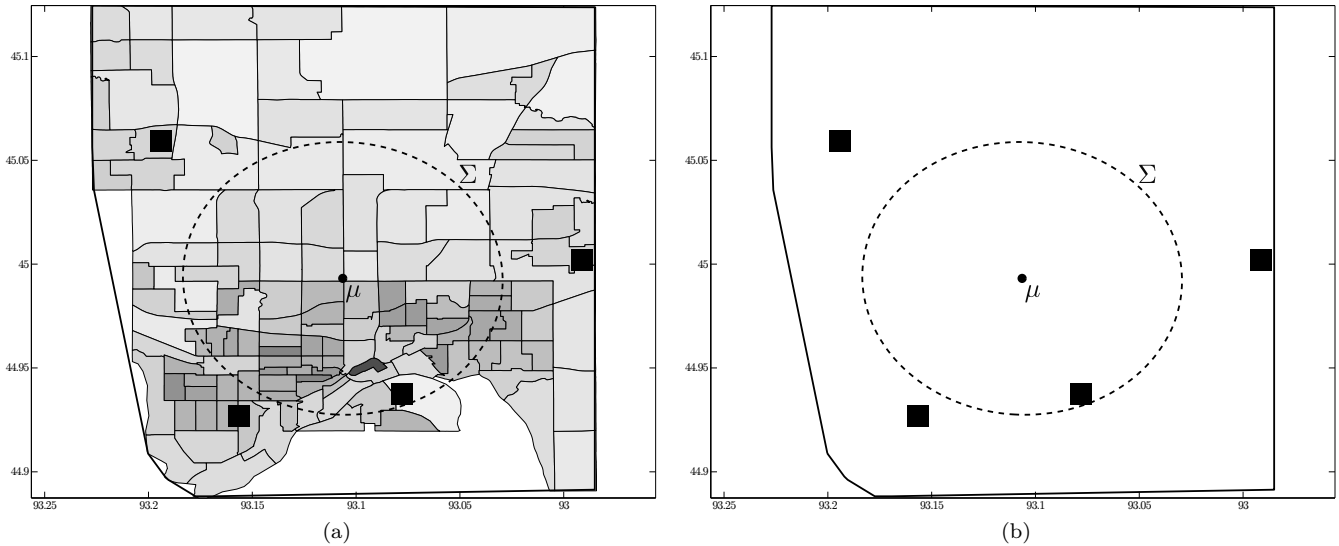\mathsf{OBJ} \sim \sqrt{2\pi} \beta (s_1 s_2)^{1/4} \sqrt{\log A}.
$$

19

Figure 5: Figure (5a) shows the population density of Ramsey County, Minnesota, the mean and covariance matrices of the population density, and four post offices. In practice, of course, we do not know the true demand density, and thus (5b) shows the input to our various optimization procedures.

5.2 A computational experiment

In this section we describe the results of a computational experiment applied to a data set obtained from the United States Census [59]. The region $\mathcal{R}$ is given by the convex hull of Ramsey County, Minnesota, and the mean μ and covariance matrix Σ are obtained from the population density thereof. In addition, we suppose that there are four vehicle depots, which are taken as the locations of four post offices; the complete setup of our experiment is shown in Figure 5, and was used previously to study the uncapacitated VRP in our paper [21]. Note that Figure (5b) shows the true input to our problem (i.e. the population density is not assumed to be known and we have only first and second moment information)

5.2.1 Theoretical preliminaries

In the following simulation, we will consider the problem of designing service districts for each of the four vehicle depots. This requires a minor generalization of Proposition 9 for the case where we are interested in the distribution

whose workload on a particular sub-region R_i is as large as possible; this problem is written as

$$\underset{f(\cdot)}{\text{maximize}} \iint_{R_i} \frac{2}{t} \|x - p_i\| f(x) + \beta \sqrt{f(x)} \, dA \qquad s.t. \tag{15}$$

$$\iint_{\mathcal{R}} x f(x) = \mu$$

$$\iint_{\mathcal{R}} xx^T f(x) \, dA \preceq \Sigma + \mu\mu^T$$

$$\iint_{\mathcal{R}} f(x) \, dA = 1$$

$$f(x) \geq 0 \quad \forall x \in \mathcal{R},$$

which differs from problem (8) only in that we have a different domain of integration in the objective function.

Proposition 12. *The optimal solution $f^*(\cdot)$ that solves (15) is of the form*

$$f^*(x) = \frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t} \|x - p_i\| \right]^2} \mathcal{I}(x \in R_i) + s^* \delta(x - \bar{x}),$$

where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, and $Q^* \succeq \mathbf{0}$ satisfy

$$\nu^* + (\lambda^*)^T x + x^T Q^* x - \frac{2}{t} \|x - p_i\| \geq 0$$

everywhere on $\mathcal{R}$, $\mathcal{I}(\cdot)$ denotes an indicator function, $\bar{x}$ is a point in $\mathcal{R}$, and s^* is a scalar.

Proof. This is almost identical to Proposition 9, which is proven in Section C of the online supplement; the only difference is the domain of integration in the objective function. □

Recall also that problems (5), (8), and (15) are all designed to maximize the *upper* bound (i.e. the right-hand side) of the original expression (4). If we want any information about the solution quality of our approach, it is necessary that we also examine the *lower* bound thereof; this problem is given by

$$\underset{f(\cdot)}{\text{maximize}} \max \left\{ \iint_{R_i} \frac{2}{t} \|x - p_i\| f(x), \iint_{R_i} \beta \sqrt{f(x)} \, dA \right\} \qquad s.t. \tag{16}$$

$$\iint_{\mathcal{R}} x f(x) = \mu$$

$$\iint_{\mathcal{R}} xx^T f(x) \, dA \preceq \Sigma + \mu\mu^T$$

$$\iint_{\mathcal{R}} f(x) \, dA = 1$$

$$f(x) \geq 0 \quad \forall x \in \mathcal{R},$$

21

whose optimal solution is characterized as follows:

Proposition 13. *The optimal solution $f^*(\cdot)$ that solves (16) takes either the form*

$$f_1^*(x) := \frac{1}{4} \cdot \frac{\beta^2}{\left[\nu^* + (\lambda^*)^T x + x^T Q^* x\right]^2} \mathcal{I}(x \in R_i) + s^* \delta(x - \bar{x})$$

where $\nu^* \in \mathbb{R}$, $\lambda^* \in \mathbb{R}^2$, and $Q^* \succeq \mathbf{0}$, or the form

$$f_2^*(x) := \sum_{i=1}^{7} s_i^* \delta(x - \bar{x}_i),$$

i.e. a mixture of at most seven atomic distributions.

Proof. Obviously, we simply analyze problem (16) by isolating the two components of the objective function. The proof that the worst-case distribution takes the form of $f_1^*(\cdot)$ when we seek to maximize $\iint_{R_i} \beta \sqrt{f(x)} \, dA$ was already derived in [21] and the proof that the worst-case distribution takes the form $f_2^*(\cdot)$ when we seek to maximize $\iint_{R_i} \frac{2}{t_N} \|x\| f(x)$ is in Section F of the online supplement. $\square$

For the sake of notational compactness, from now on, we will let $\bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ and $\Psi(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ denote the objective values of problems (15) and (16) respectively.

5.2.2 Computational results

Our objective here is to design service districts for each of the four vehicle depots whose worst-case workloads are as small as possible. In particular, we will focus on the problem of designing sub-regions R_i such that the upper bound of the maximum workload of all $n = 4$ sub-regions, i.e. $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$, is minimized. Our problem can be written as

$$\begin{aligned}
\underset{R_1,\ldots,R_n}{\text{minimize}} \ \max_i \ & \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R}) \qquad s.t. \\
\bigcup_{i=1}^{n} R_i &= \mathcal{R} \\
R_i \cap R_j &= \emptyset \quad \forall i \neq j.
\end{aligned}$$

In order to accomplish this we decide to restrict the set of sub-regions R_i (which is, of course, an infinite-dimensional family) to the set of *power diagrams* associated with the depots [9]: given a vector of weights (w_1, w_2, w_3, w_4), we assign depot i (located at point p_i) to the sub-region R_i, defined as
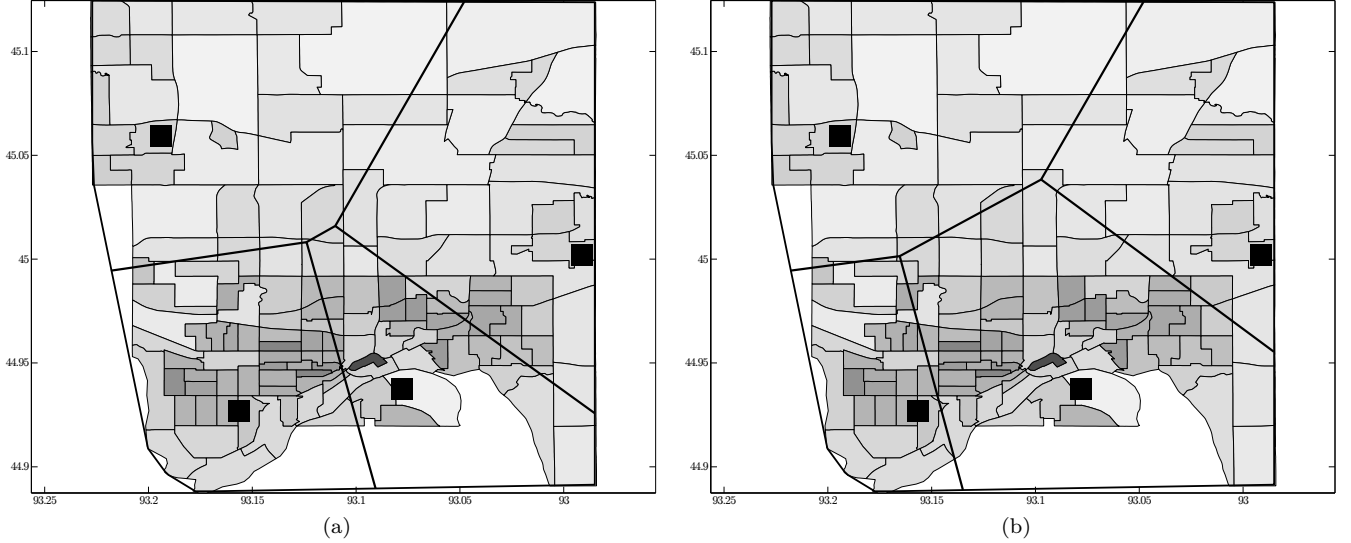
Figure 6: Two power diagram partitions of the service region $\mathcal{R}$. The weight associated with the bottom right sub-region is larger in (6a) than in (6b).

$$R_i = \left\{ x \in \mathcal{R} : \|x - p_i\|^2 - w_i \leq \|x - p_j\|^2 - w_j \quad \forall j \right\} . \tag{17}$$

A power diagram is simply a Voronoi diagram in which one also has weights associated with the depots; it has the useful property that all of the regions R_i are polygonal and convex. Of course, when all weights are equal, the power diagram simply reduces to the standard Voronoi diagram, and if we increase the weight associated with one of the depots p_i, the size of sub-region R_i will increase. Figure 6 shows two examples of power diagram partitions. In Section G of the online supplement, Algorithm 1 describes a standard branch-and-bound procedure for determining weights w_i^* that approximately minimize $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$. We refer the reader to Figure 7, which shows the worst-case workloads associated with the sub-regions R_i together with the actual workloads, as induced by the true population density. The three bands shown there are interpreted as follows: the lightest shaded band corresponds to the worst-case cost and indicates the upper and lower bounds that are obtained from the optimal objective function cost to (15), which is an upper bound of the worst possible VRP cost, and the optimal objective function cost to (16), which is a lower bound of the worst possible VRP cost. In other words, the lightest shaded band shows $\max_i \bar{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$ and $\max_i \underline{\Psi}(\mu, \Sigma, t, p_i, R_i, \mathcal{R})$, when the sub-regions R_i are power diagrams that are obtained by Algorithm 1 in the online supplement. Similarly, the second shaded band corresponds to the actual cost associated with the partition obtained by Algorithm 1 and indicates the upper and lower bounds of the VRP cost of our algorithm, i.e. $\max_i \iint_{R_i} \frac{2}{t} \|x\| f(x) + \beta \sqrt{\bar{f}(x)} \, dA$ for the upper bound and $\max_i \max \left\{ \iint_{R_i} \frac{2}{t} \|x\| f(x), \iint_{R_i} \beta \sqrt{\bar{f}(x)} \, dA \right\}$ for the lower bound, where $f(\cdot)$ represents the true population density
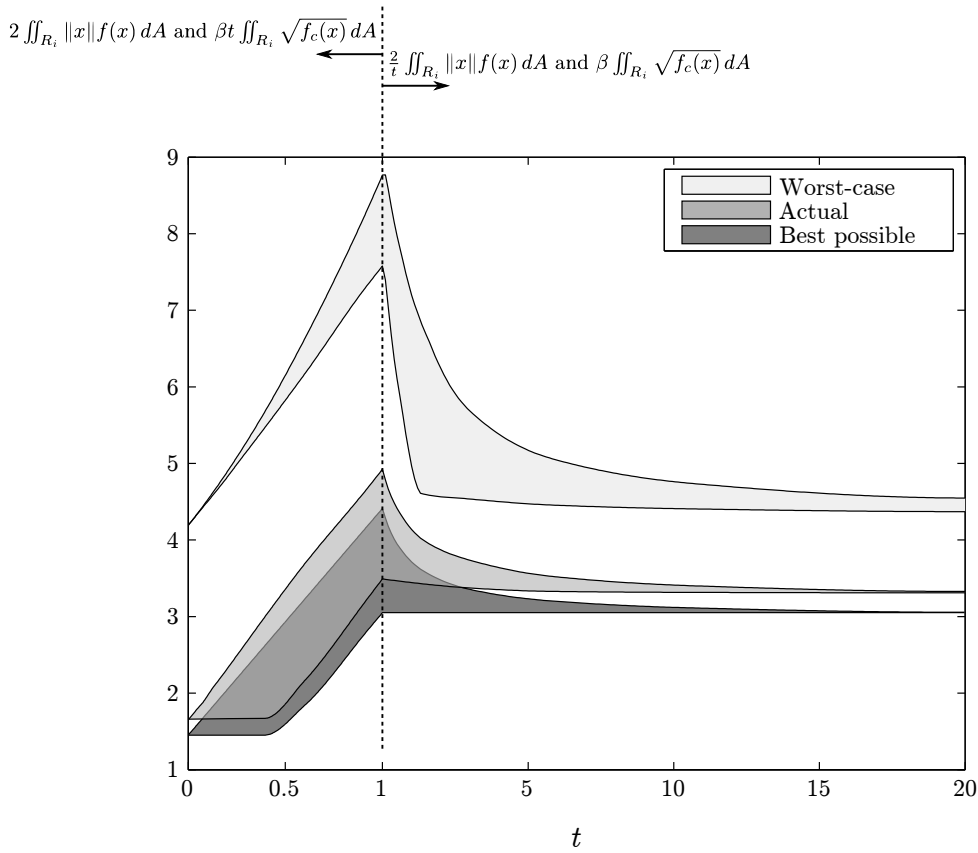
23

Figure 7: The plot above shows the workloads that are induced by Algorithm 1 for $t \in (0, 20]$, measured as the *maximum workload* of the four sub-regions R_i (rather than, say, the sum of the workloads in the four sub-regions R_i). For each value of t, we naturally require a separate run of the algorithm (we performed 100 runs overall). To the right of the dashed line (i.e. for $t \geq 1$), we plot the various upper and lower bounds to our problem, and to the left of the dashed line (i.e. for $t < 1$), we plot essentially the same quantities, but multiplied through by t in all cases. This is because all costs (the worst-case cost, the actual cost, and the best possible cost) explode towards infinity as $t \to 0$ and thus we find this to be a more useful measure of the relative costs.

as shown in Figure 5a. Finally, the darkest shaded band corresponds to the best possible actual costs (in terms of upper and lower bounds) that could be realized by partitioning the region optimally, which we obtained by using the algorithm in [22].

6 Conclusions

We have considered the asymptotic behavior of the vehicle routing problem with time windows and the capacitated vehicle routing problem in the Euclidean plane. In studying the VRP with time windows, we find that the worst possible temporal distribution of demand is either to have a large amount of demand concentrated in a single time period (when total demand is small), or uniform demand throughout all time periods (when total demand is large).

24

Thus, for upstart businesses with a small customer base, the worst possible outcome is to have a single period of concentrated activity, and for established business with a large customer base, the worst possible outcome is to have all demand spread uniformly throughout the service period.

For the capacitated VRP, we initially found that the worst-case distribution follows an "inverse square" law that closely resembles the classical gravity model of spatial interaction. In order to extend this result to a decision support system, we next computed a closed-form expression for the worst-case demand distribution when one knows the first and second moments and as well as the support of the distribution, which we used in a computational districting experiment. Such districting strategies are useful when one has limited information about the locations of demand points and seeks an assignment scheme that can be performed without full knowledge of the demand distribution.

One might also seek to understand the impact of *first* moment information in a back-of-the-envelope analysis similar to that in Section 5.1; we expect that the location of the center of mass of a distribution should impact the worst-case workloads in a distinctly different manner from that of the covariance matrix and we intend to study it in the near future. Another natural direction for further research would be to tighten the initial bounds in (4). This appears tractable because they are based on the earlier bounds in (1). Since (1) holds for any Euclidean point set, it is stronger than an asymptotic result, and it would appear that this property may be exploitable in improving (4). Finally, it is worth asking whether alternate distributional ambiguity sets (as opposed to support and first and second moment information) might be useful for the distributionally robust VRP; one possibility would be the use of statistical metrics such as the *Kullback-Leibler divergence* or *Kolmogorov-Smirnov statistic*; these have recently been applied to distributionally robust optimization problems in [12, 14] and will likely yield additional insights into the costs of providing service in VRP.

References

[1] Caviar. `http://www.trycaviar.com/`. Accessed: 2014-09-27.

[2] DoorDash Food Delivery. `http://www.doordash.com`. Accessed: 2014-10-27.

[3] Food Delivery & Restaurants Delivery - Order Food Online - BiteSquad.com. `http://www.bitesquad.com`. Accessed: 2014-10-27.

[4] Good Eggs. `http://www.goodeggs.com`. Accessed: 2014-10-27.

[5] A. Agra, M. Christiansen, R. Figueiredo, L. M. Hvattum, M. Poss, and C. Requejo. The robust vehicle routing problem with time windows. *Computers & Operations Research*, 40(3):856–866, 2013.

[6] M. Allahviranloo, J. Y. J. Chow, and W. W. Recker. Selective vehicle routing problems under uncertainty without recourse. *Transportation Research Part E: Logistics and Transportation Review*, 62:68–88, 2014.

[7] J. E. Anderson. The gravity model. Technical report, National Bureau of Economic Research, 2010.

[8] D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook. *The Traveling Salesman Problem: A Computational Study.* Princeton University Press, 2011.

[9] F. Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96, 1987.

[10] M. Barkaoui and M. Gendreau. An adaptive evolutionary approach for real-time vehicle routing and dispatching. *Computers & Operations Research*, 2013.

[11] J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327, 1959.

[12] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[13] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[14] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *arXiv preprint arXiv:1401.0212*, 2013.

[15] D. J. Bertsimas and D. Simchi-Levi. A new generation of vehicle routing research: robust algorithms, addressing uncertainty. *Operations Research*, 44(2):286–304, 1996.

[16] D. J. Bertsimas and G. van Ryzin. Stochastic and dynamic vehicle routing in the euclidean plane with multiple capacitated vehicles. *Operations Research*, 41(1):60–76, 1993.

[17] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[18] L. D. Burns, R. W. Hall, D. E. Blumenfeld, and C. F. Daganzo. Distribution strategies that minimize transportation and inventory costs. *Operations Research*, 33(3):469–490, 1985.

[19] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[20] J. F. Campbell. Location and allocation for distribution systems with transshipments and transportion economies of scale. *Annals of Operations Research*, 40(1):77–99, 1992.

[21] J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):727–744, 2013.

[22] J. G. Carlsson and R. Devulapalli. Dividing a territory among several facilities. *INFORMS Journal on Computing*, 25(4):730–742, 2012.

[23] C. Daganzo. *Logistics Systems Analysis.* Springer, 2005.

[24] C. F. Daganzo. The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. *Transportation Science*, 18(4):331–350, 1984.

[25] C. F. Daganzo. Modeling distribution problems with time windows: Part i. *Transportation Science*, 21(3):171–179, 1987.

[26] C. F. Daganzo. Modeling distribution problems with time windows. part ii: Two customer types. *Transportation Science*, 21(3):180–187, 1987.

[27] G. B. Dantzig and J. H Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

[28] L. Few. The shortest path and the shortest road through *n* points. *Mathematika*, 2:141–144, 1955.

[29] M. A. Figliozzi. Planning approximations to the average length of vehicle routing problems with time window constraints. *Transportation Research Part B: Methodological*, 43(4):438–447, 2009.

[30] J. Geunes, Z.-J. M. Shen, and A. Emir. Planning and approximation models for delivery route based services with price-sensitive demands. *European journal of operational research*, 183(1):460–471, 2007.

[31] L. A. Goddyn. Quantizers and the worst-case euclidean traveling salesman problem. *Journal of Combinatorial Theory, Series B*, 50(1):65–81, 1990.

[32] C. E. Gounaris, W. Wiesemann, and C. A. Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 2013.

[33] M. Haimovich and Thomas L. Magnanti. Extremum properties of hexagonal partitioning and the uniform distribution in Euclidean location. *SIAM J. Discrete Math.*, 1:50–64, 1988.

[34] M. Haimovich and A. H. G. Rinnooy Kan. Bounds and heuristics for capacitated routing problems. *Mathematics of Operations Research*, 10(4):527–542, 1985.

[35] D. S. Hochbaum. When are NP-hard location problems easy? *Annals of Operations Research*, 1:201–214, 1984.

[36] M. Huang, K. R. Smilowitz, and B. Balcik. A continuous approximation approach for assessment routing in disaster relief. *Transportation Research Part B: Methodological*, 50:20–41, 2013.

[37] GrubHub Inc. Form S-1: Registration Statement under the Securities Act of 1933, February 2014. `https://www.sec.gov/Archives/edgar/data/1594109/000119312514075544/d647121ds1.htm`. [Online; posted 28-Feb-2014].

[38] O. Jabali, M. Gendreau, and G. Laporte. A continuous approximation model for the fleet composition problem. *Transportation Research Part B: Methodological*, 46(10):1591–1606, 2012.

[39] H. J. Karloff. How long can a euclidean traveling salesman tour be? *SIAM Journal on Discrete Mathematics*, 2(1):91–99, 1989.

[40] S.G. Krantz and H.R. Parks. *Geometric Integration Theory*. Cornerstones. Birkhäuser, 2008.

[41] C. Lee, K. Lee, and S. Park. Robust vehicle routing problem with deadlines and travel time/demand uncertainty. *Journal of the Operational Research Society*, 63(9):1294–1306, 2011.

[42] Kalev Leetaru. The $200 Uber ride and the realtime data-driven sharing economy, January 2016. `http://www.forbes.com/sites/kalevleetaru/2016/01/02/the-200-uber-ride-and-the-realtime-data-driven-sharing-economy/`. [Online; posted 2-Jan-2016].

[43] D. G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1968.

[44] A. G. N. Novaes, J. E. S. de Cursi, and O. D. Graciolli. A continuous approach to the design of physical distribution systems. *Computers & Operations Research*, 27(9):877–893, 2000.

[45] C. S. Orloff. Route constrained fleet scheduling. *Transportation Science*, 10(2):149–168, 1976.

[46] Y. Ouyang. Design of vehicle routing zones for large-scale distribution systems. *Transportation Research Part B: Methodological*, 41(10):1079–1093, 2007.

[47] C. H. Papadimitriou. Worst-case and probabilistic analysis of a geometric location problem. *SIAM Journal on Computing*, 10:542, 1981.

[48] C. Redmond and J. E. Yukich. Limit theorems and rates of convergence for euclidean functionals. *The Annals of Applied Probability*, 4(4):pp. 1057–1073, 1994.

[49] W. J. Reilly. *The law of retail gravitation*. WJ Reilly, 1931.

[50] J.P. Rodrigue, C. Comtois, and B. Slack. *The Geography of Transport Systems*. Routledge, 2009.

[51] D. Romik. Stirling's approximation for n!: the ultimate short proof? *American Mathematical Monthly*, pages 556–557, 2000.

[52] T. L. Snyder and J. M. Steele. Equidistribution in all dimensions of worst-case point sets for the traveling salesman problem. *SIAM Journal on Discrete Mathematics*, 8(4):678–683, 1995.

[53] O. Solyalı, J.-F. Cordeau, and G. Laporte. Robust inventory routing under demand uncertainty. *Transportation Science*, 46(3):327–340, 2012.

[54] J. M. Steele. Subadditive euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, 9(3):pp. 365–376, 1981.

[55] J.M. Steele. *Probability Theory and Combinatorial Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[56] J. Q. Stewart. Demographic gravitation: evidence and applications. *Sociometry*, pages 31–58, 1948.

[57] I. Sungur, F. Ordóñez, and M. Dessouky. A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, 40(5):509–523, 2008.

[58] J. Tinbergen. Shaping the world economy; suggestions for an international economic policy. *Books (Jan Tinbergen)*, 1962.

[59] United States Census Bureau. American FactFinder. `http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml`. Accessed: 2014-05-20.

[60] A. M. Voorhees. A general theory of traffic movement. In *Proceedings of the Institute of Traffic Engineers*, pages 46–50, 1955.