The generalized TSP and trip chaining

IWSSSCM3

John Gunnar Carlsson

Epstein Department of Industrial and Systems Engineering, University of Southern California

January 7, 2016

Revolutionary Research . . . Relevant Results

The EOQ formula

Given fixed costs K, demand rate a, and holding cost h, the optimal order quantity Q^* is equal to

$$Q^* = \sqrt{\frac{2aK}{h}}\,;$$

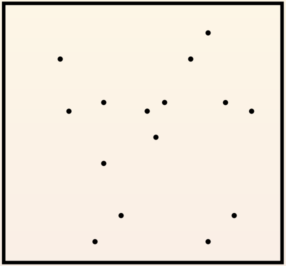
one obtains this by minimizing the cost per unit time, which is

$$\frac{aK}{Q} + ac + \frac{hQ}{2}\,,$$

and gives an optimal cost of

$$\left.\frac{aK}{Q} + ac + \frac{hQ}{2}\right|_{Q=\sqrt{2aK/h}} = \sqrt{2aKh} + ac$$

An "EOQ formula" for TSP

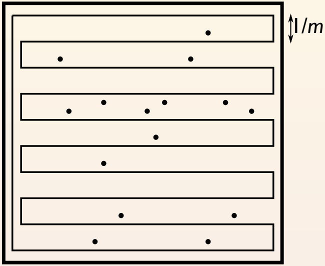


Consider n points distributed uniformly in the unit square $\mathcal{S}$:

- Let m be an even integer, and suppose that we "zig-zag" across $\mathcal{S}$ m times, which has length $\leq m + 2$
- Each point is at most $\frac{1}{2m}$ away from this path, thus we can round trip to each point with cost $\leq \frac{1}{m}$
- The cost of this tour is at most

$$m + 2 + \frac{n}{m} \implies \text{OPT} = 2\sqrt{n} + 2$$

An "EOQ formula" for TSP

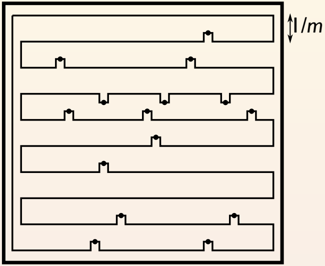


Consider n points distributed uniformly in the unit square $\mathcal{S}$:

- Let m be an even integer, and suppose that we "zig-zag" across $\mathcal{S}$ m times, which has length $\leq m + 2$
- Each point is at most $\frac{1}{2m}$ away from this path, thus we can round trip to each point with cost $\leq \frac{1}{m}$
- The cost of this tour is at most

$$m + 2 + \frac{n}{m} \implies \text{OPT} = 2\sqrt{n} + 2$$

An “EOQ formula” for TSP



Consider n points distributed uniformly in the unit square $\mathcal{S}$:

- Let m be an even integer, and suppose that we “zig-zag” across $\mathcal{S}$ m times, which has length $\leq m + 2$
- Each point is at most $\frac{1}{2m}$ away from this path, thus we can round trip to each point with cost $\leq \frac{1}{m}$
- The cost of this tour is at most

$$m + 2 + \frac{n}{m} \implies \text{OPT} = 2\sqrt{n} + 2$$

Beardwood-Halton-Hammersley Theorem (uniform case)

Theorem

Let $\{X_i\}$ *be a sequence of independent uniform samples on a compact region* $\mathcal{R} \subset \mathbb{R}^2$ *with area* 1. *Then with probability one,*

$$\lim_{N \to \infty} \frac{TSP(X_1, \ldots, X_N)}{\sqrt{N}} = \beta_{TSP}$$

where $TSP(X_1, \ldots, X_N)$ *denotes the length of a TSP tour of points* $X_1, \ldots, X_N$ *and* β_{TSP} *is a constant between* 0.6250 *and* 0.9204.

- This says that we can approximate the length of a tour as $\beta_{\mathrm{TSP}}\sqrt{N}$

Beardwood-Halton-Hammersley Theorem

Theorem

Let $\{X_i\}$ be a sequence of i.i.d. samples from an absolutely continuous probability density functionf$(\cdot)$ on a compact region $\mathcal{R} \subset \mathbb{R}^2$. Then with probability one,

$$\lim_{N \to \infty} \frac{TSP(X_1, \ldots, X_N)}{\sqrt{N}} = \beta_{TSP} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$$

where $TSP(X_1, \ldots, X_N)$ denotes the length of a TSP tour of points $X_1, \ldots, X_N$ and β_{TSP} is a constant between 0.6250 *and* 0.9204.

- This says that we can approximate the length of a tour as $\beta_{\mathrm{TSP}}\sqrt{N} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$
- We also see that the uniform distribution maximizes $\beta_{\mathrm{TSP}} \iint_{\mathcal{R}} \sqrt{f(x)}\, dA$ over all distributions $f(\cdot)$, i.e. "clustering is good"

Outline

- The generalized TSP and delivery services
- Package delivery with drones

The GTSP: Motivating example

Question

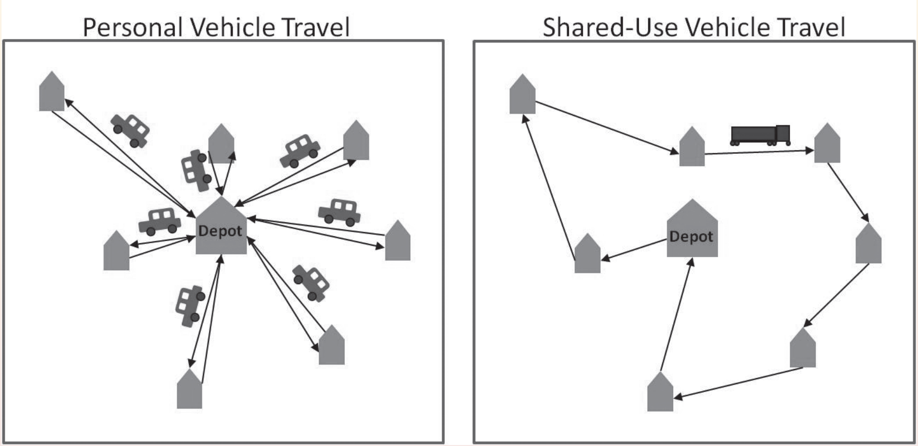
What happens to the carbon footprint of a city when its inhabitants start shopping online?

Intuition

Several things happen at once:

- Fewer trips by locals
- More work for delivery trucks, but on an *economy of scale* due to infrastructure
- The key issue: transportation that used to be ***local*** now becomes ***global***
- Is this always good? Do households have an economy of scale of their own?

Standard model



Standard model

Shopping can be part of a wider combined trip and involve only a minor detour. We assume that where a shopper undertakes trip chaining, the shopping component of the trip makes up a quarter of the overall total mileage.

–A. C. McKinnon and A. Woodburn

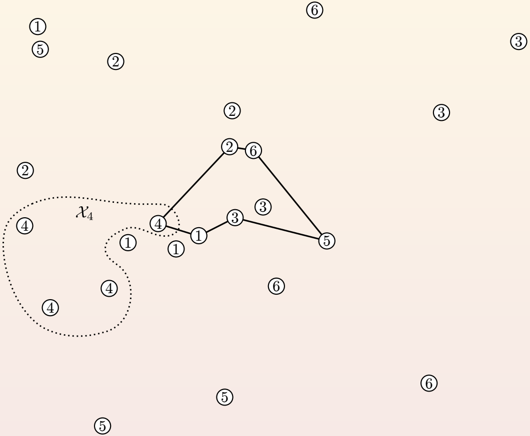
Generally, social network members will not participate or choose the burden of pickup if they have to go to a pickup point solely for the purpose of making a pickup for another person. Pickup trips for social network actors can be regarded as a chain event and is a determining variable. We assumed a 100% trip chain to additional mileage for pickup in both PLS and SPLS – in other words, the entire detour distance for pickup is attributed to the package. By contrast, previous research has applied a 0% trip chain effect for pickup.

–K. Suh, T. Smith, and M. Linhoff

A simple model

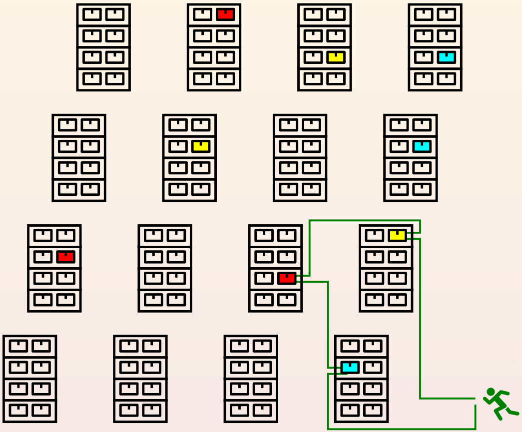
- City has area 1 and population N people
- Each person has n errands to do daily (bank, groceries, etc.)
- For each errand, there are k places to do these things
- Each person's daily route consists of a ***generalized TSP tour*** of the sets of points $\mathcal{X}_1, \ldots, \mathcal{X}_n$

The generalized TSP



Here $n = 6$ and $k = 4$

Warehouse application



Is it more efficient to stock the same good in multiple locations in a warehouse?

The generalized TSP

- What can we say about the GTSP? How long is it?
- There are two limiting cases that are interesting, either $n \to \infty$ or $k \to \infty$
- Our "gold standard" would be the BHH Theorem

The generalized TSP, limiting case 1

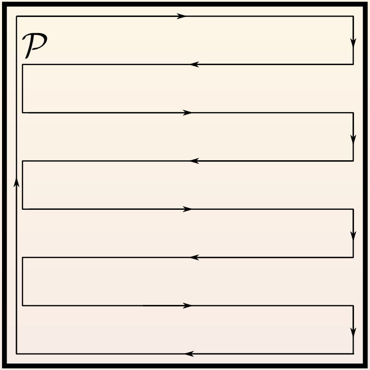
Theorem

Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ denote n sets of points, each having cardinality k, and suppose that all nk points are distributed independently and uniformly at random in a region $\mathcal{R}$ having area 1. Assume that $k \geq 1$ is fixed. Then the expected length of a generalized TSP tour of $\mathcal{X}_1, \ldots, \mathcal{X}_n$ satisfies

$$\begin{aligned} \mathbf{E}\,\mathrm{GTSP}(\mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \mathcal{O}(\sqrt{n/k}) \\ \mathbf{E}\,\mathrm{GTSP}(\mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \Omega(\sqrt{n/k}) \end{aligned}$$

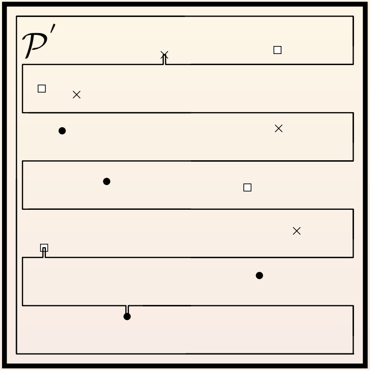
as $n \to \infty$.

Upper bound proof sketch



The path zig-zags m times, thus the length is $m+2$; here $m=8$

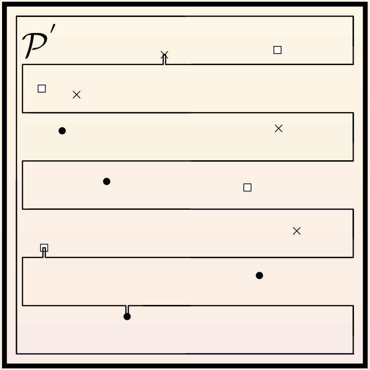
Upper bound proof sketch



Expected detour to visit a point is

$$\frac{1/(m-1)}{k+1}$$

Upper bound proof sketch



Total expected distance is

$$\underbrace{m+2}_{\text{original path}} + \underbrace{n \cdot \frac{1/(m-1)}{k+1}}_{\text{diversions}} \implies m^* \approx \sqrt{\frac{n}{k+1}} \implies \text{Total length} \propto \sqrt{\frac{n}{k}}$$

Lower bound lemma

Discretize everything, and deal with a lattice:

Theorem

Let $\mathscr{L} \subset \mathbb{Z}^2$ denote an $m \times m$ square integer lattice in the plane, let $n \geq 2$ be an integer, and let $\ell > 0$. Let $\mathscr{P}$ denote the set of all paths of the form $\{x_1, \ldots, x_n\}$, with $x_i \in \mathscr{L}$ for each i, and whose length does not exceed ℓ. Then

$$|\mathscr{P}| \leq m^2 \cdot \binom{\ell + n - 1}{n - 1} \cdot \left(\frac{8\ell}{n-1}\right)^{n-1} .$$

The generalized TSP, limiting case 2

Theorem

Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ denote n sets of points, each having cardinality k, and suppose that all nk points are distributed independently and uniformly at random in a region $\mathcal{R}$ having area 1. Assume that $n \geq 2$ is fixed. Then the expected length of a generalized TSP tour of $\mathcal{X}_1, \ldots, \mathcal{X}_n$ satisfies

$$\begin{aligned} \mathbf{E}\,\mathrm{GTSP}(\mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \mathcal{O}\left(\sqrt{\frac{n}{k^{n/(n-1)}}} \cdot (n^2 \log k + \log n)^{\frac{1}{2(n-1)}}\right) \\ \mathbf{E}\,\mathrm{GTSP}(\mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \Omega\left(\sqrt{\frac{n}{k^{n/(n-1)}}}\right) \end{aligned}$$

as $k \to \infty$.

This appears more relevant to us because we usually have $k \gg n$; numerical simulations suggest

$$\mathbf{E}\,\mathrm{GTSP}(\mathcal{X}_1, \ldots, \mathcal{X}_n) \approx \alpha\sqrt{n/k^{n/(n-1)}} = 0.29\sqrt{n/k^{n/(n-1)}}$$

A simple example

- City has area 1 and population N people
- Each person has n errands to do daily (bank, groceries, etc.):
  - A ***luddite*** performs all of their tasks by themselves and drives to each of the n locations
  - An ***early adopter*** visits $n-1$ locations and uses a delivery service for the remaining task
- There are pN early adopters in the city and $(1-p)N$ luddites

Emissions due to luddites

- Each luddite drives to n different locations, with k choices of each, thus their contribution is:

$$\psi(1-p)N\alpha\sqrt{n/k^{n/(n-1)}}$$

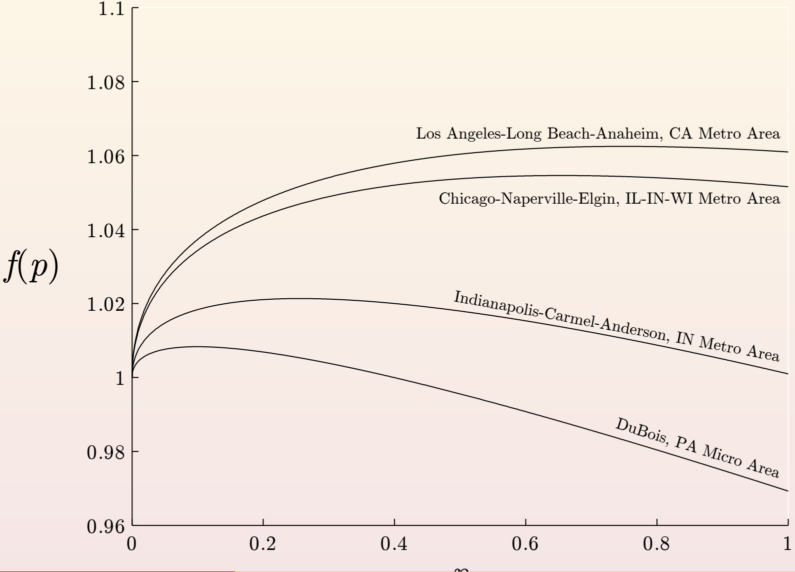
  where ψ is the CO_2/mile of their cars (we'll use $\psi = 350 \frac{\text{grams } CO_2}{\text{mile}}$)
- Each early adopter drives to $n-1$ different locations, with k choices of each, and there is also a delivery truck that visits all early adopters with a TSP, thus their contribution is

$$\underbrace{\phi\beta_2\sqrt{pN}}_{\text{delivery truck}} + \psi pN\alpha\sqrt{(n-1)/k^{(n-1)/(n-2)}}$$

  where ϕ is the CO_2/mile of a delivery truck (we'll use $\phi = 1303 \frac{\text{grams } CO_2}{\text{mile}}$)
- The overall carbon footprint is approximated by the sum of these terms:

$$\psi(1-p)N\alpha\sqrt{n/k^{n/(n-1)}} + \phi\beta_2\sqrt{pN} + \psi pN\alpha\sqrt{(n-1)/k^{(n-1)/(n-2)}}$$

Carbon footprint



Critical thresholds

Region	k	N	p^*				
			$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
Los Angeles CA	3358	13052921	> 1	> 1	> 1	> 1	> 1
Salt Lake City, UT	192	1123712	> 1	> 1	> 1	0.98	0.96
Tulsa, OK	136	951880	> 1	0.98	0.81	0.76	0.75
Albuquerque, NM	119	901700	> 1	0.86	0.72	0.68	0.68
El Paso, TX	138	830735	> 1	> 1	0.95	0.89	0.88
Colorado Springs, CO	83	668353	> 1	0.72	0.62	0.60	0.60
Boise City, ID	73	637896	0.98	0.64	0.55	0.54	0.54
Provo-Orem, UT	50	550845	0.64	0.44	0.40	0.39	0.40
Green Bay, WI	43	311098	0.90	0.64	0.59	0.58	0.60

A model "correction"

- Each person's tour is ***not quite*** a GTSP: they have to start at their house
- Let's study $\mathrm{GTSP}(\{x_0\}, \mathcal{X}_1, \ldots, \mathcal{X}_n)$:

Theorem

Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ denote n sets of points, each having cardinality k, and suppose that all nk points are distributed independently and uniformly at random in a region $\mathcal{R}$ having area 1. Assume that $n \geq 1$ is fixed. Then the expected length of a generalized TSP tour of $\{x_0\}, \mathcal{X}_1, \ldots, \mathcal{X}_n$ satisfies

$$\begin{aligned} \mathbf{E}\,\mathrm{GTSP}(\{x_0\}, \mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \mathcal{O}(\sqrt{n/k} \cdot \sqrt{\log k}) \\ \mathbf{E}\,\mathrm{GTSP}(\{x_0\}, \mathcal{X}_1, \ldots, \mathcal{X}_n) &\in \Omega\left(\sqrt{n/k}\right) \end{aligned}$$

as $k \to \infty$.

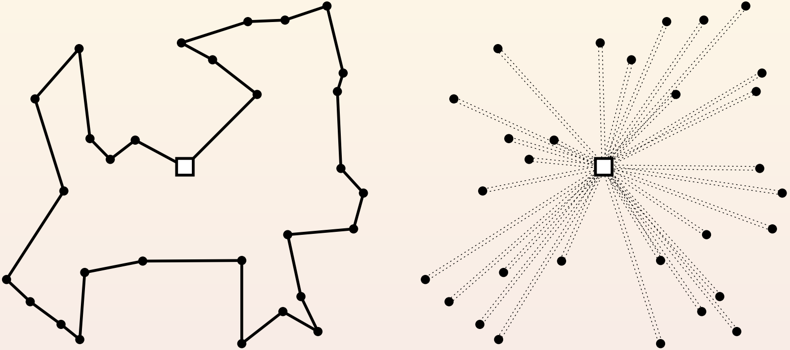
Numerical simulations suggest

$$\mathbf{E}\,\mathrm{GTSP}(\{x_0\}, \mathcal{X}_1, \ldots, \mathcal{X}_n) \approx \alpha' \sqrt{n/k} = 0.47\sqrt{n/k}$$

Revised critical thresholds

Region	k	N	p^*				
			$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
Los Angeles, CA	3358	13052921	0.10	0.14	0.17	0.20	0.23
Salt Lake City, UT	192	1123712	0.07	0.09	0.11	0.14	0.16
Tulsa, OK	136	951880	0.06	0.08	0.10	0.11	0.13
Albuquerque, NM	119	901700	0.06	0.07	0.09	0.11	0.12
El Paso, TX	138	830735	0.07	0.09	0.11	0.13	0.15
Colorado Springs, CO	83	668353	0.05	0.07	0.08	0.10	0.12
Boise City, ID	73	637896	0.05	0.06	0.08	0.09	0.11
Provo-Orem, UT	50	550845	0.04	0.05	0.06	0.07	0.09
Green Bay, WI	43	311098	0.06	0.08	0.09	0.11	0.13

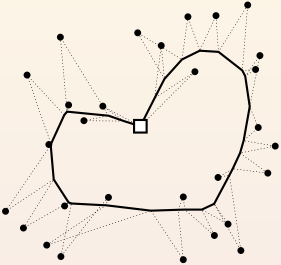
Drones



Many of the benefits and shortcomings of drone-based package delivery are obvious:

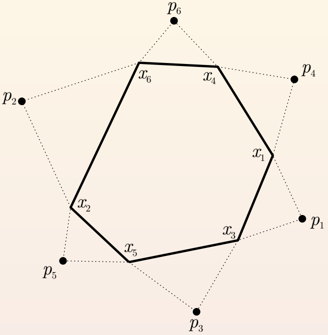
- Very cheap per-mile cost, can operate without human intervention, unaffected by road traffic
- Extremely low carrying capacity and short travelling radius

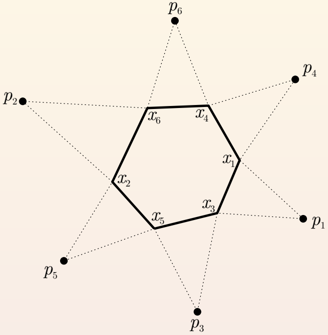
The "horsefly"



- Developed by AMP Electric Vehicles and University of Cincinnati
- Drone picks up a package from the truck, which continues on its route, and after a successful delivery, the UAV returns to the truck to pick up the next package

The "horsefly routing problem"

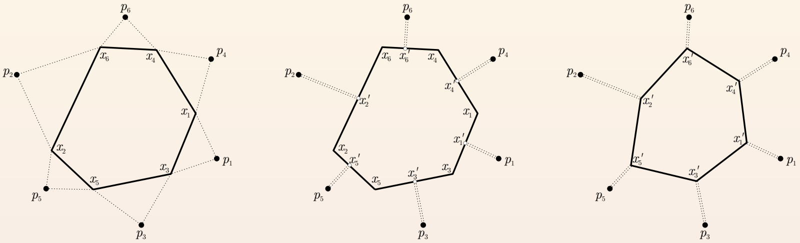




$$\underset{x_1,\ldots,x_n,\sigma\in S_n}{\text{minimize}} \sum_{i=1}^{n} \max\left\{ \frac{1}{\phi_0}\|x_{\sigma(i)} - x_{\sigma(i+1)}\|,\ \frac{1}{\phi_1}\left(\|x_{\sigma(i)} - p_{\sigma(i)}\| + \|p_{\sigma(i)} - x_{\sigma(i+1)}\|\right)\right\}$$

- $p_1, \ldots, p_n$ are customers; $x_1, \ldots, x_n$ are "launch sites"; ϕ_0, ϕ_1 are the speeds of the truck and drone
- Harder than TSP because we have to select launch sites

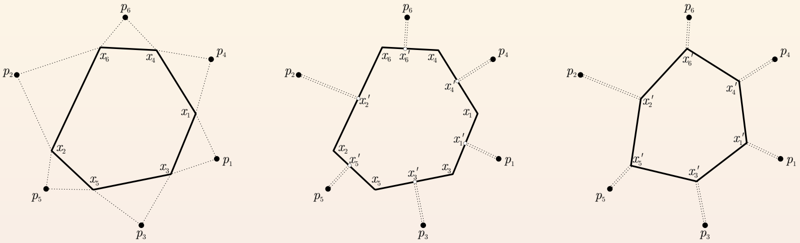
A lower bound



Exchange the summation and the $\max\{\cdot,\cdot\}$:

$$\underset{x_1,\ldots,x_n,\sigma\in S_n}{\text{minimize}} \max\left\{\frac{1}{\phi_0}\sum_{i=1}^{n}\left\|x_{\sigma(i)}-x_{\sigma(i+1)}\right\|,\ \frac{2}{\phi_1}\sum_{i=1}^{n}\left\|x_i-p_i\right\|\right\}$$

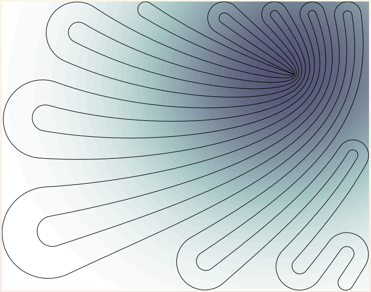
A lower bound



Take the variable over ***all loops*** $\mathcal{L}$, not the launch sites:

$$\underset{\mathcal{L}}{\text{minimize}} \max \left\{ \frac{1}{\phi_0} \text{length}(\mathcal{L}) \, , \, \frac{2}{\phi_1} \sum_{i=1}^{n} d(p_i, \mathcal{L}) \right\}$$

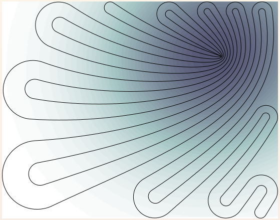
A lower bound



Assume the customers follow a continuous density f:

$$\underset{\mathcal{L} \in \mathrm{Loop}(\mathcal{R})}{\text{minimize}} \max \left\{ \frac{1}{\phi_0} \mathrm{length}(\mathcal{L}) \, , \, \frac{2n}{\phi_1} \iint_{\mathcal{R}} f(x) d(x, \mathcal{L}) dx \right\}$$

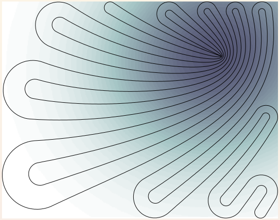
A lower bound



Theorem

$\mathrm{OPT}(\ell) \sim \frac{1}{4\ell}\left(\iint_{\mathcal{R}} \sqrt{f(x)}\,\mathrm{d}x\right)^2$ *as* $\ell \to \infty$.

A lower bound

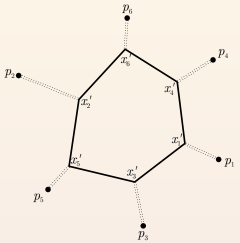


The lower bound is

$$\max\left\{\frac{1}{\phi_0}\,\mathrm{length}(\mathcal{L}^*)\,,\ \frac{2n}{\phi_1}\iint_{\mathcal{R}} f(x)d(x,\mathcal{L}^*)\mathrm{d}x\right\}\sim\sqrt{\frac{n}{2\phi_0\phi_1}}\cdot\iint_{\mathcal{R}}\sqrt{f(x)}\,\mathrm{d}x$$

as $n\to\infty$

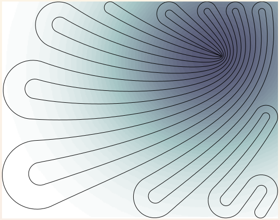
An upper bound



Just replace $\max\{\cdot,\cdot\}$ with a sum:

$$\underset{x_1,\ldots,x_n,\sigma\in S_n}{\text{minimize}} \frac{1}{\phi_0}\sum_{i=1}^{n}\|x_{\sigma(i)}-x_{\sigma(i+1)}\| + \frac{2}{\phi_1}\sum_{i=1}^{n}\|x_i-p_i\|$$

An upper bound



The upper bound is

$$\frac{1}{\phi_0}\sum_{i=1}^{n}\|x_{\sigma(i)} - x_{\sigma(i+1)}\| + \frac{2}{\phi_1}\sum_{i=1}^{n}\|x_i - p_i\| \sim \sqrt{\frac{2n}{\phi_0\phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\,\mathrm{d}x$$

as $n \to \infty$

Comparison

Our upper and lower bounds are

$$\sqrt{\frac{2n}{\phi_0\phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx$$

and

$$\sqrt{\frac{n}{2\phi_0\phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx$$

which differ from each other by a factor of 2; thus we write

$$\text{Time to perform service} \approx \beta' \sqrt{\frac{n}{\phi_0\phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx$$

for some constant β' such that $1/\sqrt{2} \leq \beta' \leq \sqrt{2}$

How much improvement?

- The BHH theorem says that if we only use a truck, then the service time will be $\beta \frac{\sqrt{n}}{\phi_0} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx$
- The improvement is therefore

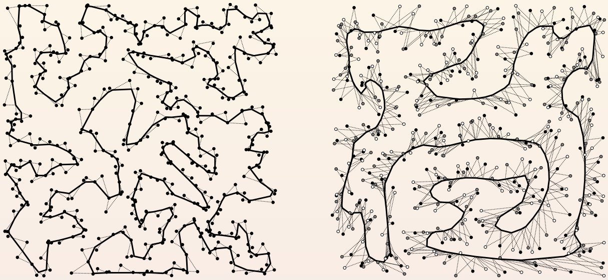
$$\frac{\text{Service time without drones}}{\text{Service time with drones}} \approx \frac{\beta \frac{\sqrt{n}}{\phi_0} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx}{\beta' \sqrt{\frac{n}{\phi_0 \phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx} = \alpha \sqrt{\frac{\phi_1}{\phi_0}}$$

with $\alpha = \beta/\beta'$ between 0.5037 and 1.0075

- If we have k drones then it's

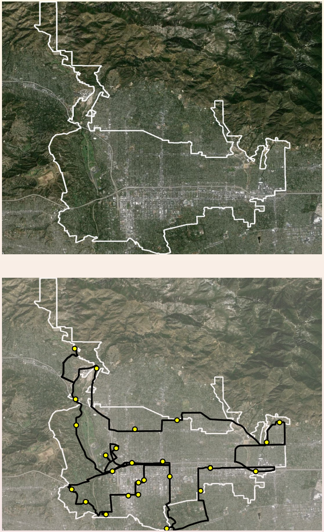
$$\frac{\text{Service time without drones}}{\text{Service time with drones}} \approx \frac{\beta \frac{\sqrt{n}}{\phi_0} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx}{\beta' \sqrt{\frac{n}{k\phi_0 \phi_1}} \cdot \iint_{\mathcal{R}} \sqrt{f(x)}\, dx} = \alpha \sqrt{\frac{k\phi_1}{\phi_0}}$$

Computational experiments, $n = 500$ points in the unit square

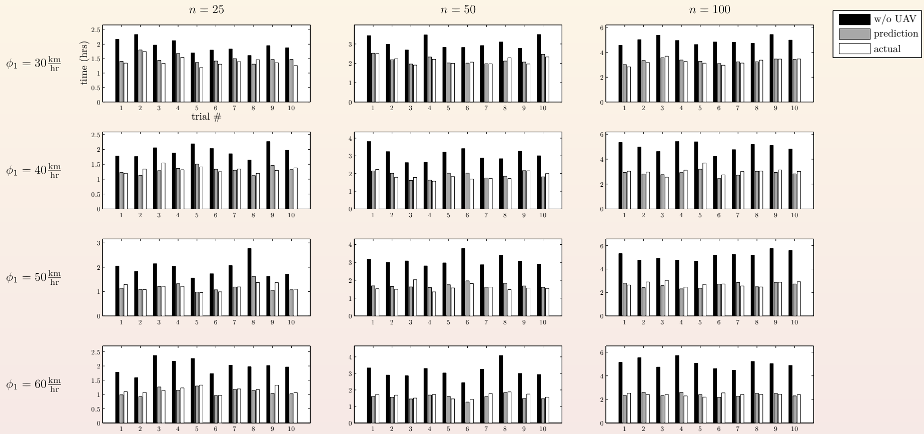


		k			
		1	2	3	5
	1.5	1.02	0.88	0.84	0.80
ϕ_1	2	1.00	0.93	0.86	0.78
	3	0.95	0.89	0.85	0.74
	5	1.02	0.92	0.83	0.80

Computational experiments, Pasadena road network



Computational experiments, Pasadena road network



Thank you!

http://www-bcf.usc.edu/∼jcarlsso/

References I