物流系统分析
Logistics Systems Analysis

模块 3　1-N 配送系统
One-to-Many Distribution System

葛乾

西南交通大学　系统所 & 交通工程系

目录 |Outline

Introduction

- We plan several lectures to address the physical distribution problems where items **produced at a single origin** are to be taken, **without transshipment**, **to a set of scattered destinations** over a service region ***R***.
- For the most part, we will focus on **delivery problems**, although it should be recognized that **collection problems** from many sources to a single destination are mathematically analogous.
- The objective is to obtain simple guidelines for the design of a set of **routes and delivery schedules** that will **minimize the total cost per unit time**.
- The CA approach for the 1-1 problem will be extended to the 1-N problem; yielding in the process simple formulae for the total resulting cost.

- The continuum approximation method is most accurate for one-dimensional point location problems if **the characteristics of the problem vary slowly along the location domain** (e.g., the time or distance line).
- The current problem is much more complex. In addition to a schedule for every customer, we must design a set of time varying routes to meet the schedule.
- It can be reduced to a point location problem in multiple (time-space) dimensions; accordingly, our solutions will be most accurate if the characteristics of the problem vary slowly over both space and time. → 本节中所涉及的问题可以约减为一个多维（时间 + 空间）中的单点选址问题。因而，只有当时空特征均缓慢变化时，所给出的解最准确。

The scenario

- A large number of destinations/customers N is distributed over a region $\boldsymbol{R}$
- The density function is a slow varying continuous function $f(\boldsymbol{x})$ of the point coordinates $\boldsymbol{x} = (x_1, x_2) \in \boldsymbol{R}$

That is, the actual number of points in a subregion of $\boldsymbol{A} \subseteq \boldsymbol{R}$, is approximately given by:

$$\int_{\boldsymbol{x} \in \boldsymbol{A}} \mathrm{Nf}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

If $f(\boldsymbol{x})$ remains nearly constant over a small $\boldsymbol{A}$, then the number can be written as:

$$\mathrm{N} \int_{\boldsymbol{x} \in \boldsymbol{A}} \mathrm{f}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx N\mathrm{f}(\boldsymbol{x}_a)|\boldsymbol{A}|.$$

where $\boldsymbol{x}_a$ is any point in $\boldsymbol{A}$, and $|\boldsymbol{A}|$ is the area of $\boldsymbol{A}$.
Note that a design approach based on expressions of this type can be used even before the actual point locations are known.

The scenario (cont.)

- In the literature, a common interpretation of $f(\boldsymbol{x})$ is as a probability density function for the coordinates of the customers, assumed to be located independently of one another.
- In that case, the above expressions represent the mean number of customers found in subareas of $\boldsymbol{R}$;
- The actual number can vary across subareas with the same mean. The standard deviation (SD) is $\{\mathrm{N}f(\boldsymbol{x}_0)|\boldsymbol{A}|[1-f(\boldsymbol{x}_0)|\boldsymbol{A}|]\}^{1/2*}$ when f(x) is nearly constant over $\boldsymbol{A}$ and points are located independently.
- The SD grows with N and $\boldsymbol{A}$ more slowly than the mean. These variations do not prevent continuous approximations to improve as N grows.
- Assume also that the cumulative number of items demanded by each customer can be expressed as a demand curve $D_n(t)(n = 1, 2, \ldots, N)$, which is assumed to vary slowly with t

*任意顾客落在区域 $\boldsymbol{A}$ 中的概率为 $|\boldsymbol{A}|$，因此区域 A 中顾客数服从二项分布。

Topics

Each topic may take one lecture.

- **Non-detailed** vehicle routing problem
- Customers with **homogeneous demand**: vehicles are filled to capacity at the depot; vehicles are not filled to capacity at the depot; detailed solution from the guidelines
- Customers with **heterogeneous demand**: symmetric strategies extended from previous discussions for the case that customers are identical; asymmetric strategies
- **Integration with production process**: adjustable production process without penalty; relaxation to handle general problems

目 录 |Outline

Non-detailed Vehicle Routing Models

- Given a set of delivery schedules for the customers in the region, one should use the vehicle routes which **minimize total distance traveled**. The toal travel distance the main determinant of transportation cost.
- It is assumed that items are distributed with identical vehicles capable of carrying $v_{\max}$ items. This definition of vehicle capacity can be used even if different item types move through the system, simply by redefining the concept of "item".
- If the maximum freight volume (or weight) that can be carried by a vehicle does not depend considerably on the mixture of item types making up its load, one can think of an "item" as a unit of volume (or weight) and of $v_{\max}$ as the vehicle's volume (or weight) capacity. Each destination can then be viewed as a consumption center for packages of unit volume (or weight) – "items" – containing an appropriate product mixture. 当一辆车辆的最大载运体积（或载重）不太依赖于所载货物的种类时，可认为一件货物对应单位体积（载重），$v_{\max}$ 为车辆的最大载运体积（或载重）。每个目的地可视为包含多种类货物合理组合的消费中心。

Vehicles are dispatched on service routes from the origin (depot) at times t_1, t_2, *etc.*, on delivery runs to particular subsets of customers (possibly the entire set each time).

- Since vehicles are identical, an operating strategy can be defined relatively easily. We seek the set t_l, as well as the **delivery lot sizes** and the **specific customers** served each "l" ; i.e., the delivery schedules for every customer.
- We also seek **the routes that minimize transportation cost** at each t_l. Our task is simplified because the combined length of all the routes is the main determinant of cost, and simple route length formulas exist

成本构成

- The cost of transportation on **one vehicle route** from one origin to several destinations was approximately a linear function of the **total size of the shipment**, **#. stops** and **the total distance traveled** (recall what we have learn in the lecture on 'Cost').
- If costs on all the vehicle routes are additive, the cost of serving all the destinations for time t_l should be the sum for the costs on each route; i.e., a linear increasing function of the total #. routes (vehicles) used, the total volume shipped, the total number of stops, and the total distance.
- For a given set of delivery schedules to each destination the total volume shipped at each t_l is obviously fixed. Thus, we only need to focus on **#. routes/vehicles, delivery stops, and vehicle-miles when seeking delivery routes for time** t_l.

- We avoid customer load-splitting among vehicles ← each destination is visited by the minimum possible number of vehicles able to hold its delivery
  - 1 vehicle if $v < v_{\max}$ items are to be delivered, and $[v/v_{\max}]^+$ otherwise *
- Although in some instances it may be possible to reduce the number of tours and the distance traveled by splitting loads†, the reductions are unlikely to be significant in most cases.
- Among all the possible strategies without load-splitting we prefer the one with the **least distance**, as this strategy should also minimize the number of vehicle routes.
  - A set of routes which minimize total distance should use vehicles to the fullest because fewer line-haul trips to and from the depot then need to be made.

---

*For customers receiving $v > v_{\max}$ items, one would dispatch $[v/v_{\max}]^-$ full vehicles exclusively to the customer, and would consolidate the remaining items with smaller deliveries to other nearby customers on a single vehicle route.

†see problem 4.1

基本假设

- Since a reasonable set of vehicle routes can be chosen on distance grounds alone, the routes can be designed **without knowing the magnitude of the cost coefficients**.
- Focusing on the difficult case when $v < v_{\max}$, the remainder of this lecture discusses **distance minimizing routing schemes** and presents simple formulas for estimating distance (and therefore transportation costs).
- The results depend on the **number of customers to be served** at time t_l, their **spatial distribution** in the region, and on the **number of stops** that vehicles can make $C = [v_{\max}/v]^-$.
- It is assumed that **the lots carried to each customer are of similar size** (a reasonable assumption for the cases with identical customers discussed in the first few sections of this chapter), so that C is the same for all vehicles; in later lectures, C will be allowed to vary.

目录 |Outline

2. The Non-detailed Vehicle Routing Models
   - Many Vehicle Tours
   - Few Vehicle Tours

主要参考文献

- Eilon, S., Watson-Gandy, C.D.T. and Christofides, N. (1971) Distribution Management: Mathematical Modelling and Practical Analysis, Hafner, New York, N.Y.
- Daganzo, C.F. (1984a) The length of tours in zones of different shapes, TR-B 18B, 135-146.
- Daganzo, C.F. (1984b) The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application, TS 18(4), 331-350.
- Newell, G.F. and Daganzo, C.F. (1986). Design of multiple vehicle delivery tours–I: A ring-radial network, TR-B 20B(5), 345-364.
- Newell, G.F. and Daganzo, C.F. (1986a). Design of multiple vehicle delivery tours–II: Other metrics, TR-B 20B(5), 365-376.
- Newell, G.F. (1986). Design of multiple vehicle delivery tours–III: Valuable goods, TR-B 20B(5), 377-390.

Many Vehicle Tours $N/C \gg C$

- In order not to introduce additional notation, we will use N to denote the number of destinations that must be visited. If tours are not being constructed for all the customers in the region, the results can be easily reinterpreted.
- Vehicles should be used to the fullest → there should be at most one vehicle that makes fewer than C stops, and none if N is an integer multiple of C.
- Our strategies are of the "**cluster-first and route-second**" type, where the service region is divided into non-overlapping zones of C customers, to be served by separate vehicles. 先划分服务区域再规划路径
- For a given set of zones, the vehicle routes are easy to construct using some simple rules. To minimize the total distance (and hence the cost), these zones should have specific shapes and orientations, dictated by the relative magnitude of N and C^2 .

Many Vehicle Tours(cont.)

- Two cases need to be considered: (i) when the number of vehicle routes N/C is much greater then the number of stops per route $C, N \gg C^2$, and (ii) when only a few vehicle routes are needed $N \ll C^2$.
- For case (i), delivery districts (or zones) should have a width comparable with the distance between neighboring points and be as long as necessary to contain C points; see Appendix A.
- The formulas are most transparent when expressed in terms of the spatial point density (#.points/unit area) evaluated at a point inside the delivery district, $\boldsymbol{x} : \delta(\boldsymbol{x}) = \mathrm{N}f(\boldsymbol{x})^*$. The factor $\delta(\boldsymbol{x})^{-1/2}$, appearing in the formulas, represents **a distance close to the average separation between neighboring points in the vicinity of $\boldsymbol{x}$.**
- For randomly scattered points, it has been found that (see Appendix A)

$$\text{zone width} \approx (6/\delta)^{1/2}; \quad \text{zone length} \approx C/(6/\delta)^{1/2}$$

*Because $\delta(\boldsymbol{x})$ varies slowly, just like $f(\boldsymbol{x})$, it does not matter which $\boldsymbol{x}$ is used

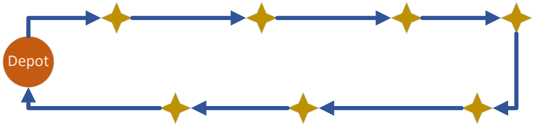
- These dimensions are close to ideal and relatively independent of the metric or underlying network.
- When δ changes over $\boldsymbol{R}$, district dimensions should also change over $\boldsymbol{R}$, although more slowly. As the solution to the EOQ problem, these expressions are robust; departures from the ideal dimensions by 20-30% are largely inconsequential, but larger departures increase distance.
- This robustness makes it easy to carve out $\boldsymbol{R}$ into delivery districts of satisfactory dimensions

服务区域的划分

- Zones should also be oriented "toward the depot", but the precise meaning of this recipe depends on the underlying metric.
- One should build equi-distance contours from the depot and design zones of the right dimensions that are perpendicular to these contours.
- For the Euclidian metric the contours are concentric circles centered at the depot, so that the zones should fan out from the depot in the radial direction.
- For the L_1 (or "Manhattan") metric, the contours are squares centered at the depot, at 45° to the metric's preferred directions （等距线应该是以出发点为形心，与量度方向，即 x, y 轴方向成 45° 夹角的一系列正方形）; in this case the zones should be perpendicular to these contours, so that they don't point exactly toward the depot. Ideal orientations can also be defined when the network includes fast/cheap roads.

服务区域的划分

- Because the zones are narrow, it is easy to construct good vehicle routes, once the region has been carved into delivery districts.
- One simply needs to *travel up one side of the zone, visiting the points in order of increasing distance to the depot, and then return along the other side visiting the remaining points in the reverse order*.
- The effectiveness of this routing scheme **improves with the slenderness of the zones –** it is exact if zones are infinitely narrow.

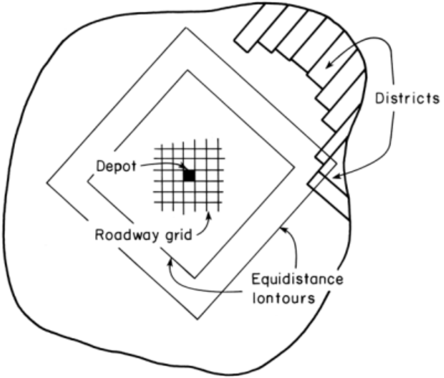


Route in L_1metric in a zone

服务区域划分的例子

Let us show how to partition a region into delivery distances with proper shape and orientation. We may draw delivery zones around the region's edge away from the depot, and then filling in the remaining space with more delivery routes, always proceeding toward the depot.

The figure depicts an intermediate point of this process for an irregular region with an internal depot and a rectangular grid network – note how most districts are perpendicular to the (square) equi-distance (L_1) contours.



As we progress toward the depot, it may become necessary to pack a few zones with the "wrong" shape, but most will have the right dimensions and orientation. Because the distance traveled is not overly sensitive to (small) deviations from the ideal design, the distance formulas about to be developed should be accurate.

单个旅程距离公式

The total distance traveled to visit the C points in a given zone containing point $\boldsymbol{x}_0$ is:

$$\text{tour distance} \approx 2\bar{r} + [k\delta^{-1/2}(\boldsymbol{x}_0)]C,$$

where $\bar{r}$ is the average distance from the C points to the depot (on the shortest path) and k is a dimensionless constant that depends on the metric ($k \approx 0.57$ for the Euclidean metric, and $k \approx 0.82$ for the L_1 metric). See Appendix A for more details.

The first term can be interpreted as the **line-haul distance** needed to reach the center of gravity of the points in the zone from the depot, and the second term as a **local distance** that must be traveled because the points are not next to one another. Note that each stop contributes toward the total a distance comparable with the separation between neighboring points, $k\delta^{-1/2}(\boldsymbol{x}_0)$. This occurs, because the vehicle must be detoured on every leg between successive deliveries. 第一项是从仓库到路径中所有点重心的每个点长途运输距离，第二项为不相邻的点之间的短途运输距离。每个停靠点对总距离的影响是其与相邻节点的间隔距离。

- In actuality, because there are only $C-1$ such legs, the factor "C" in the previous slide should be replaced by "$C-1$". Thus, a better expression is:

$$\text{tour distance} \approx 2\bar{r} + [k\delta^{-1/2}(\boldsymbol{x}_0)](C-1).$$

- The improvement afforded by this expression, particularly obvious for $C=1$, fades in importance as C grows.
- Because the previous equation is more compact, it will be used unless C is small

总距离公式

Let us now see how the total distance over $\boldsymbol{R}$ can be expressed without regard to the detailed position of points, using a continuum approximation.
Distance can be prorated to each one of the points in the zone so that if point i (located at $\boldsymbol{x}_i$) is r_i distance units away from the depot, then:

$$\text{distance prorated to } \boldsymbol{x}_i \approx \frac{2r_i}{C} + [k\delta^{-1/2}(\boldsymbol{x}_0)] \approx \frac{2r_i}{C} + [k\delta^{-1/2}(\boldsymbol{x}_i)]$$

where the second approximate equality follows from the slow varying property of $\delta(\boldsymbol{x})$.
The total distance traveled in the region is the sum of $\boldsymbol{x}_i$ across all points:

$$\text{total distance } \approx \frac{2}{C}\sum_i r_i + k\sum_i \delta^{-1/2}(\boldsymbol{x}_i)$$

For large N, the sum can be replaced by integrals independent of the specific location of all the points:

$$\sum_i \delta^{-1/2}(\boldsymbol{x}_i) \approx \int_{\boldsymbol{R}} [\delta^{-1/2}(\boldsymbol{x}_i)] \mathrm{d}\boldsymbol{x}, \quad \text{and} \quad \sum_i r_i \approx \int_{\boldsymbol{R}} r(\boldsymbol{x})\delta(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

Thus

$$\text{total distance } \approx \int_{\boldsymbol{R}} [\frac{2}{C} r(\boldsymbol{x}) + k\delta^{-1/2}(\boldsymbol{x})] \delta(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

Note that this expression is well suited for continuum approximations because the cost in any given (small) area only depends on the local conditions

Alternative Expression

An alternative expression for the total distance is obtained after replacing $\delta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ by $Nf(\boldsymbol{x})\mathrm{dx}$, it then becomes clear that these expressions can be interpreted as the product of N and the expectation of $r(\boldsymbol{x})$ or $\delta^{-1/2}(\boldsymbol{x})$, when the probability density of position is $f(\boldsymbol{x})$. Thus, letting $E(r)$ and $E(\delta^{-1/2})$ denote these expectations, the total distance can be expressed as:

$$\text{total distance } \approx N[\frac{2E(r)}{C} + kE(\delta^{-1/2})].$$

For a uniform density, $E(\delta^{-1/2}) = \delta^{-1/2} = \sqrt{|\boldsymbol{R}|/N}$ and we can write:

$$\text{total distance } \approx N[\frac{2E(r)}{C} + k\sqrt{|\boldsymbol{R}|/N}].$$

where $|\boldsymbol{R}|$ denotes the surface area of $\boldsymbol{R}$.

Interpretation

- Independent of the specific locations, *these equations are particularly useful if cost must be estimated before the point locations are known*. In that instance, it may be reasonable to view the actual locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ as outcomes of i.i.d random variables with density $f(\boldsymbol{x})$, and interpret the following equation as the average total distance over all possible locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$

$$\text{total distance} \approx N[\frac{2E(r)}{C} + kE(\delta^{-1/2})].$$

- In any specific instance there will be some discrepancy between the equation and the actual distance — for large N most of the difference typically will arise from fluctuations in $\sum_i r_i$, which are of order $O(N^{1/2})$ and comparable to the second term. If more accuracy is desired, one should wait for the point locations to become known. Comparisons made in Hall et. al. (1994) indicate that the approximation formulas just presented are fairly accurate even if the number of stops is not the same for all tours.

目 录 |Outline

Few Vehicle Tours $N/C \ll C$

- If $C^2 \gg N$, the optimal strategy must be different from the one we just explained because zones of ideal length (approximately $\frac{C}{(6/\delta)^{1/2}}$ would be too long to fit in the service region.
- It is not too difficult to design a partition of the region that will yield a distance close to a lower bound for the optimum; i.e., a near-optimal partition.
- **The lower bound is the distance for the shortest single tour visiting all the points, beginning and ending at the depot** — the "traveling salesman problem" (TSP) tour. Before describing the partitioning strategy, we must learn some basic properties of TSP tours with many points.

Few Vehicle Tours $N/C \ll C$ (cont.)

If a region with a nearly constant density of points is partitioned into a few subregions with many points each, then the length of the shortest tour in the region is close to the sum of the optimal subregional tours.

- a grand tour can be constructed by connecting the optimal tours of the subregions with a few new legs, while at the same time deleting a like number of existing legs → 将各个子区域的最优旅程通过若干小路程相连，同时删去若干小路程，即可获得一个连接所有点的旅程
- subregional tours can be constructed from a grand optimal single tour, by connecting the broken sections of the grand tour within each subregion with legs along its boundary.→ 将穿过所有点的单个最优旅程中穿过各子区域的部分的边界通过小路程相连，即可获得各个子区域的旅程

In both cases, the original (optimal) and modified (suboptimal) tours differ in total length by no more than the **combined perimeter of all the subregions**, which is a relatively small quantity when the number of points is large. Thus, the optimal grand tour should be just about as long as all the optimal subregional tours combined.

*Karp, 1977, and Eilon et al. 1971. See Appendix A for a simple proof.

Few Vehicle Tours $N/C \ll C$ (cont.)

- This property suggests that if the density of points is constant, then the TSP tour for a subregion 1/4th the region's size (with 1/4th the points) should be about four times shorter; that is, the average distance per point should be roughly constant. Since the only distance parameter of the problem is $\delta^{-1/2}$, the distance per point for large N must be of the form: $k'\delta^{-1/2}$, where k' is a dimensionless constant, independent of region shape but dependent on the metric; k' is believed to be about 0.75 for the Euclidean metric with randomly distributed points.
- The expression also holds, with a different k', for regular arrangements of points. Note that the total tour distance can be expressed as: $k'\sqrt{N|\boldsymbol{R}|}$.

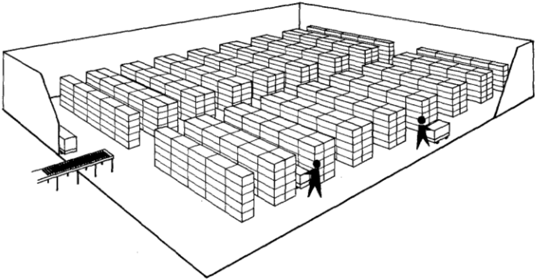
- In light of the TSP tour partitioning property, it should not matter much how the region is partitioned for the vehicle routing problem (VRP), provided that travel external to the districts is avoided by ensuring that every zone touches the depot.
- In that case each VRP tour will be similar to the TSP in the district (the TSP may not have to visit the depot), and the combined VRP length should be close to the overall TSP length; i.e., the lower bound. This means that traditional "sweep"-type algorithms for the VRP, which result in wedge shaped districts as we desire, should work well for the case with $N \ll C^2$.

- Alternatively one can build a TSP for the whole region, $\boldsymbol{R}$, and partition it into segments of C points each that would be connected to the depot.
- The length of these segments is negligible compared to the total (if $N \ll C^2$), so that the length of all the tours should be close to the length of a TSP.
- In either case, the length of all tours is close to the TSP lower bound. If the density is constant, we can write:

$$\text{total distance} \approx k' N \delta^{-1/2} = k' \sqrt{N|\boldsymbol{R}|}.$$

System with blockages

- As an aside we note that these equation, which rest on partitioning properties of TSP and VRP tours, may need to be modified for systems in which the distance metric cannot be used to define a "norm".
- An example of such a metric is a rectangular warehouse with a system of transversal aisles (横断面通道) that block travel in the longitudinal direction（纵向）, except along the sides of the rectangle.



System with blockages (cont.)

- For this type of system the length of a tour in which all the aisles with one or more service points are traversed in succession is (Kunder and Gudeus, 1975):

$$\text{total distance} \approx 2y_1 + ay_2$$

  where y_1 and y_2 are the longitudinal and transversal dimensions of the rectangle, and a is the number of aisles containing a point. y_1, y_2 分别是纵向和横向的长度，a 是包含需要访问的点的通道个数
- If N is so large that each aisle contains many points it should be clear that: (i) the traversal strategy becomes optimal and (ii) the coefficient a of the above expression can be replaced by the number of aisles.

*This shows that the above expression is not of the form since both its terms are independent of N for $N \to \infty$.

- We now return to the (usual) cases where the TSP/VRP formula can be applied, and note that for slow-varying nonuniform densities this distance expression can be approximated by the sum of expected TSP lengths over subregions with many points and (nearly) constant density. In integral form this is:

$$\text{total distance} \approx k' N E(\delta^{-1/2}).$$

  where $E(\delta^{-1/2})$ is defined previously. The uniform density is proofed to maximize the total distance; thus $k'\sqrt{N|\boldsymbol{R}|}$ is an upper bound to $k' N E(\delta^{-1/2})$.
- Notice that, unlike the many tour case, these equations are independent of C; i.e., if vehicles make so many stops that zones of ideal length cannot be packed in the service region, then travel distance is not decreased appreciably by increasing C.

- The vehicle routes within the wedge shaped zones are more difficult to develop in this case than in the previous one, which should not be surprising since the TSP problem is NP-hard.
- Nonetheless, simple algorithms such as the ones described in Daganzo (1984a) and Platzman and Bartholdi (1989) can yield tours within 20% of optimality. Simple fine-tuning corrections (see Newell and Daganzo, 1986) can then reduce its length by another 10 or 15%. Other fine-tuning approaches can yield tours even closer to optimality (see Robusté et al. 1990).
- It is not our purpose to describe here existing tour construction methods, since this is of marginal value for the theories that will follow. Suffice it to say that, in practice, it is possible to obtain tours within a few percent of optimality with an effort that only grows proportionately with the number of points to be visited.

- Now return to the one-to-many problem with identical customers.
- Recall that we are seeking the set of delivery schedules for each customer and that, given the schedules, the transportation cost at each t_l can be easily estimated with the results that have just been presented.
- The chosen schedule should strike the best balance between transportation and holding costs

目 录 |Outline

- We first consider strategies where the loads carried by each vehicle are given. Since one would then operate the smallest possible vehicles able to carry the loads, we will denote by $v_{\max}$ the load size used.
- Given $D_n(t) = D(t)$ for t in $[0, t_{\max}]$, we seek the **dispatching times** $\{t_l : l = 0, \ldots, L\}$ and **vehicle routes** which minimize the **total logistics cost**. We let $t_0 = 0$ and $t_l \leqslant t_{l+1}$.
- Because all the customers are alike, there is no compelling reason to treat some differently from others, and we shall assume that *every customer is visited with every dispatch l*. Under these conditions, the search for the t_l is facilitated considerably because, **the transportation cost only depends on the number of dispatches**, L.

Decomposition

- We now show that for a given number of dispatches L, the total transportation cost between $t = 0$ and $t = t_{\max}$ is independent of the headways: $H_l = t_l - t_{l-1}(l = 1, \ldots, L)$.
- We have already stated that the transportation cost for a given l is a linear function of # routes, # delivery stops, # items carried and the total distance.
- Clearly, the combined cost for all l must also be a function of these four descriptors. Because vehicles travel full, three of these (the **total number of items** $D(t_{\max})N$, **the number of vehicle tours** $D(t_{\max})N/v_{\max}$, and **the total number of delivery stops** NL) are fixed; they do not depend on when or how much is shipped at each t_l.

Total combined transportation cost

For a given L, the total combined distance for all dispatches is also independent of the t_l.

- As indicated by the VRP formula, it is the sum of a local distance term proportional to the total number of stops made NL, $kLNE(\delta^{-1/2})$, and a line-haul component which is proportional to the (fixed) number of vehicle tours: $2E(r) \times \#\text{tours} = 2E(r)D(t_{\max})N/v_{\max}$. Note that the line-haul component is independent of L
- With the cost coefficients, the **total transportation cost combined** between $t = 0$ and $t = t_{\max}$ is approximately:

$$c_s N \left\{ \frac{D(t_{\max})}{v_{\max}} + L \right\} + c_d kLNE(\delta^{-1/2}) + c_d 2E(r) \frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max})N,$$

  which only depends on one decision variable, L
- An expression based on few tour formula instead of the many tour one would be quite similar, and also independent of the $\{t_l\}$.

*c_s the stop cost; c_d vehicle cost for each mile traveled; c_s' added cost of carrying an extra item

Total combined transportation cost (cont.)

$$c_s N \left\{ \frac{D(t_{\max})}{v_{\max}} + L \right\} + c_d kLNE(\delta^{-1/2}) + c_d 2E(r) \frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max})N,$$

- 随着 L 不同，每个配送时间 l 各个顾客点产生的需求量不同。尽管每辆车都会用完容量 $v_{\max}$，但是所访问的顾客点数不同。这导致 local distance 的不同。
- The formula holds regardless of how many items are included in each shipping period l—even if customer lot sizes are greater than $v_{\max}$.
- It holds in particular if one decides to ship larger quantities than necessary in anticipation of future increases in the demand curves. This has a profound implication for inventory control. Given a number of shipments L to be received by a customer, their sizes and timing can be chosen to minimize holding cost without affecting the transportation cost.

目录 |Outline

3 Identical Customers and Fixed Vehicle Loads
- Very cheap items: $c_i \ll c_r$
- More expensive items: $c_i \gg c_r$
- Inventory at the origin

Very cheap items: $c_i \ll c_r$

库存/等待成本远小于租赁成本时的成本分析

- We examine first a case where items are so cheap (c_i is small) that most of the holding cost arises because of the rent paid to hold the items, $c_h \approx c_r$.
- In future lectures, with more expensive items and different customer types, the CA approach will be used to solve this problem. This is not possible now because, since the rent cost is a function of the *maximum* inventory held, said cost cannot be prorated to (small) time intervals based only on the inventories held at those times.
- Fortunately, for a given L the transportation cost is fixed, and the headways only influence the rent cost. Clearly, the headway selection problem is analogous to that examined in the 1-to-1 distribution problem. -> 给定送货次数时，运输成本为固定值，因此仅有租赁成本受到影响。

*c_r rent cost per item-year; c_i wait cost per item-unit time

Very cheap items: $c_i \ll c_r$ (cont.)

- We saw in the lot size problem with variable demand that holding cost is minimized if **all shipments are just large enough to run out before the next delivery**
- If rent costs were the dominant holding costs (so that the rent cost was proportional to the maximum lot size), then one should choose the dispatching times so as to minimize the maximum lot size →All the lot sizes should be equal, and given by $D(t_{\max})/L$.
- The same occurs here. The minimum holding cost (for L dispatching periods) is thus:

$$\text{Combined holding cost} = N[\frac{D(t_{\max})}{L}]c_r t_{\max}$$

Optimal # of dispatching times

- The total combined logistic cost consists

$$\underbrace{N[\frac{D(t_{\max})}{L}]c_r t_{\max}}_{\text{combined holding cost}} + \underbrace{c_s N\left\{\frac{D(t_{\max})}{v_{\max}} + L\right\} + c_d KLNE(\delta^{-1/2}) + c_d 2E(r)\frac{D(t_{\max})}{v_{\max}} + c_s' D(t_{\max})N}_{\text{combined transportation cost}}$$

- The optimal number of dispatching times L should be chosen by minimizing such a sum. Only the first and second terms of the transportation cost capture the local stop cost and the local distance cost and depend on L. The other terms, corresponding to the line-haul travel and the loading/handling cost do not.
- Thus, the optimal L^* is the solution of an integer constrained EOQ equation that balances the local transportation cost and the rent cost; the solution is close to:

$$L^* \approx \left[\frac{c_r t_{\max} D(t_{\max})}{c_s + c_d kE(\delta^{-1/2})}\right]^{1/2}, \text{ if } L \text{ is greater than } 1.$$

*注意书本 107 页公式 (4.9) 有误

Optimal total combined cost/item

The total combined cost per item is approximated by:

$$\frac{c_s + 2c_d E(r)}{v_{\max}} + c_s' + 2[c_r \frac{c_s + c_d k E(\delta^{-1/2})}{\bar{D}'}]^{1/2}$$

where we use $\bar{D}'$ for the average demand rate per customer, $D(t_{\max})/t_{\max}$. Remarkably, the optimal cost does not depend on the shape of $D(t)$. Not many details are needed to provide a reasonable estimate of operating cost.

3. Identical Customers and Fixed Vehicle Loads
   - Very cheap items: $c_i \ll c_r$
   - More expensive items: $c_i \gg c_r$
   - Inventory at the origin

More expensive items: $c_i \gg c_r$

- We now discuss the problems for items so expensive per unit volume that most of **the holding cost is inventory cost**. Our lectures on lot size problem showed how a CA approach could be used to locate points on the time line (the delivery times) in order to minimize approximately the sum of the holding and motion costs
- The latter was modeled by a constant c_f that represented the added cost of each dispatch. Reasonable for the one-to-one problem examined at the time, this simple formulation also applies now
- From the equation of the combined transportation cost, we notice that with each additional dispatch, the transportation cost still increases by a constant amount (对 L 求导)

$$c_f \approx [c_s + c_d k E(\delta^{-1/2})] N.$$

  This constant represents the local transportation cost induced by the N additional customer visits resulting from the extra dispatch. The line-haul cost remains unchanged.

More expensive items: $c_i \gg c_r$ (cont.)

- Consequently, the results and methods of the lot size problem for the EOQ with variable demand also apply here if one defines $c_f \approx [c_s + c_d k E(\delta^{-1/2})]N$ and replaces $D(t)$ by $ND(t)$. The CA formulation for 1-to-1 problems can then be used to estimate cost. Don't forget to add the (large) fixed components of combined transportation cost that do not depend on L
- Once the dispatch times $\{t_l\}$ and the corresponding delivery lot sizes $\{v_l\}$ have been determined, the vehicle routes can be designed as described in the non-detailed VRP, recognizing that the number of stops per vehicle $(C = n_s^l \approx v_{\max}/v_l)$ changes with l.

- For the special case with *uniform density and constant demand*, the cost formula reduces to a form analogous to formula for cheap goods, with c_i, D' and $(|\boldsymbol{R}|/N)^{1/2}$ substituted for $c_r, \bar{D}'$ and $E(\delta^{-1/2})$*.
- This approach has been used to streamline General Motors' finished product distribution procedures. The results have been compared with those of (less efficient) direct shipping strategies†.

*Burns et al. 1985
†Gallego and Simchi-Levy, 1988

目录 |Outline

3. Identical Customers and Fixed Vehicle Loads
   - Very cheap items: $c_i \ll c_r$
   - More expensive items: $c_i \gg c_r$
   - Inventory at the origin

Inventory at the origin

- The theory we have described focused on the holding cost at the destination and used cost expressions as if there were an equivalent cost at the origin.
- This assumption is reasonable for the 1-to-1 problems and is now shown also to be reasonable if the one-to-many system is operated as we described.
- However, a modification to the operating procedure can drastically reduce the origin holding costs.

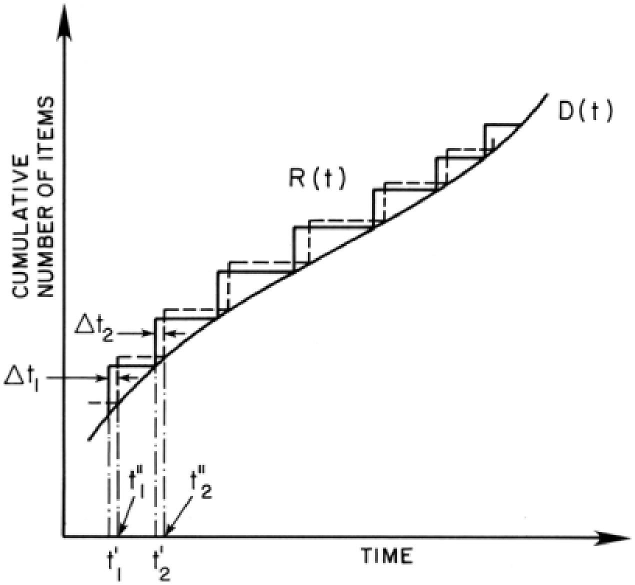
灵活生产策略下的租赁成本

- With our dispatching strategy, where all the destinations are served with each l, the number of items accumulated at the origin reaches a maximum immediately before a dispatch, and at the destinations immediately after a reception.
- If production is flexible, one will produce by dispatch l only those items that must be sent by time t_l (and no more) ; thus, the maximum accumulation at the origin is the size of the largest shipment received by any customer, times N.
- Because shipments arrive as supplies run out, this is also the maximum accumulation for all the customers. It is thus reasonable to represent rent cost by the product of a constant, c_r, and the maximum accumulation, as we have done.

灵活生产策略下的库存成本

- Inventory costs are slightly different. If one could produce the items as fast as desired, one would produce item during a short time interval prior to t_l for each combined shipment l; and would therefore avoid inventory costs at the origin. This is not likely to happen, however.
- Although the production rate can change with time to satisfy a slow varying demand $D(t)$, items are normally produced at a roughly uniform rate during each inter-dispatch interval, since most production processes benefit from a smooth production curve.
- Thus, inventory costs should not be reduced in this manner. If some destinations request more expensive items than others, then inventory cost may be reduced without altering the production rate, simply by changing the order of production. One might want to produce the cheap items at the beginning of the inter-dispatch interval and the most expensive at the end.

- In most cases, however, only a fraction of the inventory cost at the origin could be saved by exploiting these differences.
- Thus, the waiting cost at the origin should be comparable to the waiting cost at the destinations, and a strategy which assumes that both holding costs are equal should yield costs close to one which recognizes the inventory cost at the origin more accurately
  - Remember that an error in a cost parameter by a factor of 2 only increases the resulting EOQ cost by about 10%.



CUMULATIVE NUMBER OF ITEMS
D(t)
R(t)
Δt2
Δt1
t''1
t''2
t'1
t'2
TIME

Staggering production for delivery regions (交错生产策略)

- With our operating strategy, all the points in the region $\boldsymbol{R}$ are visited at each instant l.
- However, if instead of waiting for time t_l, vehicles are dispatched just as soon as their last item is produced, both the storage room and the inventory cost at the origin may be reduced. As shown below, this reduction is largest if one can produce all the items for each one of the delivery districts, in sequence.

交错生产策略

- If the delivery times to any customer are shifted by a time Δt_l smaller than one headway (i.e., the new delivery times are $t_l' = t_l - \Delta t_l > t_{l-1}$), and if Δt_l changes slowly with l so that the new headways are close to the old, then the total holding cost does not change appreciably.
- With a slow varying $D(t)$, the maximum accumulation remains virtually unchanged, and so does the total number of items-hours; see the difference between the solid and dotted $R(t)$ curves.
- This is consistent with the CA solution; the cost is sensitive to the delivery headways used as a function of time but much less so to the specific dispatching times.

交错生产策略

- Suppose that we label the tours used for the l-th shipment: $j = 1, 2, 3$, etc.
- Assume that items for destinations in tour $j = 1$ are produced first, items for destinations in $j = 2$ second, etc; and assume as well that every tour is started as soon as the orders for its customers have been completed.
- If the delivery districts do not change with every l, it would be possible to label them consistently so that all destinations would have the same label in successive dispatches. This would ensure that the l-th delivery headway to every customer is close to $(t_{l+1} - t_l)$, and that as a result the holding cost at all the destinations would remain essentially unchanged.
- The ordered production schedule, though, would cut the maximum and average inventory at the origin by a factor equal to the number of tours used for the l-th shipment, drastically reducing holding costs at the origin.

交错生产策略

Unless the demand is constant, $D(t) = \lambda t +$ constant, it is not reasonable to assume that all the delivery districts remain the same; in that case a less ambitious version of our staggered production schedule can be employed.

- The service region can be partitioned into production subregions $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_P$, where P is a number small compared with the number of tours in any l, but significantly larger than 1 (so that it can make a difference.)
- Each production subregion should contain the same number of customers (i.e., the same total demand) and require at least several tours to be covered. Under such conditions, the distance for covering $\boldsymbol{R}$ with a VRP is not much different from the collective distance of separate VRP's to cover $\boldsymbol{P}_1, \boldsymbol{P}_2$ etc.
- This is true because, like the TSP, the VRP exhibits a partitioning property. (This should be obvious, since: (i) the cost in each subregion is the sum of the costs prorated to each of its points, and (ii) the cost per point is independent of the partition).

交错生产策略

The following strategy cuts inventories at the origin by a factor P , while preserving virtually unchanged the motion and holding costs at the destination:

1. produce the items for any shipment in order of production subregion: $\boldsymbol{P}_1$ first, then $\boldsymbol{P}_2$, etc
2. On completing production for a subregion, $\boldsymbol{P}_P$, dispatch the vehicles to the subregion on VRP routes constructed for the subregion alone

交错生产策略

- As a practical matter, P does not need to be very large; once it reaches a moderate value (say $P \approx 5$) additional increases yield decreasingly small benefits.
- In fact, even if the demand was perfectly constant, it is unlikely that one would choose a P much larger than 5 because larger P's imply shorter production runs within each $\boldsymbol{P}_p$, which hinders our ability to sequence the production to meet other objectives, such as operating with smoothing worker loads and materials requirements. P 的大小决定了工厂面向不同子区域生产计划的不同，如平滑的员工工作量或者材料要求。P 越大，生产计划变化地越频繁，反而不利于工厂生产。

- If production schedules are staggered as described, then the search for the optimal dispatching times should recognize that holding costs will be lower.
- The analysis could be repeated with a changed holding cost equation (e.g., combined holding cost $= N\frac{D(t_{\max})}{L}c_r t_{\max}$ for the case $c_r \gg c_i$) but this is unnecessary; a suitable (downward) adjustment to the holding cost coefficient, either c_i or c_r, has the same effect and also preserves our results.
- If holding costs at the origin can be neglected, the coefficient should be halved; of course, there is no need to pinpoint its value very precisely, since the solution to our problem is robust to errors in the cost coefficients.

目 录 |Outline

车辆载重未定带来的问题

- In every case discussed so far, the total cost expression decreases with the vehicle load carried v_{max} ← the larger v_{max} the smaller the total number of vehicles that need to be dispatched. → In any practical, one would be well advised to use vehicles as large as the (highway, railway ...) network would allow.
- However, the analysis ignored *pipeline inventory cost* and did not consider possible *route length restrictions*. With either one of these complications, it may not always be desirable (or possible) to dispatch full vehicles all the time; vehicle load size becomes a decision variable.

内容构成

- We will discuss route length restrictions first, and will then incorporate pipeline inventory into the models.
- It will be shown that pipeline inventory cost can be ignored for freight that is neither perishable nor extremely valuable, and that it cannot be ignored for passengers.
- Were it not for this complication, the results for fixed vehicle load problems could be used for 1-to-N passenger logistics (e.g., to design a commuter rail network serving a CBD).
- We concludes with a discussion of restrictions on the delivery lot size.

Limits to Route Length

- If the optimization of the identical customer problem results in **very small delivery lot sizes**, each vehicle may have to make an **unreasonably large number of stops**.
- Very long routes may not be feasible if there are restrictions to the duration of a vehicle tour. For example, due to **labor regulations**
- We may explore the consequences of such restrictions

Limits to route length (cont.)

- Tour duration limitations essentially impose a **location-dependent limit** on the number of stops.
- Presumably, locations distant from the depot will need to be served with fewer stops than those which are nearer since more time is needed to reach their general vicinity.
- To recognize this dependence, we use $C_{\max}(\boldsymbol{x})$ for the maximum number of stops around $\boldsymbol{x}$; we assume that $C_{\max}(\boldsymbol{x})$ varies slowly with $\boldsymbol{x}$

Limits to route length (cont.)

- Assume first that N is large, so that most delivery districts do not reach all the way to the depot. Then, to minimize distance one should still attempt to design delivery districts of **width** $[6/\delta(\boldsymbol{x})]^{1/2}$, while making them long enough to include a desired number of stops at (or near) coordinate $\boldsymbol{x}$, $n_s(\boldsymbol{x}) < C_{\max}(\boldsymbol{x})$. This yields: **length** $= n_s(\boldsymbol{x})/[6\delta(\boldsymbol{x})]^{1/2}$. The total distance is then given by expressions

$$\text{Total distance} \approx [\sum_i \frac{2r_i}{n_{s,i}} + k\delta^{-1/2}(\boldsymbol{x}_i)]$$
$$\approx 2NE(\frac{r}{n_s}) + kNE(\delta^{-1/2})$$

  where $n_{s,i}$ denotes the number of stops per tour used for tours near $\boldsymbol{x}_i$;
- if $n_s(\boldsymbol{x}) \equiv C$, it coincides precisely with previous expressions. Although the line-haul distance component (the first term) is somewhat different if $n_s(\boldsymbol{x})$ varies with $\boldsymbol{x}$, *the local component remains unchanged*.

Total distance with restrictions on the route length (cont.)

$$\text{Total distance} = 2NE(\frac{r}{n_s}) + kNE(\delta^{-1/2})$$

This expression decreases with $n_{s,i}$ $\rightarrow$ # stops per tour should be made as large as practicable.

For our problem, # stops used near location $\boldsymbol{x}$ on the l-th dispatch, $n_s^l(\boldsymbol{x})$, should satisfy:

$$n_s^l(\boldsymbol{x}) = \min\{C_{\max}(\boldsymbol{x}); v_{\max}/v_l\},$$

where v_l denotes the delivery lot size used for period l. The expression indicates that the vehicle either reaches its **route length constraint**, or else is **filled to capacity**.
运输批量的两个上界：点 $\boldsymbol{x}$ 处允许的最大值，容量

Dependence on the specific headways?

- With this restriction some of the tours may carry less than a full load. As a result, it may appear that neither the total number of vehicle tours nor the line-haul transportation cost (长途运输的距离是 $2NE(\frac{r}{n_s})$) are fixed.
- We shows that, while not fixed, **the number of tours** (and thus the sum of the line-haul and stop costs) **can sometimes be approximated by an expression that only depends on the number of headways** L; then, the scheduling and routing decisions can still be decomposed.

Approximation for the number of tours

- Assume that $\boldsymbol{R}$ can be partitioned into just a few subregions, $\boldsymbol{P}_p$, with the same limitation on the number of stops: $n_s(\boldsymbol{x}) < C_{\max}(\boldsymbol{x}) \approx C_p$. Characterize each subregion by the number of destinations N_p, and their average distance to the depot $E(r_p)$.
- We will show that the number of tours in each subregion only depends on L. As a result, an expression for the total number of tours is developed.
- The number of tours in period l for subregion p is:

$$\{\# \text{ tours}; l, p\} = \max\left\{\frac{N_p v_l}{v_{\max}}; \frac{N_p}{C_p}\right\},$$

and for all periods:

$$\{\# \text{ tours}; p\} = N_p \sum_{l=1}^{L} \max\left\{\frac{v_l}{v_{\max}}; \frac{1}{C_p}\right\} \geqslant N_p \max\left\{\sum_{l=1}^{L} \frac{v_l}{v_{\max}}; \frac{L}{C_p}\right\}$$

需求接近恒定

- This inequality is a good approximation for the number of tours if rent costs dominate, as then the delivery lot size should be independent of l.
- The approximation will also be good, for the same reason, if the demand is nearly stationary. Then, we can write:

$$\{\#\ \text{tours};p\} \cong N_p \max\left[\frac{D(t_{\max})}{v_{\max}}; \frac{L}{C_p}\right] = N_p\left\{\frac{D(t_{\max})}{v_{\max}} + \max\left[0, \frac{L-L_p}{C_p}\right]\right\}$$

where $L_p = C_p D(t_{\max})/v_{\max}$.

- L_p represents a critical number of dispatching periods for subregion p. If $L > L_p$, then the lot sizes are so small that the vehicle cannot be filled in subregion p; the number of stops constraint is binding. L 越大，表示每两次配送时间间隔产生的需求量越小；如果车辆满载，则访问的顾客数会非常多，此时行程数量受最大距离限制。

† L_p 表示车辆在 p 区域所能服务的配送次数临界值。

Total transportation cost

- If the equation is a good approximation for # tours used in P_p, then the sum of **the origin stop cost plus the line-haul cost** for all tours is:

$$\begin{aligned}&\sum_{p=1}^{P}\{\#\text{ tours}; p\}[c_s + 2c_d E(r_p)]\\ &= \frac{D(t_{\max})}{v_{\max}} N[c_s + 2c_d E(r)] + \sum_{p=1}^{P}[c_s + 2c_d E(r_p)]\left[N_p \max\left(0, \frac{L - L_p}{C_p}\right)\right]\end{aligned}$$

  which only depends on the dispatching times through L.

- For small L the expression is constant, and matches the sum of the 1st and 3rd term of the total combined transportation cost. But once L exceeds some of the L_p (some tours hit the length constraint and are only partially filled), it increases with L at an increasing rate.

† Recall the expression for combined transportation cost: $c_s N\{\frac{D(t_{\max})}{v_{\max}} + L\} + c_d kLNE(\delta^{-1/2}) + c_d 2E(r)\frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max})N$。该公式对应于总运输成本的第一项(除去 $c_s NL$ 部分)和第三项

Find the optimal dispatching time L

- The optimal L can be found still as a trade-off between inventory cost* and transportation cost†, with the first and third terms revised.
- Because the revised combined transportation cost equation is piecewise linear and convex, the sum of the inventory cost and transportation cost has only one local/global minima. The revised derivative of total combined transportation cost with respect to L is now a step function:

$$\left[c_s + c_d k E(\delta^{-1/2})\right] N + \sum_{L_p < L} \frac{N_p}{C_p} \left[2 c_d E(r_p) + c_s\right],$$

where the summation only includes p's for which $L_p < L$. The second term represents **the cost increase for the extra tours that need to be sent because (some) vehicles cannot be filled to capacity**. The first term keeps unchanged.

---

*Recall the expression for combined holding cost: $N[\frac{D(t_{\max})}{L}]c_r t_{\max}$

†Recall the expression for combined transportation cost: $c_s N\{\frac{D(t_{\max})}{v_{\max}} + L\} + c_d k L N E(\delta^{-1/2}) + c_d 2E(r)\frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max})N$

Find the optimal dispatching time L (cont.)

$$\left[c_s + c_d k E(\delta^{-1/2})\right] N + \sum_{L_p < L} \frac{N_p}{C_p} \left[2c_d E(r_p) + c_s\right],$$

- In the special case where C_p is the same ($C_{\max}$) for all points, there is only one subregion, with $L_1 = C_{\max} D(t_{\max})/v_{\max}$ and $N_1 = N$. Therefore, the second term is zero if $L \leqslant C_{\max} D(t_{\max})/v_{\max}$, and equals $(N/C_{\max})(2c_d E(r) + c_s)$ otherwise.
- The optimal L can be found as follows: If there is a value of L for which the sum of this equation and the derivative of combined holding cost equals zero, then that value is optimal; otherwise, the optimal value is the L_p for which the sum changes sign.
- Because the derivative is larger than before, the optimal L will tend to be smaller and the resulting cost greater. This is intuitive; with limits to route length it may be advisable to increase the lot sizes (by reducing L) to make sure that most of the vehicles travel full.

C_p 随着子区域变化显著的情形

- Our results assume that all customers share the same L and v_l. Although this simplification facilitates production scheduling, it may also **increase logistics costs** when C_p changes significantly across subregions.
- If a different L can be used for different subregions, then **fewer dispatching intervals and larger delivery lot sizes** can be used for subregions with a low C_p; all the vehicles can be filled as a result. A strategy (a set of dispatching times and delivery districts) can then be tailored to each one of the subregions independently of the others. 对于 C_p 不均匀的区域，可为各个子区域设计不同的配送策略；C_p 越小，说明子区域 p 能访问的最大点数越小，可以通过降低配送频率和增大运输批量使得服务该区的车辆被装满。
- We will explore this point — the determination of routing/dispatching strategies that vary in time and space –more thoroughly in the following talks.

Route length restriction for few vehicle tours

- To conclude our discussion on route length restrictions, we must consider the case with few vehicle tours, $N \ll C^2$
- Very simple. The transportation cost is insensitive to # stops per vehicle for this case.
- Hence, route length restrictions do not influence either the optimal dispatching strategy or the final cost.

目录 |Outline

Accounting for Pipeline Inventory Cost

- In all the optimization problems described so far we have found a solution which minimizes the sum of the motion cost, the holding (rent) cost and the stationary inventory cost. We did not consider the pipeline inventory cost of the items in the vehicles.
- Recall that the pipeline inventory cost/item was $c_i t_m$, where t_m is the average time an item spends inside a vehicle.
- On average an item spends in a vehicle a time approximately equal to one-half of the duration of the tour. If the vehicle travels at a speed s, and takes t_s time units per stop, the duration of a tour with n_s stops and d distance units long is $d/s + (n_s + 1)t_s$; thus:

$$t_m \approx \frac{1}{2}\left[\underbrace{d/s}_{\text{途中时间}} + \underbrace{(n_s+1)t_s}_{\text{停靠时间}}\right], \text{ and } c_i t_m \approx \frac{1}{2}c_i \times \frac{d}{s} + \frac{1}{2}c_i t_s(n_s+1)$$

Compositions of the pipeline inventory cost

- Added for all items for all L shipping periods, the pipeline inventory cost becomes approximately a simple function of the total number of (item-miles), (items) and (item-stops):

$$\frac{c_i}{s} \times \# \text{ item-miles} + \frac{c_i t_s}{2} \times \# \text{ items} + c_i t_s \times \# \text{ item-stops}$$

- The total number of items is $D(t_{\max})N$. The total number of item-miles and item-stops can be obtained easily if there are no route length restrictions.
- In that case vehicles travel full (from the depot) and every stop delays on average $v_{\max}/2$ items; therefore, the total number of item-stops is $NLv_{\max}/2$. Similarly, each vehicle carries on average $v_{\max}/2$ items and the item-miles equal the product of the vehicle-miles and $(v_{\max}/2)$.

Compositions of the pipeline inventory cost (cont.)

- We have already seen that the total vehicle-miles are (recall the derivation for the total combined distance):

$$\frac{2E(r)D(t_{\max})N}{v_{\max}} + kNLE(\delta^{-1/2})$$

- Thus, the pipeline inventory cost can also be expressed as a function of the decision variables through L alone:

$$\left[\frac{c_i E(r)D(t_{\max})N}{s}\right] + \left[\frac{c_i KN}{2s} E(\delta^{-1/2}) v_{\max}\right] L + \frac{c_i t_s}{2}[D(t_{\max})N + N v_{\max} L]$$

这里仅仅计算总旅行长度，并代入 ‘item-mile’ 取值。

† 注意课本 116 页 ‘total combined distance’ 公式有误。

管道仓储成本 vs 运输成本

$$\frac{c_i E(r) D(t_{\max}) N}{s} + [\frac{c_i K N}{2s} E(\delta^{-1/2}) v_{\max}] L + \frac{c_i t_s}{2}[D(t_{\max}) N + N v_{\max} L]$$

- As a function of L, this expression is similar to the equation for transportation cost*, but it increases much more slowly: at a rate $N[c_i v_{\max}/2][t_s + kE(\delta^{-1/2})/s]$ as opposed to $N[c_s + c_d kE(\delta^{-1/2})]$
- Normally, the quantity $c_i v_{\max} t_s$ represents the cost of delay to the items in a full vehicle during a stop. It should be several orders of magnitude smaller than c_s (the truck cost and driver wages during the stop).
- Likewise, the quantity $c_i v_{\max}/s$ represents the inventory cost of a full truck per unit distance. It should be much smaller than c_d (the vehicle operating cost per unit distance, including driver wages).
- Thus, if pipeline inventory costs had been considered from the beginning, the results would not have changed.

*Recall the expression for combined transportation cost: $c_s N\{\frac{D(t_{\max})}{v_{\max}} + L\} + c_d kLNE(\delta^{-1/2}) + c_d 2E(r)\frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max}) N$

管道仓储成本不能忽略的情形

- If the items are so expensive that the pipeline inventory component cannot be neglected, then **the pipeline inventory cost, unlike the transportation cost, increases with $v_{\max}$.**
- One could thus imagine a situation where a $v_{\max}$ smaller than the maximum possible might be advantageous; the vehicle loads cannot be assumed to be known. **The transportation of people is a case** in point, where the inventory cost of the items carried (the passengers) vastly exceeds the operating cost.
- That is why airport limousine services do not distribute people from an airport to the hotels in the outlying suburbs in large buses; this would result in unacceptably large routes, with some passengers spending too much time in the vehicle *. 机场礼车服务为什么不使用大型车将顾客从机场拉到远郊的各个宾馆？因为这会使行程长的不可接受，部分旅客在车上花的时间太多。

*See Banks, et al. 1982, for a discussion

Re-examine the total transportation cost

- Let us now see how to select the routes and schedules for a system carrying items so valuable that vehicle loads are not necessarily maximal.
- Without an exogenous vehicle load, the total transportation cost no longer can be expressed as a function of L alone; the total vehicle-miles and the number of tours depend on the specific vehicle-loads used, and this has to be recognized in the optimization.
- To cope with this complication, we will consider a set of strategies more general than the ones just examined, but will analyze them less accurately.

Headway

- We will now **allow different parts of *R* to be served with different delivery headways at the same time**. To do this, we define the smooth and slow varying function $H(t, \boldsymbol{x})$, which represents the headways one would like to use for destinations near $\boldsymbol{x}$ at times close to t.
- Until now we had assumed that the headways were only a function of t: $H(t, \boldsymbol{x}) = H(t)$. As a result, the optimal dispatching times $\{t_l\}$ could be found with the exact numerical techniques; or if $D(t)$ was slow varying, with the CA approach.

stops per tour

- For the present analysis we also seek a function $n_s(t, \boldsymbol{x})$ which indicates the number of stops made by tours near $\boldsymbol{x}$ at a time close to t. Of course, this number cannot be so great that the vehicle capacity is exceeded; the following must be satisfied:
$$\{n_s(t, \boldsymbol{x}) D'(t)\} H(t, \boldsymbol{x}) \leqslant v_{\max},$$
- The quantity in braces represents the combined demand rate at the n_s destinations visited by a tour, and the left side of the inequality the load size carried by the vehicle.
- The approach we had used assumed that this equation was a pure equality, so that n_s was only a function of t, $n_s(t) = v_{\max}/[H(t)D'(t)]$, implicitly given by $H(t)$.
- Like $H(t, \boldsymbol{x})$, the function $n_s(t, \boldsymbol{x})$ will be allowed to be continuous and slow-varying during the optimization.

Decision variables

- Once $H(t, \boldsymbol{x})$ and $n_s(t, \boldsymbol{x})$ have been identified, a set of delivery districts and dispatching times consistent with these functions must be found. This will be illustrated after the optimization has been described.
- Let us write the total logistics cost per item that items at time-space point $(t, \boldsymbol{x})$ would have to pay if the parameters of the problem were the same at all other times and locations, i.e., $D'(t) = D', \delta(\boldsymbol{x}) = \delta,$ and $r(\boldsymbol{x}) = r$.
- The decision variables H and n_s that minimize such an objective function will become the sought solution, varying continuously with t and $\boldsymbol{x}$ $\big(H(t, \boldsymbol{x})$ and $n_s(t, \boldsymbol{x})\big)$.

Total motion cost per item

- The minimum value of the objective function for these coordinates $z(t, \boldsymbol{x})$, is the CA cost estimate.
- Noticing that a vehicle load consists of $D' n_s H$ items and a delivery lot of $D' H$ items, we can express the total motion cost per item as:

$$z_m = \frac{2rc_d}{D' n_s H} + c_d k \delta^{-1/2} \frac{1}{D' H} + c_s \frac{1}{D' H} + c_s \frac{1}{D' n_s H} + c_s'.$$

- Recall the expression for combined transportation (motion) cost: $c_s N[\frac{D(t_{\max})}{v_{\max}} + L] + c_d k L N E(\delta^{-1/2}) + c_d 2E(r) \frac{D(t_{\max})N}{v_{\max}} + c_s' D(t_{\max})N$. We may obtain z_m by simply putting $v_{\max} = D' n_s H, D' H = D(t_{\max})/L$ in the expression and dividing it by $D(t_{\max})N$.

Physical interpretation

$$z_m = \frac{2rc_d}{D' n_s H} + c_d k \delta^{-1/2} \frac{1}{D' H} + c_s \frac{1}{D' H} + c_s \frac{1}{D' n_s H} + c_s'.$$

This expression has an intuitive physical interpretation.

- Each tour incurs a cost $(2rc_d + c_s)$ for overcoming the line-haul distance and stopping at the origin, which prorated to all the items in the vehicle yields the first and fourth terms.
- The tour also incurs a cost $(c_d k \delta^{-1/2} + c_s)$ for each local stop and detour, which prorated to the items in a delivery lot, yields the second and third terms of the expression.
- The last term is the (constant) cost of handling each item.

Physical interpretation (cont.)

$$z_m = \frac{2rc_d}{D' n_s H} + c_d k \delta^{-1/2} \frac{1}{D' H} + c_s \frac{1}{D' H} + c_s \frac{1}{D' n_s H} + c_s'.$$

Thus, the first two terms are the cost of overcoming line-haul and local distance (assuming that many tours are needed); the third term is the cost of stopping at the destinations; the fourth the cost of stopping at the origin, and the last one the handling/loading cost.

Holding costs

- The holding costs can be expressed in a similar manner. For the pipeline inventory cost per item, we have the following expression when $C = n_s$:

$$z_p = c_i\frac{r}{s} + c_i k\delta^{-1/2}\frac{n_s}{2s} + c_i\frac{t_s}{2}n_s + \frac{1}{2}c_i t_s.$$

- As with the expression for the total motion cost, the four terms correspond to times spent in **line-haul travel, local travel, destination stops, and at the origin**. The stationary inventory cost per item averages $z_s = c_i H$ if we count it both at the origin and the destination.
- The rent cost can be ignored because if items are expensive compared to transportation costs, they will certainly satisfy $c_i \gg c_r$; thus $c_h = (c_i + c_r) \approx c_i$, and we can write $z_s = c_h H$. .

*Inclusion of rent costs would pose a problem because rent does not depend only on local characteristics such as H and n_s. An exception arises if the demand is stationary in time, $D'(t) = D'$, because then the optimal solution is also stationary; i.e., $H(t, \mathbf{x})$ is independent of t, and the rent cost is $c_r H$

Total logistics cost

If instead of H (and as is often done in the literature) we use the delivery lot size $v = D'H$ as a decision variable, keeping ns as the other variable, then the sum of costs can be expressed as:

$$z = \alpha_0 + \alpha_1 \frac{1}{n_s v} + \alpha_2 \frac{1}{v} + \alpha_3 n_s + \alpha_4 v.$$

where the $\alpha_0, \dots, \alpha_4$ are the following interpretable cost constants, which will be used from now on:

$\alpha_0 =$ $(c_s' + c_i r/s + c_i t_s/2)$; handling and fixed pipeline inventory cost per item,

$\alpha_1 =$ $(2rc_d + c_s)$; transportation cost per dispatch,

$\alpha_2 =$ $(c_d k\delta^{-1/2} + c_s)$; transportation cost added by a customer detour,

$\alpha_3 =$ $1/2c_i(k\delta^{-1/2}/s + t_s)$; pipeline inventory cost per item caused by a customer detour and the ensuing stop,

$\alpha_4 =$ c_h/D'; stationary holding cost of holding one item during the time $(1/D')$ between demands.

Total logistics cost (cont.)

$$z = \alpha_0 + \alpha_1 \frac{1}{n_s v} + \alpha_2 \frac{1}{v} + \alpha_3 n_s + \alpha_4 v.$$

- z is a "logistics cost function" (LCF) that relates the cost per item distributed to the decision variables of our problem.
- With the new notation, we have: $n_s v \leqslant v_{\max}$. In addition, we require $n_s \geqslant 1$. Clearly, the LCF is constrained by these inequalities.
- We will see that the determination of a realistic LCF is perhaps the most important step in the design of a logistics system with the CA approach. In the present case, the minimum of the total cost subject to these inequalities is the solution to our problem.

Total logistic cost (cont.)

- Note that α_0 can (and often will) be omitted for optimization purposes. Note as well that, with a small modification to the expressions for α_1, α_2, and α_3, This expression also applies to the VRP case with few tours*; k should be replaced by k' and the term $2rc_d$ should be omitted.
- We will assume for the remainder of this section that the α_1, α_2, and α_3 for large N are used in the optimization; if the resulting n_s found in an application is inconsistent with these values, then the α's should be changed to recognize that N is "small". Our qualitative discussion also applies to this case, which is very similar.

*Recall N is small compared with n_s^2

The full vehicle condition

- We now identify a condition under which the pipeline inventory term ($\alpha_3 n_s$) can be neglected, and show that in that case $n_s v = v_{\max}$.
- For any integer n_s, a feasible solution to the LCF is $v = v_{\max}/n_s$, which (ignoring α_0) yields:

$$z(n_s) = \frac{\alpha_1}{v_{\max}} + \frac{\alpha_2 n_s}{v_{\max}} + \alpha_3 n_s + \alpha_4 \frac{v_{\max}}{n_s}$$

- An upper bound, z^u, to the minimum of the LCF, z^*, is obtained from $z(n_s)$, using $n_s \approx \max\{1, v_{\max}(\alpha_4/\alpha_2)^{1/2}\}$ that is:

$$z^* \leqslant z^u \approx \begin{cases} \frac{\alpha_1}{v_{\max}} + 2(\alpha_2\alpha_4)^{1/2} + \alpha_3 v_{\max}(\alpha_4/\alpha_2)^{1/2*}, & \text{if } v_{\max}(\alpha_4/\alpha_2)^{1/2} \geqslant 1 \\ \frac{\alpha_1}{v_{\max}} + \frac{\alpha_2}{v_{\max}} + \alpha_3 + \alpha_4 v_{\max}, & \text{otherwise} \end{cases}$$

*注意课本 121 页笔误

Lower bound

- A lower bound to the optimal cost is obtained by neglecting the pipeline inventory term $\alpha_3 n_s$ of the LCF, and optimizing the problem. We see at a glance that LCF decreases with n_s for any v; thus, one will always choose the largest n_s satisfying constraints: $n_s \leqslant v_{\max}/v$. (Note that if $v < v_{\max}$, then $n_s \geqslant 1$ holds.)
- If this value is substituted for n_s in the LCF, without its first and fourth terms, we obtain a function

$$z(v) = \frac{\alpha_1}{v_{\max}} + \frac{\alpha_2}{v} + \alpha_4 v$$

whose minimum (subject to $v < v_{\max}$) is a lower bound, z^l. Its expression is:

$$z^* \geqslant z^l \approx \begin{cases} \frac{\alpha_1}{v_{\max}} + 2(\alpha_2\alpha_4)^{1/2}, & \text{if } v_{\max}(\alpha_4/\alpha_2)^{1/2*} \geqslant 1 \\ \frac{\alpha_1}{v_{\max}} + \frac{\alpha_2}{v_{\max}} + \alpha_4 v_{\max}, & \text{otherwise} \end{cases}$$

*注意课本 121 页公式 4.21 错误

Gap between z^u and z^l

- Notice that the expressions for z^u and z^l are almost identical: $z^u - z^l = \alpha_3 v_{\max}(\alpha_4/\alpha_2)^{1/2}$ if $v_{\max}(\alpha_4/\alpha_2)^{1/2} \geqslant 1$, and $z^u - z^l = \alpha_3$, otherwise.
- The relative difference between any two of z^u, z^* and z^l should be lower than $\epsilon = \frac{\alpha_3 v_{\max}}{2\alpha_2}$, the ratio of the maximum value of $(z^u - z^l)$ to $2(\alpha_2\alpha_4)^{1/2}$, which is the second term of z^l when $v_{\max}(\alpha_4/\alpha_2)^{1/2} \geqslant 1$. It bounds z^l from below.
- The numerator of this constant, $\alpha_3 v_{\max}$, is the pipeline inventory cost accruing to a full vehicle for one delivery detour; the denominator is double the vehicle motion cost per detour. For most commodities this ratio is orders of magnitude smaller than 1, so that the lower and upper bounds will nearly coincide.
- In summary, if $\epsilon \ll 1$, then filling the vehicles (as done with the strategy leading to z^u) is near optimal; the resulting cost is close to the lower bound, obtained without pipeline inventory costs.

Problems with large gaps

- The incentive to fill vehicles, used so far, does not apply if $\epsilon = \alpha_3 v_{\max}/(2\alpha_2)$ is large compared with 1. The LCF minimization problem then yields a strict inequality for $n_s v \leqslant v_{\max}$. We now examine the solution to this minimization problem with varying conditions in time-space.
- The unconstrained minimum of LCF can be obtained numerically, and it can also be expressed analytically as a function of one single parameter β. To see this, let n_s be close to the unconstrained minimum of LCF: $n_s \approx (\alpha_1/\alpha_3 v)^{1/2}$; then $z^*(v) = 2(\alpha_1\alpha_3/v)^{1/2} + \alpha_2/v + \alpha_4 v$. This expression reflects an achievable cost if $n_s > 1$.

Problems with large gaps (cont.)

- Because $z^*(v)$ is convex, its minimum is the root of $dz^*(v)/dv = 0$. Using $v' = (\alpha_1\alpha_3 v)^{1/2}/\alpha_2$, we can express this equation in terms of v' as follows:

$$\beta \times (v')^4 = 1 + v';\ \text{i.e.,}\ \beta = (v')^{-4} + (v')^{-3}$$

  where $\beta = \alpha_4\alpha_2^3/(\alpha_1\alpha_3)^2$
- When v' is small compared with 1 the second term in the last expression can be neglected; in this case the solution is: $v' \approx \beta^{-1/4} \ll 1$ for $\beta \gg 1$.
- Conversely, if v' is large compared with 1, i.e., $\beta \ll 1$, the first term can be neglected and the solution becomes $v' \approx \beta^{-1/3}$.

Problems with large gaps (cont.)

- The largest of the two extreme solutions can be used as a rough approximation when $\beta \approx 1$.
- The optimal vehicle load is $n_s v = v' \alpha_2/\alpha_3$, and $n_s = \alpha_1/\alpha_2 v'$. If the vehicle load is smaller than $v_{\max}$ and $n_s > 1$, then the solution can be accepted. (This happens if $\alpha_2 v' < \alpha_1$ and $\alpha_3 v_{\max}$). The optimal H and z can also be expressed as a function of v', and thus of β.

Ignoring the pipeline inventory cost

- Without pipeline inventory, the solution z^* is as the derivation for the lower bounder. $z^* \approx z^l \approx \begin{cases} \frac{\alpha_1}{v_{\max}} + 2(\alpha_2\alpha_4)^{1/2}, & \text{if } v_{\max}(\alpha_4/\alpha_2)^{1/2} \geqslant 1 \\ \frac{\alpha_1}{v_{\max}} + \frac{\alpha_2}{v_{\max}} + \alpha_4 v_{\max}, & \text{otherwise} \end{cases}$ increases linearly with α_1*
- Because there is an intercept, both z^* and the total cost/ unit time $ND'z^*$ increase "less-than-proportionately" with r; the ratio of cost to distance decreases.
- We also see that z^* decreases with the demand rate/customer D'†, but increases with the spatial density of customers δ‡ if their aggregate demand rate ND' (i.e., $\delta D'$) is constant. However, the total cost/unit time $ND'z^*$ is non-decreasing with D'.
- While not so obvious, these scale economies are also shared by the solution to LCF minimization problem as just described. While $ND'z^*$ increases with D', z^* decreases; the optimal cost also increases less than proportionately with distance from the depot.

*which also increases linearly with the distance from the depot r since $\alpha_1 = 2rc_d + c_s$

†$\alpha_4 = c_h/D'$

‡$\alpha_2 = c_d k\delta^{-1/2} + c_s$

Extensions

- To estimate cost for a problem with varying $D'(t), \delta(\boldsymbol{x})$ and $r(\boldsymbol{x})$, one would need to **average the analytical solution over t and $\boldsymbol{x}$**. Although it may be possible to do this in closed form using statistical approximation formulas for expectations (these indicate that cost increases with variable conditions), a few numerical calculations should suffice.
- One could calculate z^* for all the $D(t_{\max})N$ items demanded, using their respective t and $\boldsymbol{x}$, but this would be too laborious. Instead, one can partition the time axis into $m = 1, \ldots, M$ intervals and $\boldsymbol{R}$ into $p = 1, \ldots P$ subregions so that each (m, p) combination includes roughly the same amount of demand. We use any interior point $(t, \boldsymbol{x})$ of each combination to calculate both the parameters of the optimization and the resulting cost, z^{mp}. The estimated cost is then the arithmetic average of the z^{mp}.

目 录 |Outline

- We now describe how specific solutions can be designed from the optimization results in prior sections. It also discusses systematic ways for fine-tuning the designs.
- We already know that changes in the input parameters of an EOQ optimization have a dampened effect on the decision variables; this is also true for the objective function now at hand.
- Thus, if $D(t)$ and $\delta(\boldsymbol{x})$ change slowly, the decision variables H (or v) and n_s will change even more sluggishly over t and $\boldsymbol{R}$. Because, as with the EOQ optimization, the decision variables themselves do not need to be set very precisely, it should be possible to identify large regions of the time-space domain where the decision variables can be set constant without a serious penalty.

- For our problem with identical customers, the partition is easily developed: (i) divide the time axis into $m = 1, 2, \dots, M$ periods with nearly constant demand rates; and (ii) partition $\boldsymbol{R}$ into $p = 1, 2, \dots P$ subregions with similar customer density and distance to the depot.
- The subregions and time periods should be large enough to include respectively several delivery districts and several headways. This ensures that the number of stops in each district can be close to ideal, and that the theoretical headway $H(t, \boldsymbol{x})$ can be approximated with an integer number of dispatches.
- We anticipate now that, by designing a different spatial partition for every time period, this method can be extended to situations with different customers and time varying customer densities.

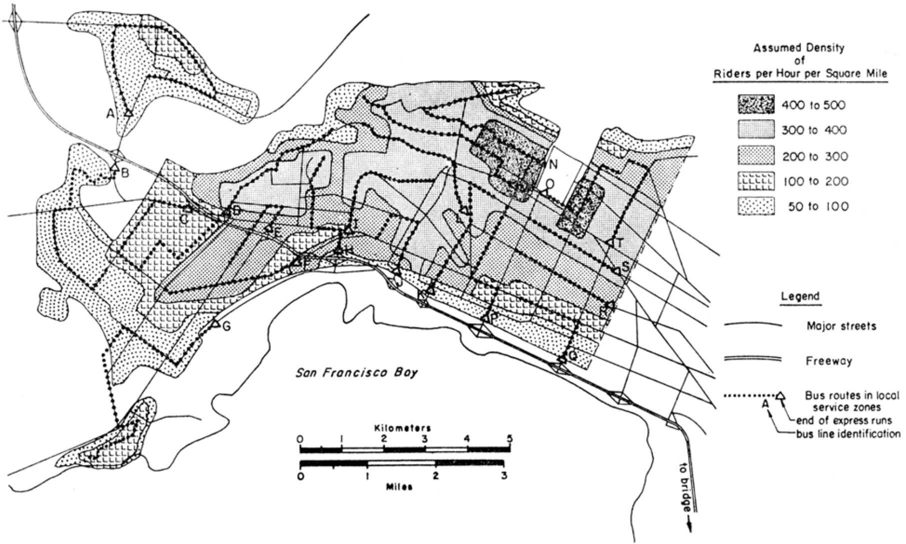
- An application of the technique for a very similar problem has been reported by Clarens and Hurdle (1975).
- These authors explored the best way of laying out transit routes from a CBD to its outlying suburbs. They assumed that the demand was stationary and changed with position.

- They describe the solution in terms of slightly different variables and inputs, but the differences are only superficial. They define the vehicle operating cost as a function of time (and not distance), c_t, and do not explicitly account for the number of stops; instead they assume that one knows from empirical observations the time that it takes for a bus to cover one unit area — a constant, $\tau(\boldsymbol{x})$, that can vary with position.
- They define the demand as a density per unit area and unit time, $\lambda(\boldsymbol{x})$, which changes with position. Instead of a distance from the CBD, $r(\boldsymbol{x})$, they define an express (line-haul) travel time, $T(\boldsymbol{x})$, and as a decision variable they use the area of a bus service zone, $A(\boldsymbol{x})$, instead of $n_s(\boldsymbol{x})$. Thus, they work with the following logistic cost function, which is equivalent to LCF:

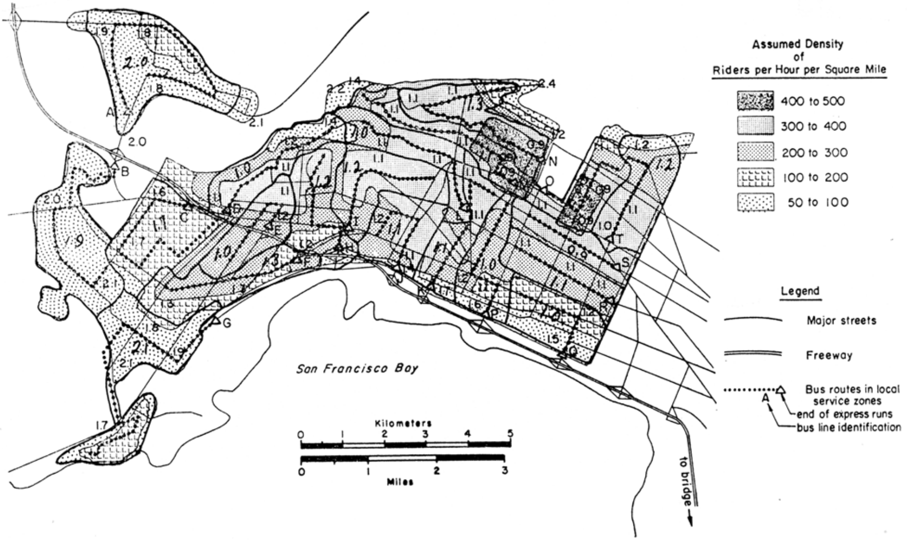
$$z = \frac{2c_t T}{A\lambda H} + \frac{\tau c_t}{\lambda H} + c_i\{T + \tau A/2\} + c_h H/2$$

where the bus load, $A\lambda H$ is restricted to be below $v_{\max} = 45$ passengers. Note that the constraints are also similar.

Demand distribution for a transit line design problem



Worksheet for a transit design problem



Comparison of the actual and ideal zone size

Results of the transit line design process
(Source: Clarens and Hurdle, 1975)

Zone	Area (square miles)		Headway (minutes)	Average Load On Bus (persons)	Load Factor	$T(x,y)$ (minutes)	$r(x,y)$ (min./sq mi)
	Actual	$A^*(x,y)$					
A	2.0	1.9	13	35	78	27	9
B	1.9	1.9	14	31	69	27	10
C	1.7	1.3	10	43	96	25	10
D	1.0	1.1	9	39	87	26	11
E	1.0	1.2	11	40	89	24	9
F	1.3	1.4	14	36	80	26	10
G	2.1	1.9	8	27	60	21	9
H	1.2	1.5	7.3	38	84	22	8
I [a]	1.2	1.2	7	45	Full	26	8
J	1.1	1.2	7	36	80	20	8
K	1.1	1.1	9.0	38	84	19	10
L	1.0	1.0	6.7	45	Full	24	13
M [a]	1.1	0.9	6.7	45	Full	26	13
N [a]	1.3	0.8	5.8	45	Full	29	16
O [a]	0.9	1.0	8	43	96	25	11
P	1.0	1.0	9	30	67	17	11
Q	1.0	1.3	10	23	51	15	8
R	1.1	1.3	8	35	38	20	8
S	0.9	1.0	7	40	89	21	10
T	1.2	1.5	7	39	87	22	7

[a] Zones where $A^*(x,y) = A_e(x,y)$.

- Given the close agreement between these two columns of figures and the robustness of the CA solution to small departures from the recommended settings, one would expect to have a cost that is very close to the minimum.
- The Clarens-Hurdle case study was an published example where the CA guidelines have been translated into a proposed design for a two-dimensional problem.
- On reviewing the procedure, it becomes clear that a great deal of human intuition is required to complete a design. Furthermore, careful efforts notwithstanding, the designer may miss opportunities for small improvements at the margin that depend on specific details (e.g., stop locations, street intersections, etc.) of the particular problem. It might be worthwhile to use fine-tuning software to find these possible improvements if any exist.

目录 |Outline

- The rest of this section describes the results of some experiments where fine-tuning software was used to improve detailed VRP solutions developed quickly from the guidelines of TSP/VRP.
- These authors tested simulated annealing (SA) as a technique that is well suited for fine-tuning purposes. The brief discussion of simulated annealing provided in this reference is included as Appendix B. The technique is attractive because:
    - A prototype computer program can be developed quickly for most problems since the SA logic is very simple. (These authors developed software for the VRP, from scratch, in about three mandays.)
    - The optimization can be controlled by means of input variables (called initial "temperature" and "cooling rate" or "annealing schedule") which determine how much the algorithm is allowed to increase (worsen) the objective function at different stages of the process in the hope of finding larger reductions later.

- Simulated annealing is known to converge in probability to the global optimum of combinatorial optimization problems, such as those arising when designing in detail logistics systems.
- Unfortunately, convergence is slow. To be guaranteed, the initial temperature has to be very large and the cooling rate very slow; the computer time required rapidly becomes prohibitively long with increasing problem size. However, with an overall idea of the system's structure, and a near optimal initial solution as would be obtained with nondetailed methods, the scope of the annealing search can be restricted. As demonstrated in Robusté et al. (1990), a low initial temperature achieves that.
- It prevents the search from wandering away from the initial solution, while systematically testing variations that exploit the details (specific locations of customers, for example.)

- One of the examples in this reference considers a VRP problem with $N = 500$ points (randomly located according to a uniform density in a 6-inch by 10-inch rectangle), $C = 45$ stops per tour and a centrally located depot; distances are Euclidean.
- For this test the VRP formula with $k \approx 0.57$, predicts a total distance averaging 179 inches. With a high initial temperature, the SA approach yielded tours that were very long in reasonable times; after one day of computation it obtained a set of tours 180.4 inches long. This was reasonable, but longer than the hand constructed tours using the VRP guidelines presented earlier.
- When the hand constructed tours were used to initiate SA with a low initial temperature, the SA algorithm found enough modifications to reduce the total length by about four percent — to 173.6 inches.

SA solution

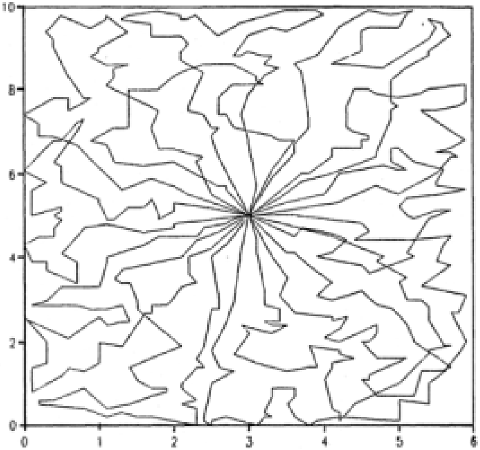


Figure: 500 point VRP. C= 45. 12 tours with total length = 180.4inches.

Manual solution

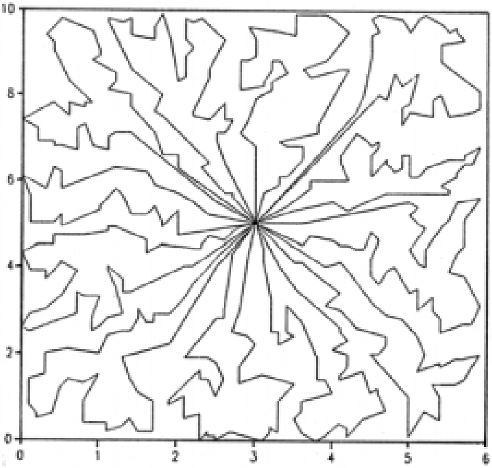


Figure: 500 Point VRP. C=45. 12 tours with total length = 179.8 inches.

Other tests performed in this reference show that the non-detailed approach, fine-tuned with SA, can obtain solutions with objective functions as low as those currently believed to be optimal. The efficiency of the twostep approach has also been demonstrated in practice — the (non-detailed) results in Burns and Daganzo (1987) were used in conjunction with SA to schedule the assembly lines in some GM plants

- These observations are in agreement with our philosophical conclusions. Like the evolution processes in nature, to design a complex logistic system it seems best to develop a preliminary design based on the overall characteristics of the problem, and use the details later to fine-tune the preliminary design. This view has been adopted in the recent works of Langevin and StMleux (1992) and Hall et. al. (1994).
- Although the CA approach and the SA algorithm seem to be ideal companions for this twostep approach, other methods may also be useful. The critical thing is not the specific approach for each step, but the fact that the first step disregards details in searching over all possible solutions, and the second step—restricted to a small subset of possible solutions—incorporates all the details.
- Perhaps other computer fine-tuning methods will improve on SA (Neural Networks and Tabu Searches...etc.). But the improvement should not be measured only on computation grounds; the ability to develop the software quickly is just as important.

目 录 |Outline

Overview

- Symmetric strategies are extensions from strategies for identical customers
- Asymmetric strategies allow different customer types to be served differently.
- Conditions under which these more complex strategies are likely to be of benefit are also discussed here

Symmetric Strategies

- Let us allow $D_n(t)$ to vary across customers n, and possibly to be non-stationary. With this generalization, even if the demand is stationary, D'_n can vary across n.
- With many customers the individual demand rates should be treated as "details", which we try to avoid. To this end, an expected **demand density rate per unit area** $\lambda(t, \boldsymbol{x})$ is used instead of the specific $D_n(t)$'s.
- $\lambda(t, \boldsymbol{x})$ is assumed to vary slowly with time and location so that the demand in a subregion, $\boldsymbol{P}_p$* of $\boldsymbol{R}$ during a time interval $[t_{m-1}, t_m)$ is:

$$\int_{t_{m-1}}^{t_m} \int_{x \in \boldsymbol{P}_p} \lambda(t, \boldsymbol{x}) \mathrm{d}\boldsymbol{x}\boldsymbol{d}t.$$

* $\boldsymbol{P}_p$ is large enough to contain several destinations but of small dimensions relative to $\boldsymbol{R}$

顾客密度与需求不确定性

- Similarly, we define a **customer density**, $\delta(t, \boldsymbol{x})$, which is also allowed to vary with time. Note that we are allowing here for the number and locations of customers to change with time; all we require is that these changes can be approximated with functions $\delta(t, \boldsymbol{x})$ and $\lambda(t, \boldsymbol{x})$, that vary smoothly with t and $\boldsymbol{x}$.
- **Demand uncertainty** is an important phenomenon when **the tours have to be planned before the demand is known at the destinations**. It will be captured by an index of **dispersion function**, as described below.

需求分散度的衡量

- Take a partition $\{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_p, \ldots, \boldsymbol{P}_P\}$ of $\boldsymbol{R}$ and a partition of time into consecutive intervals $\tau_m = [t_{m-1}, t_m)$, and let D_{mp} represent the actual number of items demanded in $\boldsymbol{P}_p$ during τ_m.
- The parameter $\lambda(t, \boldsymbol{x})$ can then be defined as the average demand rate density, so that $\int_{t_{m-1}}^{t_m} \int_{\boldsymbol{x} \in \boldsymbol{P}_p} \lambda(t, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \mathrm{d}t$ now gives the mean of D_{mp}.
- We assume that, for any partition, the variables D_{mp} are independent, and identically distributed. Then their variance can be expressed as:

$$\mathrm{var}\{D_{mp}\} \cong \int_{\tau_m} \int_{\boldsymbol{P}_P} \lambda(t, \boldsymbol{x}) \gamma(t, \boldsymbol{x}) \mathrm{d}\boldsymbol{x}\boldsymbol{d}t$$

where $\gamma(t, \boldsymbol{x})$ is an "index of dispersion", with "items" as its physical dimension. $\gamma(t, \boldsymbol{x})$ 是衡量分散度的指标，其物理量纲为数量

需求分散度的衡量 (cont.)

$$\text{var}\{D_{mp}\} \cong \int_{\tau_m} \int_{\boldsymbol{P}_P} \lambda(t, \boldsymbol{x})\gamma(t, \boldsymbol{x})\text{d}\boldsymbol{x}\boldsymbol{d}t$$

- A special case of this model arises if each customer's demand fluctuates independently of other customers, either like a stochastic process with independent increments — such as a compound Poisson process or a Brownian motion process.
- Although in most cases a fixed γ should capture demand fluctuations well, we allow $\gamma(t, \boldsymbol{x})$ to vary slowly with t and $\boldsymbol{x}$. An index equal to zero represents known demand; no uncertainty. This case will be examined next.

物理系统成本构成

Recall the LCF:

$$z = \alpha_0 + \alpha_1 \frac{1}{n_s v} + \alpha_2 \frac{1}{v} + \alpha_3 n_s + \alpha_4 v.$$

where the $\alpha_0, \ldots, \alpha_4$ are the following interpretable cost constants:

$\alpha_0 =$ $(c_s' + c_i r/s + c_i t_s/2)$; handling and fixed pipeline inventory cost per item,

$\alpha_1 =$ $(2rc_d + c_s)$; transportation cost per dispatch,

$\alpha_2 =$ $(c_d k \delta^{-1/2} + c_s)$; transportation cost added by a customer detour,

$\alpha_3 =$ $1/2 c_i (k\delta^{-1/2}/s + t_s)$; pipeline inventory cost per item caused by a customer detour and the ensuing stop,

$\alpha_4 =$ c_h/D'; stationary holding cost of holding one item during the time $(1/D')$ between demands.

The constraints are $n_s v \leqslant v_{\max}$ and $n_s \geqslant 1$.

决策变量与目标函数

- For consistency with the literature, we continue to use $H(t, \boldsymbol{x})$ and $A(t, \boldsymbol{x})$ as the decision variables instead of n_s and v. Both formulations are equivalent, since there is a 1:1 correspondence between two sets of variables—the number of stops in a tour is the number of customers in its district, which is given by $n_s \approx \delta(t, \boldsymbol{x})A(t, \boldsymbol{x})$, and the delivery lot size is the consumption during a headway in the area around a customer: $v \approx \lambda(t, \boldsymbol{x})H(t, \boldsymbol{x})/\delta(t, \boldsymbol{x})$.
- Making these substitutions in the LCF and the constraints, and recognizing that $D' = \lambda/\delta$, the cost per item at $(t, \boldsymbol{x})$ can be expressed as:

$$z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0$$

where α_1, α_2 and α_3 are the constants defined in connection with LCF optimization problem, which now can vary in both time and space; the constraints become: $\lambda AH \leqslant v_{\max}, \delta A \geqslant 1,$ and $\lambda H \leqslant v^{\circ}\delta$.

- The important thing to remember here is not new expression for LCF, but the process followed to derive them and use them. This process is quite general and can be used for problems involving various peculiarities.
- Because it is impossible here to discuss all possible situations, the process is only illustrated with three examples involving stochastic phenomena and requiring some modifications to the equations.
- The first example arises where **items are indivisible and the expected demand per customer per headway is less than one item**;
- The second when **the customer demands are not known until the vehicles make the stop**
- The third when **the vehicles make coordinated adjustments to their routes as demand information becomes known**.

目录 |Outline

Random Demand: Low Customer Demand

$$\min \quad z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0$$
$$\text{s.t.} \ \lambda AH \leqslant v_{\max}, \delta A \geqslant 1, \text{ and } \lambda H \leqslant v^{\circ}\delta$$

- It implicitly assume that each customer is visited each time – the number of stops is equal to δA. But if items are indivisible (as opposed to fluids, or very small items) and the demand by individual customers is so low that **some have no demand during a headway, their stops can be skipped**. 低需求率的特点是每次配送不必访问所有顾客。
- For some demand processes, the proportion of stops that can be skipped should decrease with H as $exp^{(-H/H_0)}$, where H_0 is a constant that depends on t and $\boldsymbol{x}$. If the customers in a subregion are alike and their demand is well described by Poisson processes*, then the parameter H_0 is the average time between successive demands at one destination; i.e., $H_0 = D'^{-1} = \delta/\lambda$. For other processes the relationship is similar.

*当需求的变动遵从泊松过程时，需求之间的时间间隔服从指数分布

Random Demand: Low Customer Demand (cont.)

- The effective density of stops is only $\delta[1 - exp^{-H/H_0}]$.
- This expression must be substituted for the parameter δ in the expressions for $(\delta\alpha_2)$ and $(\delta\alpha_3)$ (remember that δ also appears in α_2 and α_3). The optimization and design process can be carried out as described earlier. Although the resulting optimum is slightly more complicated, two extreme cases are quite simple.
- First, if $H \gg H_0$ then the density of stops is δ as before; the solution does not have to be changed. The opposite extreme case with $H \ll H_0$, arising for example if $\delta \to \infty$ but $D' \to 0$, also admits a simple expression for the stop density, even if the demand varies across customers. The expression is $\delta H/H_0 \approx \lambda H$* if items are not demanded in batches; then the number of vehicle stops per tour, $(\lambda H)A$, equals the vehicle load λHA as one might expect.

*每次配送每个顾客被停靠的概率，同样也是每两个配送间隔每个顾客产生的需求量

Random Demand: Uncertain Customer Requests

$$\min \quad z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0$$
$$\text{s.t. } \lambda AH \leqslant v_{\max}, \delta A \geqslant 1, \text{ and } \lambda H \leqslant v^{\circ}\delta$$

- If α_3 is small* we have seen that the minimum logistics cost will be such that $\lambda AH = v_{\max}$. There is an incentive to dispatch totally full vehicles.
- Let us now see what modifications are needed if the exact demand on a vehicle route is not known accurately when the vehicles are dispatched†.
- The system of interest operates with a **headway** (e.g., daily, weekly, etc.) to be determined, and advertised to customers as a service schedule that is to be met even if the volumes to be carried change with every headway. This scenario can arise for both collection and distribution problems, although for distribution problems of destination-specific items the demand will normally be known.

*It implies that items are cheap

†A case with expensive items is not considered here because if time is of the essence, it is unlikely that one would operate with imperfect information

Random Demand: Uncertain Customer Requests (cont.)

$$\min \quad z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0$$
$$\text{s.t. } \lambda AH \leqslant v_{\max}, \delta A \geqslant 1, \text{ and } \lambda H \leqslant v^{\circ}\delta$$

- If the size of each delivery v_n is both known and small compared with $v_{\max}$ it should not be difficult to partition the service region into delivery districts of nearly ideal shape with $\sum_n v_n \approx v_{\max}$. Then, the distance formulae hold and the LCF can be used without modification.
- If some delivery lots are comparable to the vehicle's capacity, the routing problem is more difficult because one needs to balance the incentive for filling a vehicle by delivering a right lot size to an out-of-the-way customer with the extra distance that one would have to travel.

- In view of the above, our discussion is phrased in terms of collection, although hypothetical distribution problems with uncertain demand would be mathematically analogous.
- For collection problems some of the vehicles may be filled before completing their routes, which would cause some of the demands to go unfulfilled.

Random Demand: Uncertain Customer Requests (cont.)

- The overflow customers (still needing visits) could be covered in the same headway by collection vehicles with unused cargo space or, failing that, by vehicles dispatched from the depot.
- Clearly, if some vehicles can be rerouted before returning to the depot, some distance can be saved. Dynamic routing introduces modeling complexities that will be discussed later. For now we assume that **all the overflow customers are visited by a separate set of secondary vehicle routes based at the depot and planned with full information**.
- This information is gathered by the original (primary) vehicles, which are assumed to visit all the customers. Because items are "cheap", secondary vehicles should also travel full.

Random Demand: Uncertain Customer Requests (cont.)

- The decision variables are A and H, as before, but now the capacity constraint must be replaced by an **overflow cost** which depends on A and H. A new trade-off becomes clear.
- If the average demand for a tour satisfies $\lambda AH \ll v_{\max}$, then the overflow cost will be negligible, but most primary vehicles will travel nearly empty.
- On the other hand, if $\lambda AH \approx v_{\max}$, a larger number of customers will overflow on average—the actual number will depend on the variability of demand as captured by its index of dispersion, γ.

Random Demand: Uncertain Customer Requests (cont.)

$$\begin{aligned} \min \quad & z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0 \\ \text{s.t.} \quad & \lambda AH \leqslant v_{\max}, \delta A \geqslant 1, \text{ and } \lambda H \leqslant v^{\circ}\delta \end{aligned}$$

- Instead of a total cost per item, we work with a **cost per unit time and per unit area**. For given A and H, the transportation cost per unit time and unit area for primary tours is approximately independent of the overflow; it is well approximated by the product of the constant factor λ, and the first two terms of z:

$$\frac{\alpha_1}{AH} + \frac{\delta\alpha_2}{H}$$

  Strictly speaking, this expression is an upper bound because it **ignores the local delivery distance** that it is saved by the stops that are skipped.

Random Demand: Uncertain Customer Requests (cont.)

- Note that, especially when the fraction of tours overflowing is small, the **overflow customers will tend to be geographically distributed in widely spaced clusters of customers corresponding to overflowing tours**. Because the overflow transportation cost formulas with clustered destinations are more complicated, two simple bounds will be used instead to approximate the secondary distance traveled*. 超量顾客的分布更稀疏
- It should be intuitive without a formal derivation that **smearing the clusters uniformly over *R*** increases the distance traveled, while **collapsing them into a single point** decreases it. Upper and lower bounds for secondary distance are derived below, imagining that clusters are either spread or fused in this manner.

*Blumenfeld and Beckmann, 1984, have developed formulas for VRP's with clustered demand points

辅助运输发生的次数

- An expression for f_0, the fraction of items that must be delivered or collected as overflow, will be derived shortly.
- Assume for now that f_0 is given. Then the number of secondary (overflow) tours per unit area is $\lambda H f_0 / v_{max}$*, and the number of stops is close to $f_0 \delta$.
- This expression implies that **the fraction of items overflowing is the same as the fraction of customers**; the expression is exact if primary vehicles don't deliver (or collect) partial lots, and is also a good approximation in other cases.

*λH 是区域内的需求量，除以 v_{max} 表示主要配送所服务的顾客数，再乘以 f_0 表示辅助配送所服务的顾客数

Random Demand: Uncertain Customer Requests (cont.)

超量顾客分散/集中分布时，辅助运输的距离

- With **de-clustered overflowing customers**, the upper bound to the secondary distance per unit area is thus:

$$\frac{2r\lambda H f_0}{v_{max}} + k(f_0\delta)^{1/2}.$$

  We are assuming here that the total number of customers is greater than the squared number of stops per vehicle: $Nf_0 \gg (v_{max}\delta/\lambda H)^2$

- With **perfectly clustered groups** the density of stops equals the density of incomplete primary **tours**. If we let g_0 denote the probability that a tour overflows, then this density is g_0/A; thus a lower bound for the distance per unit area is:

$$\frac{2r\lambda H f_0}{v_{max}} + k(g_0/A)^{1/2}.$$

Random Demand: Uncertain Customer Requests (cont.)

- The secondary transportation cost per unit area and unit time is obtained by multiplying either distance bound by c_d/H, and adding to the result the cost of stopping. For the upper bound we have:

$$\begin{aligned}\text{overflow transport cost} \approx & c_d\left[\frac{2r\lambda f_0}{v_{\max}} + \frac{k(f_0\delta)^{1/2}}{H}\right] + c_s\left(\frac{\lambda f_0}{v_{\max}} + \frac{f_0\delta}{H}\right) \\ = & \alpha_1(\frac{\lambda f_0}{v_{\max}}) + \frac{k(f_0\delta)^{1/2}c_d}{H} + \frac{f_0\delta c_s}{H}\end{aligned}$$

- For the lower bound, the factor $(f_0\delta)^{1/2}$ of the second term should be replaced by $(g_0/A)^{1/2}$. If the overflow is so small that only a few secondary tours are used, $Nf_0 < [v_{\max}\delta/\lambda H]^2$, then k should be replaced by k' and r should be set to 0, regardless of position.

Random Demand: Uncertain Customer Requests (cont.)

- Either on primary or secondary tours, items reach the destination at regular intervals, as required, approximately H time units apart. Thus, the stationary holding cost per unit time and unit area is:

$$\text{holding cost} \approx c_h(\lambda H)$$

- We are now ready to write the logistic cost function for our problem. In practical situations one would expect the difference between the upper and lower bound to be small. Therefore, we will use one of these bounds (the upper bound) below.

Random Demand: Uncertain Customer Requests (cont.)

- In terms of total cost per unit time and unit area (the sum of primary and secondary transportation costs, plus the holding cost), the upper bound is

$$\lambda z \cong \frac{(\alpha_1)}{AH} + \frac{(\delta\alpha_2)}{H} + (\alpha_1 \frac{\lambda}{v_{\max}}) f_0 + (k\delta^{1/2} c_d) \frac{f_0^{1/2}}{H} + (\delta c_s) \frac{f_0}{H} + (\lambda c_h) H,$$

  where the parenthetical items are constants and the rest $(A, H,$ and $f_0)$ are decision variables. Note that the constant handling cost, α_0, has been omitted from the LCF.

Random Demand: Uncertain Customer Requests (cont.)

- The fraction of items that overflow is related to A and H.
- Recall the following equations

$$\text{mean}\{D_{mp}\} \cong \int_{t=t_{m-1}}^{t_m} \int_{x \in \boldsymbol{P}_p} \lambda(t, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \mathrm{d}t.$$

$$\text{var}\{D_{mp}\} \cong \int_{\tau_m} \int_{\boldsymbol{P}_P} \lambda(t, \boldsymbol{x}) \gamma(t, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \mathrm{d}t$$

The mean and variance of the number of items to be carried by a primary vehicle are λAH and $\lambda A \gamma H$. The expectation of the excess of this random variable over $v_{\max}$ is the average overflow for the vehicle.

Random Demand: Uncertain Customer Requests (cont.)

Assuming that the demand is approximately normally distributed, and letting Φ denote the standard normal CDF (and Φ' its derivative —the PDF), we can therefore write:

$$
\begin{aligned}
f_0 \cong & \frac{1}{\lambda AH} \int_{v_{\max}}^{\infty} (x - v_{\max}) d\Phi \left(\frac{x - \lambda AH}{(\lambda A \gamma H)^{1/2}} \right) \\
= & (\lambda AH/\gamma)^{-1/2} \Psi \left[\frac{(\lambda AH - v_{\max})}{(\lambda A \gamma H)^{1/2}} \right]
\end{aligned}
$$

where

$$
\Psi(z) = \int_{-\infty}^{z} \Phi(w) dw = \Phi'(z) + z\Phi(z)
$$

which is a convex function increasing from zero (when $z \to -\infty$) to ∞ (when $z \to \infty$). Note that f_0 may depend on position and time.
本式计算的实际上是：假设配送量服从正态分布时，其取值大于等于 $v_{\max}$ 的概率。该分布的参数为 λ, A, H, γ。

Random Demand: Uncertain Customer Requests (cont.)

$$\lambda z \cong \frac{(\alpha_1)}{AH} + \frac{(\delta\alpha_2)}{H} + (\alpha_1 \frac{\lambda}{v_{\max}}) f_0 + (k\delta^{1/2} c_d) \frac{f_0^{1/2}}{H} + (\delta c_s) \frac{f_0}{H} + (\lambda c_h) H,$$

- Thus, λz should be minimized, subject to the expression of f_0.
- The procedure is simple. Conditional on AH, i.e. on the average vehicle load per district, f_0 is fixed and λz only depends on H; the optimal headway can be obtained in closed form from the expression of λz as an EOQ trade-off involving the 2nd, 4th, 5th and 6th terms of that expression. The resulting cost is only a function of AH, which can be minimized numerically.
- The procedure also works for the lower bound, and when the number of secondary tours is low. For the lower bound one should replace the fourth term of λz by $kc_d(g_0/A)^{1/2}/H$, where $g_0 = \Phi(z)$. Note that g_0 is fixed if AH is fixed, like f_0.

- Cost estimates and guidelines for the construction of a detailed strategy can be obtained as usual, by repeating the minimization for a few combinations of $(t, \boldsymbol{x})$.
- We could also verify that the final strategy and the resulting cost do not change much if the overflow local distance term is replaced by the lower bound.

Dynamic Response to Uncertainty

很多情况下，车辆的路径可以动态调整

- In many applications, vehicle routes can be adjusted dynamically during the course of operation. For example if a collection truck of an express package carrier falls behind schedule, central dispatch can reassign some of its remaining customers to currently underutilized trucks. 当某辆用于收集快件的卡车滞后于时刻表时，配送部门可将其剩余顾客分配给目前未被占用的卡车
- If a firm can do this systematically with an efficient control strategy, it should be able to operate with fewer vehicles.

决策尺度

- To design such a system we must make a single set of **planning*** **decisions** at the beginning of the planning period, e.g., choosing **# trucks**; and then a stream of **control decisions** that change dynamically as information is revealed over time.

*or configuration

规划阶段的策略要求

To minimize the combination of fixed and operating costs, configuration decisions must anticipate and accommodate the long-run needs of the control strategy; that is, the system should be planned for control. This is difficult to do exactly but can be achieved approximately if we can find a family of control strategies that is:

1. parametrizable (describable in terms of **just a few parameters**);
2. appealing (containing a **near-optimal strategy** for the configuration of **every reasonable system**);
3. simple (with a **predictable expected cost**).

Properties (1) and (3) guarantee we can write an LCF that captures approximately all fixed and recurring costs in terms of the configuration variables and control parameters. Property (2) guarantees that good control parameters exist for every reasonable configuration.

规划阶段的策略要求 (cont.)

- Hence, the minimum of the LCF is an "appealing" plan. Since an analytic expression exists the minimum can be searched effectively with conventional optimization methods, even if the number of variables and parameters is considerable.
- The selection of a proper family is more an art than a science. The temptation is always to look for the most efficient control strategies, excelling at (2), even if they fail the simplicity test (3). The problem with this approach is that a search for the optimum configuration cannot then easily incorporate the effects of control. The result can be gross sub-optimization.
- For planning purposes we prefer to look for **idealized*** **control strategies** that can be systematically analyzed. This allows us to explore a much larger solution space when configuring the system. The idealized strategies play the role of approximations to the more refined strategies during the optimization process, but the refined strategies can still be used when the system is operated.

*less efficient

配送区域的划分与决策变量

- Let us consider again the load-constrained system, but assume now that $H = 1$ day as in package collection systems. We want to configure a system where vehicles that are partially filled at the end of their runs can cover the overflow customers of other vehicles. Although very complex dynamic routing strategies can be designed to achieve this goal, we shall be satisfied with a simple one that is obviously sub-optimal but improves significantly on the static approach
- We partition the service region into an **inner region close to the depot** (region 2) and an **outer fringe** (region 1). Only customers in region 1 are allocated to primary tours. We use only one planning variable: **# of primary service zones in region 1**, which equals the number of vehicles m. The **radius of the inner region**, r_T, is our **control parameter**.

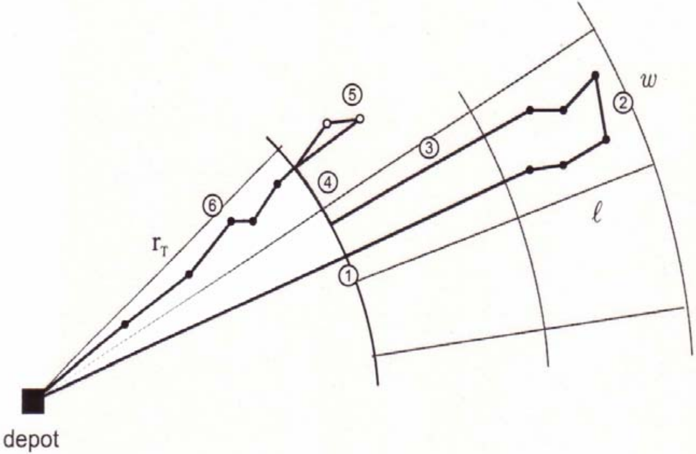


Figure: The idealized control strategy has two phases with several steps

规划阶段的服务过程

- In phase one vehicles travel to their service zones (step 1), serve their customers (step 2), and either return to the depot, if filled, or else stop at the boundary between regions 1 and 2 (step 3). Unfilled vehicles wait there for the start of the second phase, until all vehicles are done.
- Then, they are repositioned along the boundary in anticipation of serving carefully designed groups of remaining customers (step 4). The size of these groups is chosen to be consistent with each vehicle's available capacity. Vehicles first serve the part of their group in region 1(step 5), then the part in region 2 (step 6). Region 2 customers are arranged in wedges that can be served efficiently as vehicles return to the depot. Finally, if any customers remain unserved, they are served with a set of secondary tours (step 7). Note that virtually no customers require such secondary tours when systems are configured optimally.

规划阶段的服务过程 (cont.)

- This strategy generalizes the static procedure, since the effects of the latter can be essentially achieved by setting $r_T = 0$. Although the new strategy is sub-optimal, it has clear efficiencies over the static procedure; thus, it is "appealing" in the sense of (ii). The strategy also has properties (i) and (iii), since it* is parameterized by the inner radius r_T and is simple.
- An analytic approximation for the LCF is given in Erera (2000). The approximations in this reference were designed to be most accurate for intermediate values of r_T, where the optimum was expected to be. The formulae are not given here because they would take too long to explain, but the qualitative results are interesting.

*注意课本 142 页最后一段倒数第五行 t 应为 it

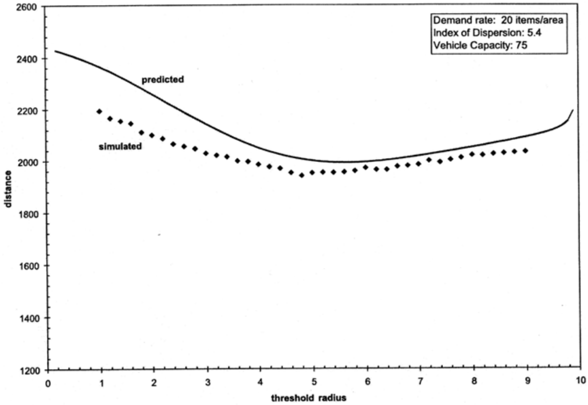


Figure: This figure shows how the approximate total distance per day varies as a function of r_T for a test problem, after the number of vehicles m was optimized. The figure also includes a dotted line from a simulation that used the recommended values of m and r_T, and a more sophisticated control algorithm. This curve gives the actual distance that could be expected in an implementation.

- Reassuringly, the value of r_T recommended by the optimization (the minimum of the solid line) yields a near-minimum actual distance. Note from the figure that this distance is considerably smaller than that achieved with the static strategy ($r_T = 0$).
- Erera (2000) shows with a battery of 20 problems that the reduction in the required number of vehicles is even greater.
- The portion of the vehicle fleet required by uncertainty (the "fleet penalty" in Erera's lingo) was reduced by 50% or more in 19 out of 20 cases and by more than 70% in half of the cases. The median reduction in the "distance penalty" due to uncertainty, on the other hand was only about 30%.

目 录 |Outline

The scenario

- We now explore the advantages of offering different service levels to customers with **different consumption rates** and/or **different holding costs**.
- Because these differences are likely to be most notable for collection problems, our discussion will be phrased in these terms — **factories and manufacturing plants typically consume a wide selection of parts and raw materials even if their product line is homogeneous**.
- Before explaining how asymmetric collection strategies can be designed, we introduce why they are desirable with a very simple example with two customer types.

The LCF

$$\min \quad z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + \delta\alpha_3 A + c_h H + \alpha_0$$
$$\text{s.t. } \lambda AH \leqslant v_{\max}, \delta A \geqslant 1, \text{ and } \lambda H \leqslant v^\circ \delta$$

- Consider a problem with stationary conditions (i.e. λ and δ independent of time) obeying the LCF for which it is desirable to fill the vehicles. More specifically, we assume that: (i) the third (**pipeline inventory**) term of LCF **can be neglected** because items are "cheap", and (ii) that only constraint $\lambda AH \leqslant v_{\max}$ plays a role because **storage room at the origins is plentiful** and the customer density is so large that **the ideal # of vehicle stops is sure to exceed 1**. We also assume that the stop cost c_s can be neglected.

The LCF (cont.)

$$\min z = \frac{\alpha_1}{A\lambda H} + \frac{\delta\alpha_2}{\lambda H} + c_h H + \alpha_0 \text{ s.t. } \lambda AH \leqslant v_{\max}$$

- Let us now examine how the optimal system cost depends on λ and δ. Because z decreases with A for any H, its minimum is reached for as large a district area A as possible.
- Therefore, as expected, the vehicle capacity constraint must hold strictly: $A = v_{\max}/(\lambda H)$. On making this substitution and minimizing the resulting EOQ expression with respect to H, a simple formula for the cost per item z^*, is obtained.

- If α_4 $(c_h H = \alpha_4 v)^*$ is replaced by its expression in terms of δ and λ (i.e., $\alpha_4 = c_h \delta/\lambda$), and the result is expressed in cost units per unit time and unit area, the formula becomes:

$$\lambda z^* = \beta_1 \lambda + (\beta_2 \lambda)^{1/2} \delta^{1/4}.$$

where $\beta_1 = \alpha_0 + \alpha_1 / v_{\max}$ and $\beta_2 = 4 c_h c_d k$. Notice that λz^* increases at a decreasing rate with λ, δ and β_2; this concavity encourages discrimination

*Recall $\alpha_4 = c_h/D' = c_h H/v$.
*注意课本 144 页公式 4.27a 笔误

Two different types of customers

- Suppose that there are two customer types, $n = 1, 2$, with demand characteristics (λ_n, δ_n) and with different c_h, so that β_2 is different for the two customer types: $\beta_2^{(1)}$ and $\beta_2^{(2)}$. (We use n to index customer classes, instead of customers.)
- Note then that $\lambda = \lambda_1 + \lambda_2$ and $\delta = \delta_1 + \delta_2$.

Separate delivery

If the two customer classes are treated completely separately, as if the other did not exist, the combined cost per unit time and unit area, instead of being given by $\lambda z^* = \beta_1 \lambda + (\beta_2 \lambda^{1/2}) \delta^{1/4}$, would be:

$$\lambda z^* = \sum_{n=1}^{2} \left[\lambda_n \beta_1 + \left(\beta_2^{(n)} \lambda_n \right)^{1/2} \delta_n^{1/4} \right]$$

$$= \lambda \beta_1 + \sum_{n=1}^{2} \left(\beta_2^{(n)} \lambda_n \right)^{1/2} \delta_n^{1/4}.$$

When this strategy is best?

It is best if:

$$\sum_{n=1}^{2}(\beta_2^{(n)}\lambda_n)^{1/2}\delta_n^{1/4} < (\sum_{n=1}^{2}\beta_2^{(n)}\lambda_n)^{1/2}(\sum_{n=1}^{2}\delta_n)^{1/4} \quad \text{Cauchy-Schwarz Inequality}$$

If the two customer types are similar, this inequality does not hold. Therefore, a symmetric strategy is best: items should be shipped together because with the higher demand density resulting from amalgamation vehicle tours can cover smaller zones and save operating costs. This is not always the case, however.

*The inequality holds when $\beta_2^{(1)}\lambda_1/\delta_1^{1/2} \neq \beta_2^{(2)}\lambda_2/\delta_2^{1/2}$

When this strategy is best?

$$\sum_{n=1}^{2}(\beta_2^{(n)}\lambda_n)^{1/2}\delta_n^{1/4} < (\sum_{n=1}^{2}\beta_2^{(n)}\lambda_n)^{1/2}(\sum_{n=1}^{2}\delta_n)^{1/4}$$

- This will hold if one set of suppliers is highly concentrated $\delta_1 \approx 0$ while producing many items that are expensive to store ($\lambda_1\beta_2^{(1)}$ large), and the other set has opposite characteristics (δ_2 is large but $\beta_2^{(2)} \approx 0$). 第一类供应商集中且存储费用高；第二类供应商更分散且存储费用更低
- Separate service for the two sets is then reasonable because the distribution strategies for both sets should be different. For the second set one would like to save operating costs at the expense of holding cost (one would use a large H in order to reduce the area served by each vehicle) and for the first set one would do the opposite. 第二类供应商应该用更低频率配送，以通过适量提高存储费用节省运输费用，另一类反之。

Combined delivery v.s. separate delivery

- In both cases the local operating costs plus the holding cost $((\beta_2^{(n)}\lambda_n)^{1/2}\delta_n^{1/4})$ would be close to zero. However, if both items types are combined together, neither of the factors on the right side $((\sum_{n=1}^{2}\beta_2^{(n)}\lambda_n)^{1/2}(\sum_{n=1}^{2}\delta_n)^{1/4})$ of is close to zero — service has to be moderately frequent because some of the items are expensive to store, and tours must cover moderate size areas because all destinations have to be visited.
- Clearly, the requirements of the two sets of customers interfere with each other, increasing cost dramatically.

现实世界的情况

- This phenomenon explains why **separate logistic systems are used to carry widely different items in real life**, even if from a transportation standpoint alone it would seem wise to combine them.
- It should not be surprising to find several transportation modes (taxis, limousines, buses, etc.) at the disposal of passengers exiting an airport. For freight transportation, the differences in the requirements of various customers are less likely to merit discriminating service; but the possibility should be considered.

差异化策略

- For general problems, the example just described suggests that cost may be reduced if the set of all customers is divided into classes with different characteristics, served with separate collection systems .
- For a given set of classes, total cost can be easily estimated — the cost and structure of near-optimal symmetric strategies would be used within each of our subsystems.
- The tricky part is **defining the customer subsets that will minimize total cost**. Daganzo (1985) presents a simple dynamic programming procedure to achieve this goal without detailed customer information — the method only uses the frequency (probability) distribution of customer characteristics — and shows in the process that the optimal solution would rarely exhibit more than 2 or 3 classes. When it is found that cost is minimized with only one class, discriminatory service is not cost-effective.

何时有效?

- Although we have ignored the pipeline inventory cost in this lecture, and have also assumed that the same transportation mode is used for all the subsystems, this is not a prerequisite for discriminatory service to be attractive.
- It is impossible to discuss here all the possible cases that can arise in detail, but a general statement can be made: if customers are very different, then we should check **whether dividing them into a few classes with (highly) different characteristics** — and serving them separately — can reduce cost; this is unlikely to result in much gain when customers are not very different, though.

- With the approach just described, each customer class n is designed separately and is characterized by design parameters A_n and H_n.
- By restricting these design parameters somewhat, Hall (1985) has developed a strategy that **allows customers from all classes to share the transportation fleet** while **being visited at different frequencies**. He requires A to be the same for all customers and each H_n to be an integer multiple of the time between dispatches H; that is, $H_n = m_n H$, for an integer m_n. He assumes that vehicles are dispatched at times $t = 0, H, 2H$, etc., visiting each time $(1/m_n)$th of the customers in every class n. This allows the effective stop density, $\sum_n \{\delta_n / m_n\}$, to be greater than for any class alone while ensuring that individual customers are only visited every m_n dispatches; it decreases the local transportation cost.

单个顾客的不同策略

- With the help of f_0, a variable denoting the fraction of customers served in each period, Hall's strategy can be defined without resorting to classes.
- Accordingly, the symbol "n" now reverts to its original meaning, indexing individual customers. We seek the optimal m_n for individual customers, as well as the optimal H and f_0. As done at the outset, let us assume that the conditions are such that vehicles will be dispatched full.

- Then, the line-haul motion cost per item is $\alpha_1/v_{\max}$, and does not depend on the allocation scheme for customers. The local motion cost per unit time and unit area is:

$$c_d k \frac{(f_0\delta)^{1/2}}{H} + c_s f^0 \frac{\delta}{H}.$$

- This somewhat conservative estimate assumes that stops are randomly and uniformly distributed within subregions of $\boldsymbol{R}$ larger than a collection district; it may be on the high side if customers of a similar kind cluster together.
- The holding cost per unit time in a subregion of unit are $\boldsymbol{P}$ is:

$$\sum_{n \in \boldsymbol{P}} c_h^{(n)}(m_n H) D'_n.$$

分解方法

- The system can be designed with a simple decomposition method. Conditional on f^0 and H, the local motion cost is fixed; thus, cost is minimized by the m_n's that minimize the holding cost. These m_n's, to be consistent with f^0, must satisfy:

$$\sum_{n \in \boldsymbol{P}} 1/m_n = f^0 \delta.$$

  Once the m_n have been found, the conditional total cost is obtained. Testing various values of f^0 and H, we can identify a near-optimal solution.
- Alternatively, if one replaces the constraint $[m_n = 1, 2, 3, \dots]$ by $[m_n > 1]$, a simple approximation for the minimal holding cost for a given f^0 and H can be obtained. The optimal strategy is then defined by the minimum over f^0 and H of the sum of this approximation and the local motion cost expression.

目录 |Outline

One of the reasons for the very extensive literature on algorithms to vehicle routing problems is that in actual applications almost every problem has some peculiarity that renders it unique. We have already seen that there can be a variety of cases depending on:

1. the relative size of the number of tours and the maximum number of stops per tour.
2. the relative cost of rent, inventory, and operating costs.
3. limitations to route length and storage space
4. dissimilarity in the values of items and the demand rates at different destinations
5. amount of uncertainty as to the customer lot sizes.

In addition (and this is not an exhaustive list) one might find situations in which time enters the problem because customers request service during certain “time windows”, or there is a limit to the amount of time an item can spend in transit (perishable items). There also are situations where vehicles do both distribution and collection (routing with backhauls), and situations where vehicle loading considerations make it advantageous to visit customers in an order which does not minimize the total distance traveled.

目录 |Outline

- At the core of our proposed two-step method for solving general distribution problems there should be a simple and efficient routing algorithm, whose performance can be quantified by means of simple formulas using average density as an input, instead of detailed customer locations. It is then a simple matter to add holding and pipeline inventory costs to the motion cost to define a logistic cost function. If routing/scheduling strategies can be defined in terms of a few decisions variables that are constrained only locally in the time-space domain, then the minimum of the (constrained) logistic cost function will approximate the cost generated by items in different portions of the time-space domain. The CA approach can be used.

- Some routing cost models that allow this to be accomplished already exist. They are now briefly reviewed. Simple transportation cost formulas have been proposed for time-window problems (Daganzo, 1987a,b). The results show how cost increases with the narrowness of the windows, and with the proportion of customers with tight requirements. The proposed routing strategy uses a different set of delivery districts for the customers in each time window, and staggers the zones in such a way so as to leave most vehicles in favorable locations at the beginning of each new window period.

- Perishable items such as newspapers (Han, 1984, and Han and Daganzo, 1986), lead to VRP structures which are similar to those arising from the vehicle route length limitations discussed in Sec. 4.4.1. The main difference is that service districts that are far away form the depot should be (i) more elongated than usual and (ii) covered in a one-way pass that begins at the end of the district that is close to the depot and terminates at the far end. Although this modification increases the line-haul distance traveled, it also allows distribution to begin sooner and the districts to include more stops.

- Models with both pick-ups and deliveries have been constructed for public transportation systems (Daganzo, Hendrickson and Wilson, 1977, Hendrickson, 1978) serving one focal point and a surrounding area. The strategies examined in these early works, however, are not as general as possible; they only consider two extreme cases for a partition of the surrounding area into service zones. More recently, Daganzo and Hall (1990) present an improved cost model for routing with backhauls, emphasizing cases where the total flow in one direction (e.g. outbound from the depot) is a few times larger than in the other direction.

- The basic idea is briefly summarized below for the case where the dominant flow is outbound; the reverse situation is similar. One simply constructs distribution tours as if there were no pickups, allocates each pickup to the nearest return leg of a distribution trip (or "spoke"), and finally modifies the vehicle tours in recognition of the newly assigned stops. Because the density of spokes increases rapidly toward the depot, significant vehicle deviations are only required for pickups near the outer fringe of the region. Pickup miles on the fringe can be reduced by ending the outermost delivery tours at the far end of their districts and by other modifications that are geared to optimize the spatial distribution of spokes. In fact, it is shown in Daganzo and Hall (1990) that under some conditions it is almost as if the secondary stops added only a stop cost and no distance cost. Hall (1993) has applied the concept of spokes to the VRP problem for deliveries only, in which customers demand large and small items.

- Another complication that deserves attention involves the interaction of vehicle loading and routing. When items have awkward shapes and are large, so that only a few fit in a vehicle, vmax may not be fixed; it may depend on the specific customers that are visited or even the order in which they are visited. The latter phenomenon may arise if weight distribution restrictions, for example, dictate that some items (and thus some stops) must be handled before others. This topic is very complex and hard to handle generally; see Hall (1989) and Ball et al. (1995a) for example.

目录 |Outline

Another area where further results may be desirable involves the interaction of physical distribution with production schedules. This interaction sometimes offers an opportunity for further cost reductions.

- This subject was broached in Sec. 4.3.3 (Inventory at the origin), where it was suggested that production of (destination-specific) items should be rotated among geographical customer regions every headway H. Dispatching the vehicles to a region immediately after its production run was completed greatly reduced the holding costs at the origin. It was assumed that production would be coordinated with transportation in this manner without much of a penalty.
- More likely, though, there may be a set-up cost associated with each switch in production item types. In this case production costs may be reduced by switching less frequently and holding higher inventories at the origin. An integrated solution can then be obtained by including in the logistic cost function the production set-up costs, e.g., as explained below.

- If no attempt is made to coordinate the production schedule with the physical distribution schedule, then the inventory at the origin of items of a certain type can be decomposed as shown in Figure into a (shaded) component which depends on the time between setups for that item type, H_s , and a (dotted) component which depends on the transportation headway, H:

$$\text{average inventory cost per item at origin} \approx \frac{c_i}{2} + \frac{c_i}{2}H.$$

  We are assuming that the number of item types is large and, therefore, the steps of the production curve are nearly vertical. Similar conclusions can be reached for few item types.

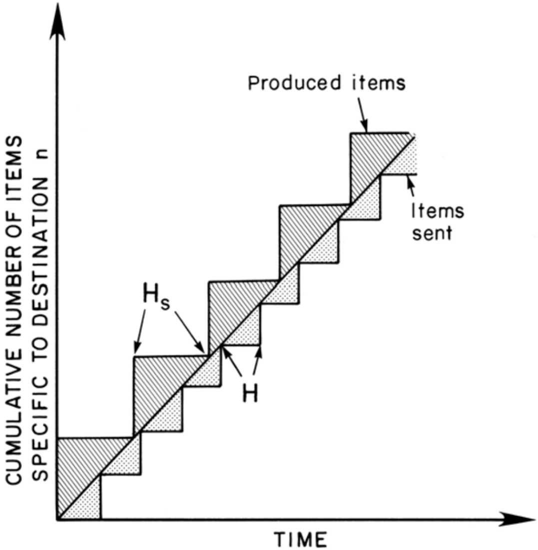


Figure: Inventory accumulation when no attempt is made to coordinate production and distribution

- The maximum accumulation also decomposes in a similar manner:

$$\text{maximum accumulation} \approx H_s D' + HD$$

- Because production costs depend on H_s and not on H, the sum of the production and logistics costs is made up of two components: (i) a production component with only production decision variables (including H_s), and (ii) a logistic component with only logistics variables (including A and H). Logistics and production decisions, thus, can be made independently of each other.

By selecting H to be an integer submultiple of H_s, or vice versa, it is possible to reduce the inventory time at the origin by an amount equal to the smallest of H and H_s, and the maximum accumulation becomes the difference between the maximum and the minimum of $H_s D'$ and HD'.

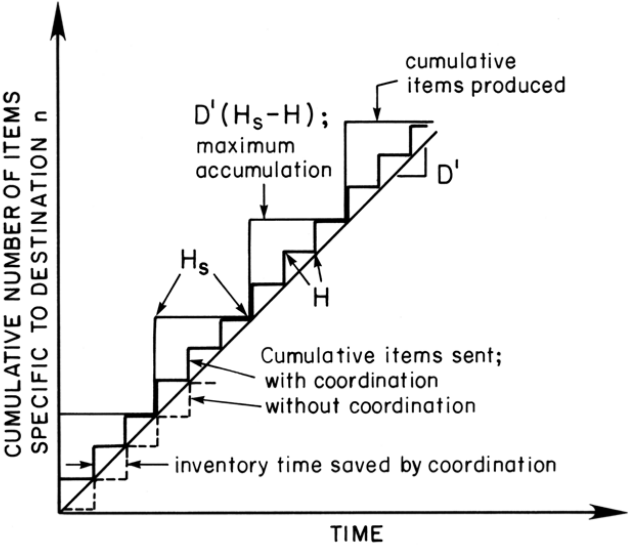


Figure: Inventory accumulation with coordinated schedules with $H_s = 3H$

- If this kind of coordination is feasible, the sum of the production and logistics costs no longer decomposes, and a coordinated production and distribution scheme should be considered.
- Blumenfeld et. al. (1985a) and (1986) have examined the case where each district is constrained to contain only one destination and all shipments are direct ($n_s = 1$). They illustrated situations where coordination of production and distribution is most conducive to cost savings, and provided a bound on the maximum possible benefit.
- Further research may be worthwhile to relax the $n_s = 1$ assumption and to allow more destinations than item types.

- Throughout this talk it was assumed that the total production rate (not just the schedule by item type) could be adapted to the changing demand without penalty. In practice, though, this is rarely so, even if the items produced are generic. (It is more costly to change the quantity of items produced than the kind of items produced because to adjust the production rate one needs to hire extra labor, pay overtime or fire labor as needed — and the penalty for these actions is large; Newell, 1990, has examined the production rate adjustment process.) To conclude this lecture, we show that this seemingly strong assumption can often be relaxed.

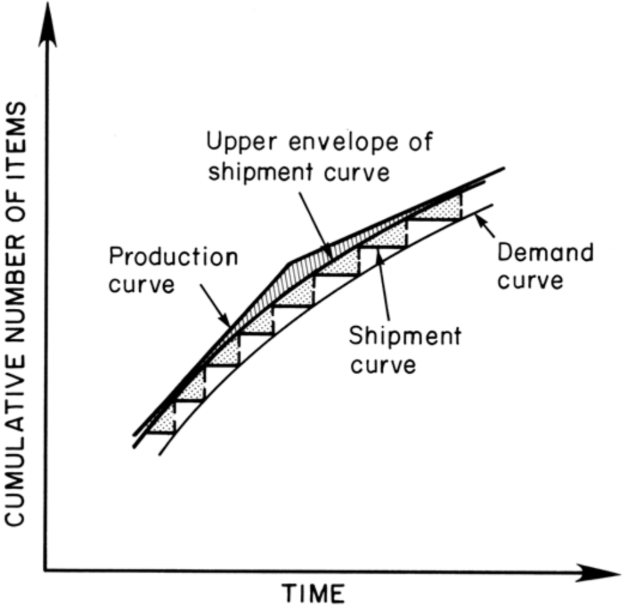


Figure: Production for a gradually decreasing demand

- The figure shows how a production curve may be adapted to a gradually decreasing demand; the objective is tracking the smooth envelope to the crests of the shipment curve (which varies like the demand curve) as closely as possible, without many production rate changes. We had already known that for a similar model described previously, lot size decisions were independent of production decisions; fortunately, this is also true now.
- In this figure, the inventory at the origin decomposes in two components: (i) a (shaded) component, which is due to the discreteness in the production rate changes and is independent of the shipping schedule, and (ii) a dotted component which is the same as if the production schedule was adjusted continuously as assumed in this chapter. Thus, costs can be divided into two components affected respectively only by production, or only by logistics decision variables.

Any questions?

Readings

- Daganzo. Logistics System Analysis. Ch.4.