物流系统分析
Logistics Systems Analysis

模块 2　1-1 配送系统
One-to-One Distribution Systems

葛乾

西南交通大学　系统所 & 交通工程系

Revisit of ideas of CA

- Accurate cost estimates can be obtained without precise, detailed input data,
- Departures from an optimal decision by a moderate percentage do not increase cost significantly. Since there is no need to seek the most accurate estimate of the optimum, there may be little use for highly detailed data
- Detailed data may get in the way of the optimization, actually hindering the search for an optimum
- Thus, the CA advocates a two-step solution approach to logistics problems: the first (analytical) step involves little detail and yields broad solution concepts; the second (or fine tuning) step leads to specific solutions, consistent with the ideals revealed by the first — it uses all the relevant detailed information.

本讲学习内容

1. One-to-one systems with constant production and consumption rates -> the robustness and accuracy of the results (Daganzo's work)
2. One-to-one systems with variable demand over time -> numerical methods and a continuous approximation (CA) analytical approach that is based on summarized data (Newell's work)
3. Extension of the CA approach to a location problem that has an analogous structure
4. The accuracy of the CA solutions
5. Extension of the CA approach to multidimensional problems with constraints
6. Network design issues.

目 录 |Outline

Let us now explore the optimization problem for the optimum shipment size, v^*:

$$z = \{Av + \frac{B}{v} : v \leqslant v_{\max}\} \tag{1}$$

- $A = c_h/D'$ 表示单位货物保管费用*, $B = c_f$ 表示每批货物固定运输费用。
- Consider first the case $v_{\max} = \infty$. Then v^* is the value of v which minimizes the convex expression $Av + B/v$. $v^* = \sqrt{B/A}$
- The optimum cost per item is: $z^* = (cost/item)^* = 2\sqrt{AB}$, which is easy to remember as "twice the square root of the product" of the terms in 1
- As a function of c_f, c_h and D', the optimum cost per item increases at a decreasing rate with c_f and c_h and decreases with the item flow D'. There are economies of scale, since higher item flows lead to lesser average cost.

*D' 表示单位时间生产量，单位货物的保管费用为 $c_r H_1 + c_i \bar{H} + c_i t_m$，$c_h = c_r + c_i$ 为租赁成本和在途等待成本之和。忽略常数项 $c_i t_m$，引入 $H_1 = \bar{H} = v/D'$ 得到表达式

We now examine the sensitivity of the resulting cost to errors in

- the decision variable, v
- the inputs (A or B)
- the functional form of the equation

Robustness in the Decision Variable

- Suppose that instead of v^* , the chosen shipment size is $v^0 = \gamma v$, where γ is a number close to 1, capturing the relative error in v^0. Then, the ratio of the actual to optimum cost z^0/z^* will be a number, γ', greater than 1, satisfying:

$$\gamma' = [A\gamma\sqrt{B/A} + \frac{B}{\gamma\sqrt{B/A}}]/[2\sqrt{AB}] = \frac{1}{2}[\gamma + \frac{1}{\gamma}] \tag{2}$$

- Independent of A and B, this relationship between input and output relative errors holds for all EOQ models.

Robustness in the Decision Variable (cont.)

- If γ is between 0.5 and 2, so that the optimal shipment size is approximated to within a factor of 2, then $\gamma' < 1.25$. If γ is between 0.8 and 1.25, then $\gamma' < 1.025 \rightarrow$ A cost within 2.5% of the optimum can be reached if the decision variable is within 25% of optimal.
- If γ is several times larger (or smaller) than 1, then the cost penalty is severe, i.e., $\gamma' \approx \gamma$ (or $\gamma' \approx 1/\gamma$)
- Obviously, while it is important to get reasonably close to the optimal value of the decision variable (say to within 20-40%), from a practical standpoint it may not be imperative to refine the decision beyond this level.

Robustness in the data errors (cont.)

- Let us now assume that one of the cost coefficients A (or B) is not known precisely. If it is believed to be $A' = \delta A$ (or $B' = \delta B$) , for some $\delta \approx 1$, then the optimal decision with this erroneous cost structure is:

$$v'^* = \begin{cases} \sqrt{B/A}\delta^{-1/2} = v^*\delta^{-1/2} & \text{if } A' = \delta A \\ v^*\delta^{1/2} & \text{if } B' = \delta B \end{cases}$$

- Because the actual to optimal shipment size ratio, v'^*/v^*, is either $\delta^{-1/2}$ or $\delta^{1/2}$, the cost penalty paid is as if $\gamma = \delta^{1/2}$. Thus, the resulting cost is even less sensitive to the data than it is to the decision variables

Robustness in the Data Errors (cont.)

- If the input is known to within a factor of 2 ($0.5 \leqslant \delta \leqslant 2$), then $0.7 \leqslant \gamma \leqslant 1.4$ and $\gamma' \leqslant 1.1$. The cost penalty would be about 10%, whereas before it was 25%. The penalty declines quickly as δ approaches 1
- This robustness to data errors is fortunate because the cost coefficients (for waiting cost especially)are rarely known accurately

Robustness in the Model Errors

- A cost penalty is also paid if the EOQ formula itself is inaccurate.
- To illustrate the impact of such functional errors, we assume that the actual cost, a complicated (perhaps unknown) expression, can be bounded by two EOQ expressions; the cost penalty can then be related to the width of the bounds.
- Suppose, for example, that the actual holding cost $z_h(v)$ is not exactly equal to the EOQ term (Av), but it satisfies:

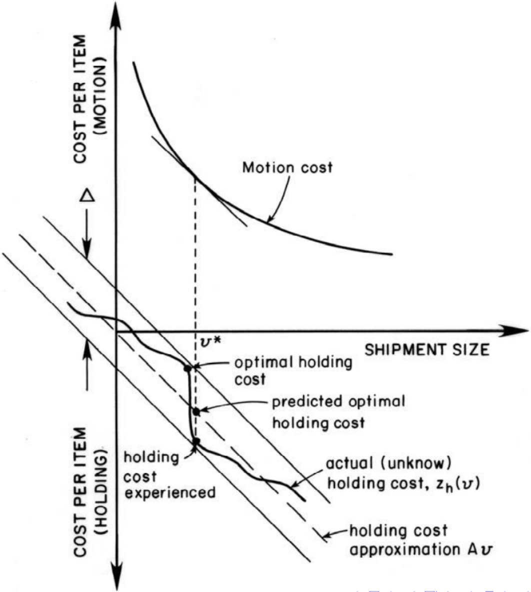
$$Av - \Delta/2 \leqslant z_h(v) \leqslant Av + \Delta/2 \tag{3}$$

for some small Δ. Such a situation could happen, for example, if storage space could only be obtained in discrete amounts. Because Δ is small, the EOQ lot size v^* is adopted.

Robustness in the Model Errors (cont.)

- The absolute difference between the actual cost $[z_h(v^*) + B/v^*]$ and the predicted EOQ cost z^* cannot exceed $\Delta/2$. It is also easy to see that the difference between the optimal cost with perfect information, $\min\{z_h(v) + B/v\}$, and z^* cannot exceed $\Delta/2$ either. As a result, the difference between the actual and theoretical minimum costs — the cost penalty is bounded by Δ.
- Usually, this penalty will be significantly smaller than the maximum possible
- If Δ is small compared to z^* (e.g., within 10%) the functional form error should be inconsequential. The same conclusion is reached if the motion cost is also inaccurate.
- In general, the EOQ solution will be reasonable if it is accurate to within a small fraction of its predicted optimal cost.

Unusual conditions generating the largest penalty



Error Combinations

多种误差的组合带来的总误差，并不一定是简单叠加

- If errors of the three types exist, one would expect the cost penalty to be greater. Fortunately though, when dealing with errors the whole (the combined penalty) is not as great as the sum of its parts
- Suppose for example that the lot size recipe is not followed very precisely (because, e.g., lots are chosen to be multiples of a box, only certain dispatching times are feasible, etc.) and that as a result 40% discrepancies are expected between the calculated and actual lot sizes. We have already seen that such discrepancies can be expected to increase cost by about 10%.

- Let us assume that one of the inputs (A or B) is suspected to be in error by a factor of 2, which taken alone would also increase cost by about 10%. Would it then be reasonable to expect a 20% cost increase? The answer is no; it should be intuitive that the penalty paid by introducing an input error when the lot size decision does not follow the recipe accurately should be smaller than the penalty paid if the decision follows the recipe.
- In our example, the combined likely increase is 14% [the square root of the sum of the squared errors: $0.14 = (0.1^2 + 0.1^2)^{1/2}$]. Statistical analysis of error propagation through models reveals similar composition laws in more general contexts.

Error Combinations (cont.)

- The previous example illustrated how input and decision errors propagate. Although model errors follow similar laws – the whole is still less than the sum of the parts – for some approximate models the results are surprising. The composed (data and model) error can be actually smaller than the data error alone with the exact model!
- This fortuitous （巧合的） phenomenon has a special significance because it arises when certain discontinuous models with discrete inputs are approximated by continuous functions and data*.

*Daganzo, C.F. (1987) "Increasing model precision can reduce accuracy" Trans.Sci. 21(2), 100-105.

Problems with constraints

前面分析了在忽略 $v < v_{\max}$ 条件时 EOQ 模型的稳健性和误差，同样的结论也适用于带约束的模型和拓展模型。

- The constrained EOQ solution is now presented rather briefly, before turning our attention to the lot size problem with variable demand.
- If we find that $v^* > v_{\max}$ in solving the unconstrained EOQ problem, then the solution is not feasible. Choosing $v = v_{\max}$ is optimal. Hence, the optimal EOQ solution can be expressed as:

$$v^* = \min\{\sqrt{B/A}, v_{\max}\}$$

and the optimal cost per item

$$z^* = \begin{cases} 2\sqrt{AB} & \text{if } \sqrt{B/A} \leqslant v_{\max} \\ Av_{\max} + B/v_{\max} & \text{if } \sqrt{B/A} > v_{\max} \end{cases}$$

Problems with constraints (cont.)

- Note that z^* is an increasing and concave function of A, and also of B.
- As $A = c_h/D'$, z^* is decreasing a function of D' and convex; the economies of scale continue to exist for all ranges of D'.
- The total cost per unit time, $D'z^*$, is proportional to $D'^{1/2}$ until the capacity constraint is reached, and from then on increases linearly with D'. The critical point is $D'_{crit} = (v_{\max})^2 c_h/c_f$*

*固定 B，则总成本为 A 的函数，分段函数的间断点为 $\sqrt{B/A} = v_{\max} \rightarrow A = \frac{B}{v_{\max}^2}$。代入 $A = c_h/D'$，可得 $D'_{crit} = (v_{\max})^2 c_h/c_f$

Optimal EOQ cost as a function various parameters

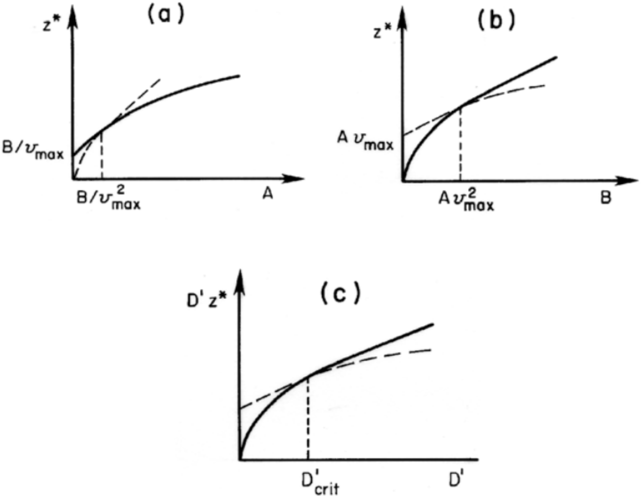


Figure: Optimal EOQ cost as a function various parameters: (a) holding cost per item, A; (固定 B) (b) fixed motion costs, B; (固定 A) and (c) demand rate, D'. Dashed lines are the unused branches of z^*

目 录 |Outline

Problem with variable demand （变需求下的 EOQ 问题）

接下来考虑在有限时间区间内，消费量以可预测的方式变化的情况下的 EOQ 问题。

- The demand pattern is characterized by a function $D(t)$ that gives the cumulative number of items demanded between times 0 (the beginning of the study period) and t. The time derivative of this function $D'(t)$ represents the variable demand rate.
- We then seek the set of times when shipments are to be received $(t_0 = 0, t_1, \ldots, t_{n-1})$, and the shipment sizes $(v_0, v_1, \ldots, v_{n-1})$, that will minimize the sum of the motion plus holding costs over our horizon, $t \in [0, t_{\max}]$.
- As previously, we also define as **inputs** to our problem a **fixed (motion) cost per vehicle dispatch** c_f, a **holding cost per item-time** $c_h = c_r + c_i$, and a **maximum lot size** $v_{\max}$. With an infinite horizon and a constant demand, $D(t) = D't$, this formulation reduces to the EOQ problem examined in previous sections.

Solution when holding cost $\approx$ rent cost

- If inventory cost is negligible, $c_i \ll c_r$, then holding cost approximately equals rent cost $c_h \approx c_r$. We have already mentioned that rent cost increases with the maximum inventory accumulation*, and that otherwise the cost is rather insensitive to the accumulations at other times. This property of holding cost simplifies the solution to our problem.
- Recall that given a set of n shipments, the motion cost during the period of analysis, $c_f n$, is independent of the shipment times and sizes†. The problem is then to find the sets of shipment times and sizes that will minimize holding cost.

*$c_r \times D' H_1$ 与保存所发生的时间无关

†移动的总费用与次数无法

Solution when holding cost ≈ rent cost (cont.)

- A lower bound to the maximum accumulation at the destination is the size of the largest shipment received. (Why?*)
- This lower bound minimized when all the shipments are equal.(Why?†)
- Hence, the largest shipment – and, thus, the maximum accumulation – must exceed or at least equal $D(t_{\max})/n$, the set is an optimal way of sending n shipments with rent cost per unit time: $c_r D(t_{\max})/n$‡.
- Each shipment is just large enough to meet the demand until the next shipment; the consumption between consecutive receiving times, the same in all cases, is $D(t_{\max})/n$§.

*在目的地的积累量的一个下界为最大批量，意味着该批货物到达时刚好无存货

†每个批量相等，意味着最大批量即为平均批量。此时最大批量最小

‡每个批量相等且 n 次配送的总量至少为 $D_{\max}$，则显然每次批量 $\geqslant D(t_{\max})/n$

§每次的送货量刚好能满足两次配送之间的需求量

Solution when holding cost $\approx$ rent cost (cont.)

Clearly the following strategy is optimal:

- Divide the ordinate axis between 0 and $D(t_{\max})$ into n equal segments and find the times t_i for which $D(t)$ equals $(i/n)D(t_{\max})$ for $i = 0, \ldots, n-1$. These are the shipment times,
- Dispatch barely enough to cover the demand until the following shipment.

One must now find the optimal n by minimizing the resulting cost

$$cost/time = c_r[D(t_{\max})/n] + c_f[n/t_{\max}]$$

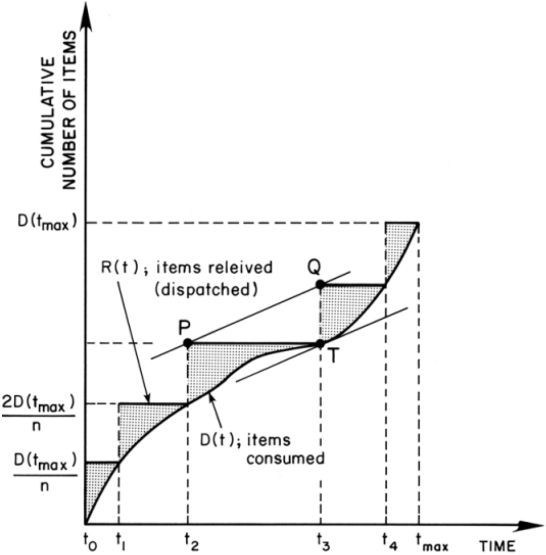
$$cost/item = (\frac{c_r}{\bar{D}'})(\frac{D(t_{\max})}{n}) + c_f[n/D(t_{\max})]$$

where D' is the average consumption rate $\bar{D}' = D(t_{\max})/t_{\max}$

Solution when holding cost $\approx$ rent cost (cont.)

- Note that the formulation is the EOQ expression with $v = D(t_{\max})/n$. The solution now requires that n be an integer (there are constraints on v), but we have already seen that any v close to the unconstrained v^* is near optimal. As a result, unless the time horizon is so short that $n^* = 1$ or 2, the optimal cost per item should be close to the cost with constant demand
- If $v_{\max} < \infty$, the solution procedure does not change. It is still optimal to have equal shipment sizes, but the number of shipments should be large enough to satisfy: $D(t_{\max})/n < v_{\max}$. The solution is still of the same form, with v^{-1} restricted to being an integer multiple of $D(t_{\max})$-1.

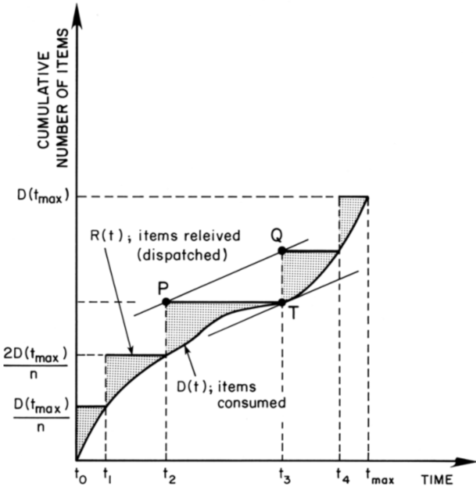
Solution procedure



Solution when rent cost is negligible

另一种情况与之前相反，即租金可忽略，但是在途的保管费用不可忽略。

- This situation occurs when items are so small and expensive, that most of the holding cost arises from the item-hours spent in inventory, and not from the rent for the space to hold them.
- In this case the destination's holding cost should be proportional to the shaded area of right figure

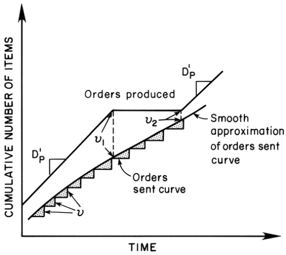


Solution when rent cost is negligible

The combined origin-destination holding cost will also be proportional to the shaded area:

- if the origin holding cost can be ignored <— if the origin produces generic items for so many destinations that the part of its costs that would be prorated (按比例分配) to each destination is negligible.

- if the origin holding cost is proportional to the area. <— if the production strategy at the origin is as described in figure below. The total wait at the origin that can be attributed to the shipping strategy must be similar to that of the destination; i.e., it would also be proportional to the shaded area

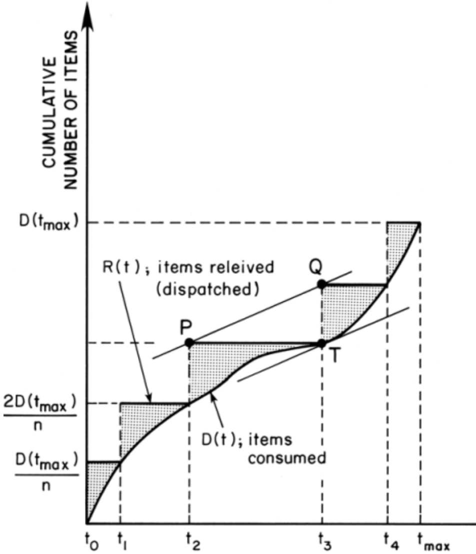


- for typical passenger transportation systems

Solution when rent cost is negligible

- When holding costs are proportional to the shaded area, they are no longer a function of n alone.
- For a set of points $(t_1 \dots t_{n-1})$ to be optimal, each line $\overline{PQ}$ must be parallel to the tangent line to $D(t)$ at the receiving time[a] (point T in the figure)
- We may verify that if this condition is not satisfied, then it is possible to reduce the total shaded area by either advancing or delaying the receiving time by a small amount.

[a] Newell (1971)



Solution when rent cost is negligible (cont.)

- Unfortunately, the smallest shaded area - and thus the waiting cost - no longer can be expressed as a function of n alone, independently of $D(t)$.
- Thus, it seems that a simple expression for the optimal cost cannot be obtained for any $D(t)$

Numerical solution – 滚动时域优化思路

- It can be formulated as a rolling horizon optimization problem in which a shipment time, t_i, is chosen at each stage $(i = 1, \ldots, n-1)$, and where the state of the system is the prior shipment time, t_{i-1}. The optimization procedure yields an optimum holding cost for a given n, $z_i^*(n)$, which can be substituted for the first term of the following equation to yield n^*.

$$cost/item = z_i^*(n) + c_f(n/D(t_{\max}))$$

Numerical solution — Newell's method

The following procedure is less laborious and works particularly well if $D(t)$ is smooth, without bends or jumps (refer to figure for the explanation)

1. Choose a point P_1 on the ordinates axis and move across to T_1
2. Draw from P_1 a line parallel to the tangent to $D(t)$ at T_1, and draw from T_1 a vertical line. Label the point of intersection P_2

Steps (i) and (ii) identify a point P_2 from a point P_1. They should be repeated to identify P_3 from P_2, P_4 from P_3, etc., defining in this manner a receiving step curve, $R(t)$. If $R(t)$ does not pass through the end point, $(t_{max}, D(t_{max}))$, the position of P_1 should be perturbed until it does.

Numerical Solution — Newell's method (cont.)

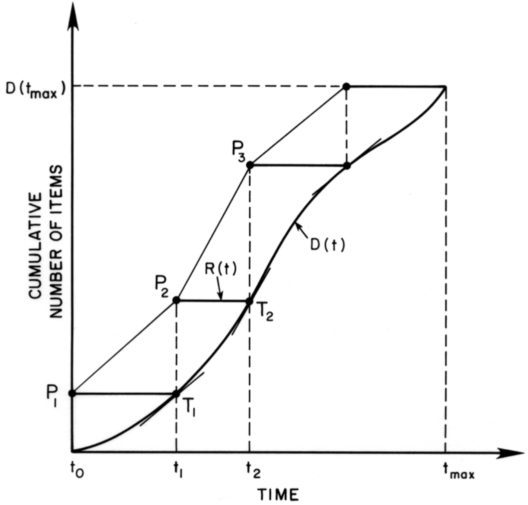


Figure: Construction method for the cumulative number of items shipped versus time

The Continuous Approximation Method

- The CA method replaces the search for $\{t_i\}$ by a search for a continuous function, whose knowledge yields a set of t_i with near minimal cost
- It works well when $D'(t)$ does not change rapidly; i.e., if $D'(t_i) \approx D'(t_{i+1})$ for all i. A byproduct is a simple expression and decomposition principle for the total cost

The Continuous Approximation Method (cont.)

- Let us assume that an optimal solution has been found, and denote by I_i the $i-$th interval between consecutive receiving times: $[t_{i-1}, t_i), i = 1, 2, \ldots.$
- Then divide the total cost during the study period into portions "$cost_i$" corresponding to each interval. That is, "$cost_i$" includes the cost, c_f, of dispatching one shipment plus the product of c_i and the shaded area for interval I_i

$$cost_i = c_f + c_i \times area_i$$

- Clearly, the sum of the prorated costs will equal the total cost. Since $D'(t)$ is continuous, it should be intuitive that there is a point t_i' in each interval I_i for which the area above $D(t)$ satisfies:

$$area_i = \frac{1}{2}(t_i - t_{i-1})^2 D'(t_i')^*$$

*阴影区域的面积等于需求函数在 t_i' 处取值时的，其中横轴长度为 $t_i - t_{i-1}$，纵轴为 $D'(t_i')(t_i - t_{i-1})$

The Continuous Approximation Method (cont.)

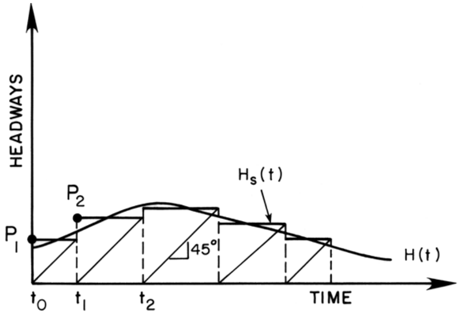
为什么 $[t_{i-1}, t_i)$ 区间内一定存在这样的 t_i'?

- Consider the triangle defined by the horizontal and vertical lines passing through a point P_i in the figure and a straight line passing through T_i with a slope that yields "$area_i$" for the triangle; i.e. slope $D'(t_i')$.
- Since such a slanted line must intersect $D(t)$ (otherwise the areas above $D(t)$ and above the slanted line could not be equal) there must be a point between T_i and the point of intersection where the two lines have the same slope. The abscissa (横坐标) of this point is t_i'.

Therefore we can write:

$$area_i = \frac{1}{2}(t_i - t_{i-1})^2 D'(t_i') = \int_{t_{i-1}}^{t_i} \frac{1}{2}(t_i - t_{i-1}) D'(t_i') \mathrm{d}t$$

The Continuous Approximation Method (cont.)



If we now define $H_s(t)$ as a step function such that $H_s(t) = t_i - t_{i-1}$ if $t \in I_i$ (see the figure above for example), then the cost per interval can be expressed as:

$$cost_i = \int_{t_{i-1}}^{t_i} [\frac{c_f}{H_s(t)} + \frac{c_i H_s(t)}{2} D'(t_i')]\mathrm{d}t.$$

Note that this is an exact expression.

The Continuous Approximation Method (cont.)

If we now approximate $D'(t_i')$ by $D'(t)$ – which is reasonable if $D'(t)$ varies slowly – the total cost over the whole study period can be expressed as the following integral:

$$cost_i = \int_{t_{i-1}}^{t_i} [\frac{c_f}{H_s(t)} + \frac{c_i H_s(t)}{2} D'(t)] \mathrm{d}t.$$

We seek the function $H_s(t)$, which minimizes the equation above. Unfortunately, this is akin to determining the $\{t_i\}$ themselves. A closed form solution can be obtained if $H_s(t)$ is replaced by a smooth function, $H(t)$. That is:

$$cost_i \approx \int_{t_0}^{t_{\max}} [\frac{c_f}{H(t)} + \frac{c_i H(t)}{2} D'(t)] \mathrm{d}t.$$

Now, instead of finding $H_s(t)$, we can find the $H(t)$ which minimizes the new equation - a much easier task - and then choose a set of shipment times (i.e., $H_s(t)$) consistent with $H(t)$.

The Continuous Approximation Method (cont.)

Clearly, the $H(t)$ which minimizes the RHS minimizes the integrand （被积项）at every t; thus:

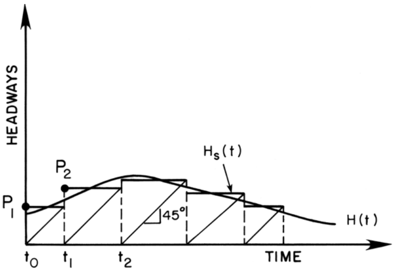
$$H(t) = [2c_f/(c_i D'(t))]^{1/2}.$$

This is the time between dispatches (headway) for the EOQ problem with constant demand $D' = D'(t)$.
A set of shipment times consistent with $H(t)$ can be found easily since $H(t)$ varies slowly with t.

The Continuous Approximation Method (cont.)

The figure suggests how this can be done systematically.

- Starting at the origin (point t_0) draw a 45°line and find a horizontal segment from a point on the vertical axis, such as P_1 in the figure, to the intersection with the 45°line.
- The elevation of P_1 should be such that the area below the segment equals the area below $H(t)$.
- The abscissa of the point of intersection is the next shipment time, t_1. This locates t_1, given t_0.
- The construction is then repeated from t_1 to locate t_2, from t_2, to locate t_3, etc.



In practice one does not need to be quite so precise, since we have already seen that small deviations from optimality have a minor effect.

The Continuous Approximation Method (cont.)

Now we may calculate the total cost given the optimal $H(t)$.

$$\text{Total cost} \approx \int_{t_0}^{t_{\max}} [2c_i c_f D'(t)]^{1/2} \mathrm{d}t.$$

The integrand of this expression is the optimal EOQ cost per unit time if $D' = D'(t)$. Note that the integrand in the equation can be written as

$$[2c_i c_f / D'(t)]^{1/2} [D'(t) \mathrm{d}t]$$

where the first factor represents the optimal cost per item for an EOQ problem with constant demand, $D'(t)$. The average cost per item (across all the items) is obtained by dividing the total cost by the total number of items $D(t_{\max}) = \int_{t_0}^{t_{\max}} D'(t) \mathrm{d}t$.

The Continuous Approximation Method (cont.)

The result is:

$$(\text{cost}/\text{item})^* = \frac{\int_{t_0}^{t_{\max}} [2c_i c_f D'(t)]^{1/2} \mathrm{d}t}{\int_{t_0}^{t_{\max}} D'(t) \mathrm{d}t}$$

- In practical terms the (cost/item)* expression indicates that the average optimal cost per item can be obtained by averaging the cost of all the items, as if each one of these was given by the EOQ formula with a (constant) demand rate equal to the demand rate at the time when the item is consumed*.
- The total cost expression indicates that, given a partition of $[0, t_{\max}]$ into a collection of short time intervals, the optimum cost can be approximated by the sum of the EOQ costs for each one of the intervals considered isolated from the others*.
- These equations are so simple that they can be used as building blocks for the study of more complex problems in following lectures.

This is one of the attractive features of the CA approach; it yields cost estimates without having to develop, or even define, a detailed solution to the problem.

*商品的最优平均成本可以通过对所有商品成本的均值求出，而每件成本的成本由 EOQ 公式在该商品被消费时的取值得出。

*总成本的解释也是类似。假设 $[0, t_{\max}]$ 之间被划分成若干短时间段，最优的总成本可通过每个孤立时间段的 EOQ 成本之和近似

The Continuous Approximation Method (cont.)

- The CA approach can also be used to locate points on any line (time or otherwise) provided that the total cost can be prorated approximately to (short) intervals on the line, while ensuring that the prorated cost to any interval only depends on the characteristics of said interval. In the previous discussion, the integrand in the cost equation $cost_i \approx \int_{t_0}^{t_{\max}} [\frac{c_f}{H(t)} + \frac{c_i H(t)}{2} D'(t)] \mathrm{d}t$ is the prorated cost in $[t, t + \mathrm{d}t)$, which does not depend on the demand rate outside the interval
- The CA approach can also be used to locate points in multidimensional space, when the total cost can be expressed as a sum of neighborhood costs dependent only on their local characteristics. Newell (1973) argues that the CA approach is comparatively more useful then, because in the multidimensional case it is much more difficult for exact numerical methods to deal with the complex boundary conditions that arise. Because the CA approach will be used in forthcoming lectures repeatedly, the next section discusses two additional (one-dimensional) examples.

目录 |Outline

Bus departure schedule problem

The CA technique was originally proposed to find a near-optimal bus departure schedule from a depot (Newell, 1971).

Given the cumulative number of people $D(t)$ demanding service by time t, the fixed cost of a bus dispatch c_f, and the cost of each person-hour waited c_i, the objective was to minimize the sum of the bus dispatch (motion) and waiting (holding) costs.

With an unlimited bus capacity, this problem is almost identical to the one we have just solved; except for $D(t)$, which now represents the cumulative number of people (items) entering the system and not the number leaving. The cost equations still hold.

Freight terminals location on a distance line

- This problem locates freight terminals on a distance line between 0 and $d_{\max}$. This interval contains origins, which send items to a depot. The distance line extends from the origin, O, to a depot, located at $d = \tilde{d} \geqslant d_{\max}$.
- The flow of freight (number of items per day) that originates between O and d is a function of d, $D(d)$, which increases from 0 to v_{tot}. Items are individually carried to the terminals at a cost c'_d per unit distance per item. Each day a vehicle travels the route collecting the items accumulated at each terminal and takes them to the depot.

The access cost

- The motion cost for this operation has three components: the handling cost at the terminals, assumed to be constant and therefore ignored, the access cost to the terminals, and the line-haul cost of operating the vehicle from the terminals to the depot. 构成：枢纽处的处理成本，可视为常数并忽略；到达枢纽的成本；干线运输的成本，即从枢纽直运到目的地的费用
- The access cost is given by the product of c'_d and the total item-miles of access traveled per day; it increases with the separation between stops.

The line-haul cost

The line-haul cost has the form of

$$\text{line-haul cost/day} = c_s(1 + n_s) + c_d(\tilde{d}) + c_s'(v_{\text{tot}})$$

- c_s, is the cost attributable to each trip, regardless of distance and shipment composition; it includes the cost of stopping the vehicle and having it sit idle while it is being loaded and unloaded. Think of it as the fixed cost of stopping, independent of what is being loaded and unloaded.
- n_s is the number of stops (excluding the depot)
- c_d is the cost of vehicle-mile. It is the vehicle cost (including the driver) for distance traveled regardless of the vehicle's contents;
- c_s', represents the cost of carrying items. It represents a penalty for delaying the vehicle while loading and unloading the items, as well as the cost of handling the items within the vehicle. v_{tot} is the total size of the shipment arriving at the depot.

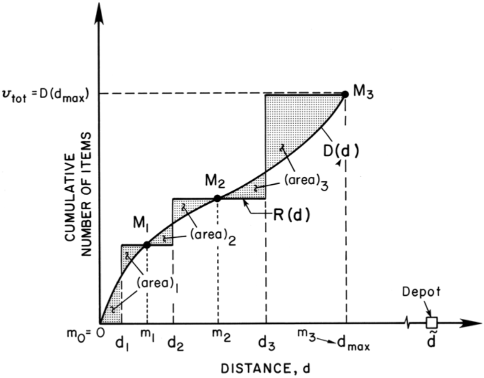
The line-haul cost (cont.)

Note that the line-haul cost does not depend on the specific stop locations and that in contrast to the access cost, it increases with n_s.

$$\text{line-haul cost} = c^o + c_s n_s$$

where c^o is a constant that will be ignored for design purposes.
As the problem has been formulated, with one trip per day, the sum of the holding costs at all stops can be ignored - consideration reveals that the sum is constant. Pipeline inventory costs do depend on the decision variables (they should increase with n_s) but for cheap freight the effect is negligible. Thus, all inventory and holding costs are neglected.
The stops will be located as the result of a trade-off between line-haul and access costs. Without this simplification, which is inappropriate for passenger transportation, the problem is equivalent to the transit stop location problem.

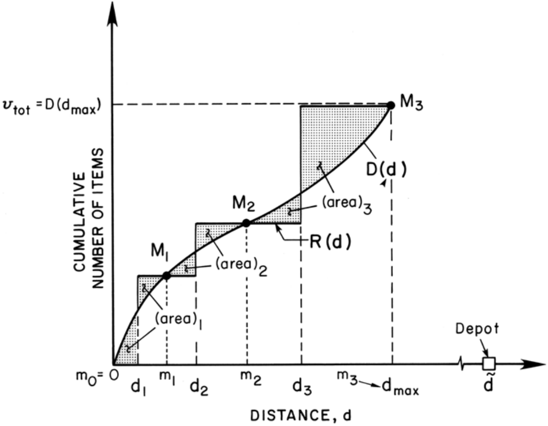
示意图



The figure depicts the location of three terminals (at points d_1, d_2, and d_3) and a curve, $R(d)$, depicting the number of items in the vehicle as a function of its position. This curve increases in steps at each terminal location. The size of each step equals the number of items collected.

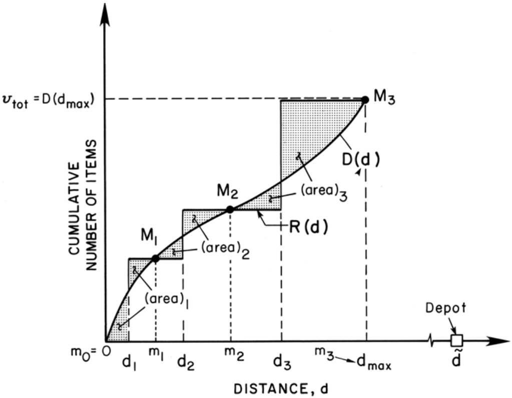
To minimize access (and total) cost each item is routed to the nearest terminal, and as a result the step curve passes through the midpoints, M_i, shown in the figure.

思考一下，为什么需要考虑中转站的中点?



计算中转站的中点是为了划分每个枢纽的服务范围. 例如, $[m_0, m_1]$ 范围内的货物由在 d_1 处中转, $[m_1, m_2]$ 范围内的货物由在 d_2 处中转等, 因为对应的枢纽距离货物位置最近。

The coordinates of M_i are $m_i = (d_i + d_{i+1})/2$ and $D(m_i)$; with $m_0 = 0$ and $m_{n_s} = d_{\max}$



Let us see how the total cost can be prorated to short intervals, by considering the partition of $(0, d_{\max}]$ into the following intervals surrounding each terminal: $I_1 = (0, m_1], I_2 = (m_1, m_2], \ldots, I_{n_s} = (m_{n_{s-1}}, d_{\max}]$. Each interval, I_i, adds an access cost proportional to the daily item-miles traveled for access to terminal i.

This is given by the shaded area on the two quasi-triangular segments next to the location of the terminal, $(area)_i$, thus

$$\text{access cost}_i = (area)_i c'_d. \qquad \text{平均的运送距离是}(m_i - m_{i-1})/4$$

For slowly varying $D(d)$, the access cost can be rewritten as:

$$\text{access cost}_i \approx \frac{1}{4}(m_i - m_{i-1})^2 D'(d_i) c_d'.$$

Since each terminal adds c_s to the daily line-haul cost, the share of the total cost prorated to I_i is:

$$(\text{Total cost per day})_i \approx c_s + \frac{c_d'}{4}(m_i - m_{i-1})^2 D'(d_i).$$

Since $D'(d) \approx D'(d_i)$ for $d \in I_i$ (we stated that $D'(d)$ varied slowly), the above expression can be approximated by:

$$(\text{Total cost per day})_i \approx \int_{m_{i-1}}^{m_i} \{\frac{c_s}{m_i - m_{i-1}} + \frac{c_d'}{4}(m_i - m_{i-1}) D'(d)\} \mathrm{d}d.$$

If we now let $s(d)$ denote a slowly varying function such that $s(d_i) = m_i - m_{i-1}$ (the function, used later to locate the terminals, indicates the size of a terminal's influence area depending on location), then we can rewrite the last expression once again, using $s(d)$ instead of $mi - m_{i-1}$:

$$\text{(Total cost per day)}_i \approx \int_{m_{i-1}}^{m_i} \{\frac{c_s}{s(d_i)} + \frac{c_d'}{4}(s(d_i))D'(d)\}\mathrm{d}d.$$

The total cost for the system is then:

$$\text{(Total cost per day)} \approx \int_{0}^{d_{\max}} \{\frac{c_s}{s(d)} + \frac{c_d'}{4}(s(d))D'(d)\}\mathrm{d}d.$$

Analytical form of EOQ

The least cost $s(d)$ minimizes the integrand at every point; given its EOQ analytical form, we find

$$s(d) \approx 2[\frac{c_s}{c'_d D'(d)}]^{1/2}.$$

Note that if D' varies slowly, $s(d)$ will vary slowly as we had assumed.
The expressions for the minimum total and average (per item) cost as follows

$$\text{(Total cost per unit time)}^* \approx \int_0^{d_{\max}} [c_s c'_d D'(d)]^{1/2} \mathrm{d}d$$

$$\text{cost per item}^* \approx \int_0^{d_{\max}} [c_s c'_d D'(d)]^{1/2} D'(d) \mathrm{d}d / \int_0^{d_{\max}} D'(d) \mathrm{d}d$$

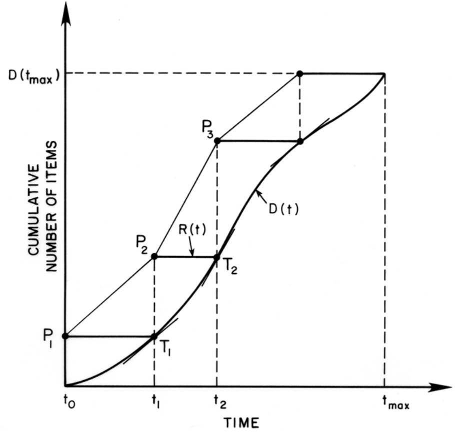
Locate the terminals

- To locate the terminals, one first divides $(0, d_{\max}]$ into non-overlapping intervals of approximately correct, length I_1, I_2, etc.. . . , by starting at one end and using the EOQ formula repeatedly.
- If the last interval is not of correct length, then the difference can be absorbed by small changes to the other intervals.
- If $d_{\max}$ is large (so that there are at least several intervals), then the final partition should satisfy $s(d) \approx m_i - m_{i-1}$ if $d \in I_i$, and the approximations leading to the equations should be valid.
- With the influence areas defined in this manner, the terminals are located next. They should be positioned within each interval so that the boundary between neighboring intervals is equidistant from the terminals.
- For a general sequence of intervals (e.g., of rapidly fluctuating lengths) this may be difficult (even impossible) to do, but for our problem with $|I_i| \approx |I_{i+1}|$ the best locations should be near the center of each interval; in fact little is lost by locating the terminals at the centers.

目录 |Outline

- Although a systematic analysis of its errors has not been reported, experience indicates that the CA approach is very accurate when the descriptive characteristics of the problem ($D'(t)$ in the text's examples) vary slowly as assumed.
- The approach is also robust. It is effective even if the variation in conditions is fairly rapid –in our case, accurate results are obtained even if $D'(t)$ varies by a factor of two within the influence areas. This conclusion is not surprising in light of the EOQ robustness discussed in the previous lecture
- When conditions are unfavorable, the CA method can both over- and under-predict the optimal cost. The textbook provides two examples identify said conditions, with the first example illustrating overestimation and the second underestimation. The basis for comparison will be the exact solution, which for our problem can be obtained readily.

Recall the construction method



Recall our previous construction method for the cumulative number of items shipped versus time

- Choose a point P_1 on the ordinates axis and move across to T_1
- Draw from P_1 a line parallel to the tangent to $D(t)$ at T_1, and draw from T_1 a vertical line. Label the point of intersection P_2 .

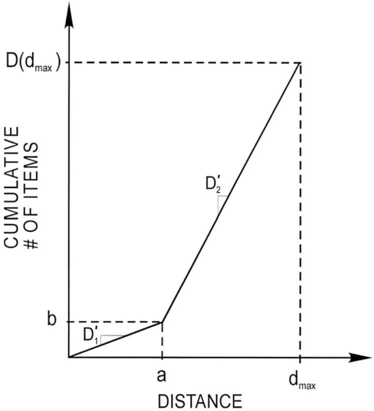
之前例子中横轴是时间，中转枢纽问题中横轴是距离

Construction method for the terminal location problem

- Given # of stops n_s, for a set of locations to be optimal the line $D(d)$ of the distance figure must bisect in two equal halves every vertical segment of $R(d)$. Otherwise, the terminal (e.g., terminal 3) could be moved slightly to decrease access cost. The optimal solution can then be found by comparing all the possible $R(d)$ with the above property.
- For a given d_1, draw a vertical step that is bisected by $D(d)$, and move across horizontally so that the horizontal segment is also bisected by $D(d)$. This identifies d_2. Repeat the construction to find d_3, d_4, etc. (Only those values of d_1 for which the last vertical segment is bisected by $D(d)$ need to be considered seriously.) The optimal solution corresponds to a d_1 which minimizes the sum of the stop cost and access cost.
- The procedure is so simple that it can be implemented in spreadsheet form*.

*The user selects d_1 and the spreadsheet returns the graphs, and the cost; it is then easy to find the solution either interactively or automatically with the computer.

Example 1: Overestimation



Terminals are to be located on two adjoining regions with high and low demand. The left figure depicts a generic piece-wise linear cumulative demand curve of this type. The coordinates of the break-point (distance, item number) are given by parameters "a" and "b". They must be consistent with the specified values for $d_{\max}$, $D(d_{\max})$, D'_1 and D'_2. For this problem the continuum approximation approach yields

$$TC^* \approx (c_s c'_d)^{1/2}\{a\sqrt{D'_1} + (d_{\max} - a)\sqrt{D'_2}\}$$

Parameters

A possible set of parameters is $d_{\max} = 500, D(d_{\max}) = 1700, D'_1 = 1, D'_2 = 5, a = b = 200, c'_d = 1$ and $c_s = 160,000$. This choice has been made because a systematic analysis shows that it produces the largest overprediction error in%age terms. The predicted cost is: $TC^* = 348,328$.

In actuality the least possible cost is 8% smaller. It arises when a single terminal is located at $d = 330$. The reader can verify that the exact access cost for this location is 160,500 units. Since the terminal cost is 160,000 units (for one terminal), the grand total is 320,500 $<$ 348,328.

Insights from the example

This rather extreme example illustrates that the CA approach can overestimate the optimum cost.

- To understand why this happens let us decompose the CA costs into its components. Note first that the ideal spacing between terminals predicted by the CA method with is:
  $s(d) = 2 * [160,000/(1 * 1)]^{1/2} = 800$ in the low demand section, and
  $s(d) = 2 * [160,000/(1 * 5)]^{1/2} \approx 357$ in the high demand section.
- The CA access cost is calculated as if the average access distance was $s(d)/4 = 200$ in the low demand section and 89.25 units in the high demand section. Since there are 200 items in the low density region and 1500 in the high density region, the total CA access cost is approximately: $200 \times 200 + 89.25 \times 1500 = 173,875$.
- The CA stop cost (line-haul cost) is calculated by integrating the density of terminals over the service region, $(200/800 + 300/357) \approx 1.09$, and multiplying this result by the cost of a terminal: $1.09 \times 160,000 = 174,400$.
- The grand total is therefore: $173,875 + 174,400 = 348,275 \approx 348,328$.

Insights from the example (cont.)

It turns out, however, that just a single terminal in the high density region can serve both, the low density points with an average distance barely greater than the CA access distance, and the high demand section with an average access distance considerably inferior to the corresponding CA distance. For our chosen location ($d = 330$) the actual average access distances are: 230 units for the low density section (200 with the CA method)and 76 for the high density section (89 with CA method). Since we are using only one terminal, the final cost is lower.
The overprediction effect arises because the demand curve varies significantly and very favorably between the terminal and the edge of the service region, and the CA approach does not exploit this variation. The variation is so favorable that it allows a terminal provided for the high density points to double up efficiently as a terminal for the low density points.
Favorable conditions are unusual, however. When the demand does not vary rapidly the CA approach consistently underestimates demand.

Example 2: Underestimation example

- By its nature, the CA approach ignores that the number of terminals must be an **integer**; any situation with a finite region size (or time horizon) will exhibit this error type.
- To exclude the overprediction error type illustrated by example 1, the demand per unit length of region is set constant: $D'(d) = D'$. This also allows closed form comparisons to be made.
- The CA solution is $TC^* = \sqrt{c_s c'_d}\sqrt{D'} d_{\max}$. Without losing generality, we choose the units of distance, item quantity and money so that $d_{\max} = 1, D(d_{\max}) =$ 1 and $c_s = 1$. Thus, $D' = 1$ and only the parameter c'_d remains. The above expression becomes: $TC^* = \sqrt{c'_d}$
- If the exact optimal solution has n_s terminals, the distance line will be partitioned into n_s intervals of equal length: $I_i = ((i-1)/n_s, i/n_s]$. The total cost is then

$$TC(n_s) = n_s + 2n_s\{(\frac{1}{2n_s})^2 \frac{c'_d}{2}\} = n_s + \frac{c'_d}{4n_s}$$

  which is an EOQ expression in n_s*. Its minimum over $n_s = 1, 2, 3, \dots$ is the optimal cost.

*第一项为到达成本，第二项为干线运输成本

- This least cost will always be greater or equal to $\sqrt{c'_d}$ because it is the minimum of $n_s + \frac{c'_d}{4n_s}$ with unrestricted n_s obtained for $n_s^* = (c'_d/4)^{1/2}$.
- Clearly, the underprediction will be most significant when n_s^* is close to an odd multiple of 0.5, or close to zero. We have tested the sensitivity of the EOQ cost expression to errors in the decision variables, which also quantifies this underprediction; as n_s^* increases the underprediction quickly vanishes
- Once $c'_d > 16$ (n_s^* is greater than 2) the difference is below 1%. If $c'_d > 4$ (the value at which $n_s^* = 1$) then the maximum difference stays below 6%. Although for smaller c'_d the difference can grow arbitrarily large as $c'_d \to 0$, that is not the case that is likely to be of interest; the large spacing between terminals recommended by the CA method (much larger than $d_{\max}$) indidates that operating line-haul vehicles is probably an overkill（多此一举）.
- If it were of interest, and a terminal had to be provided, terminal had to be provided, one could force the solution to the CA approach to satisfy the constraint $n_s > 1$. The next section will discuss how more involved constraints can be accommodated within a general CA framework.

- Although exhibiting different errors types, both examples shared a common trait when their errors were largest: the ideal terminal spacing in an interval with constant demand exceeded the length of the interval; i.e., demand varied significantly within the spacing.
- Errors arose because this property violates the stated requirement for the CA approach: $D'(d)$ should vary slowly over distances comparable with $s(d)$. Conversely, the numerical results prove that an error below one% results if $D(d)$ is piece-wise linear with segments at least three times as long as each $s(d)$.
- Thus, any demand function that can be approximated in this manner should also yield accurate results.

目 录 |Outline

- The CA method can be applied to more complex problems - even problems that defy exact numerical solution. In forthcoming chapters it will be used to locate points in multidimensional (time-space) domains while satisfying decision variable constraints.
- All that is needed is that the input data vary slowly with position, either in one or multiple dimensions, that the total cost can be expressed as a sum of costs over non-overlapping (small) regions of the location domain, and that these component costs (and constraints) depend only on the decisions made in their regions. If this is true, the decomposition principle holds and the CA results approximate the optimal cost accurately.

Constraints

As a one-dimensional illustration, let us return to the inventory control problem, and let us assume that there is a capacity constraint on shipment size:

$$D(t_i) - D(t_{i-1}) \leqslant v_{\max}$$

This constraint has a local nature because it only involves quantities determined by events close to the time of shipment; i.e., by two neighboring dispatching times and by the amount of consumption between them. For any time t it should be possible to write the constraint approximately as an inequality including only variables and data specific to time t.
该约束的物理意义是：两次配送时间的发生的需求量应该小于最大批量

EOQ formula

Recalling the definition of $H_s(t)$*, and using the slow-varying property of $D'(t)$, we can write:

$$D(t_i) - D(t_{i-1}) \approx H_s(t)D'(t) \approx H(t)D'(t)$$

and the constraint can be replaced by the approximation based only on conditions at t:

$$H(t)D'(t) \leqslant v_{\max}, \quad \text{or} \quad H(t) \leqslant v_{\max}/D'(t)$$

which must be satisfied for all t.

An approximate solution to our problem, thus, is an $H(t)$ that minimizes the total cost subject to this constraint. The optimal $H(t)$ is the least of: (i) $(2c_f/c_iD'(t))^{1/2}$, and (ii) $v_{\max}/D'(t)$. Letting Ψx denote the increasing concave function $\{x^{1/2}$ if $x < 1$; or $[1+x]/2$ if $x > 1\}$, we can express the minimum cost per unit time concisely in terms of the dimensionless quantity, $2c_fD'(t)/(c_iv_{\max}^2)$:

$$c_iv_{\max}\Psi\{2c_fD'(t)/(c_iv_{\max}^2)\}$$

* $H_s(t)$ as a step function such that $H_s(t) = t_i - t_{i-1}$ if $t \in I_i$

Total cost and average cost

Integrated from 0 to $t_{\max}$, this expression approximates the optimal total cost. Note that when the argument of Ψ is less than one, as would happen if $v_{\max}$ is very large, then the expression becomes $[2c_i c_f D'(t)]^{1/2}$. An average cost per item can also be obtained in a similar way; its interpretation as a cost average across items (calculated as if each item was part of a problem with constant conditions, equal to the local conditions for the item) is still valid. In practical cases, a per-item cost estimate can be obtained easily with the following two-step procedure:

- Solve the problem with constant conditions for a representative sample of items and input data,
- Average the solution across all the sampled items to obtain the result.

Note that the cost estimate can be obtained even without defining the decision variables in the first step.

Practical Considerations

- While for simple problems, such as the one solved above, the solution can be easily automated, more complex situations may benefit from decision support tools with substantial human intervention. The following two-step human/machine procedure is recommended:
  - recognizing that its recommendations may need fine-tuning adjustments, the CA (or other simplified) method is applied to a basic version of the problem without secondary details;
  - trained humans develop implementable solutions that account for the details, perhaps aided by numerical methods that can benefit from the output of the first step.
- In some cases, when time is of the essence humans alone may have to carry out this second step because efficient numerical methods capturing peculiar details may not be readily available, and developing them may be prohibitively time consuming.

- Even without time pressures, if the details are so complex (or so vaguely understood) that they cannot be quantified properly, pursuing automation for the fine-tuning step would seem ill-advised. Fortunately, this is not a serious drawback, significant departures from ideal situations should not increase cost significantly, leaving humans considerable latitude for accommodating details.
- The cost of the two-step procedure (fine-tuned by hand) is compared to the ideal cost without restrictions, and (optionally) to the exact optimal cost obtained with dynamic programming.
- We may find that the fine-tuning step often identifies the exact optimum, and when it does not, the difference between the two-step and the exact optimal costs is measured by a fraction of a%age point. Furthermore, the two-step and one-step (or ideal) costs are very close; of course, provided that n_s^* is not greater than 50.

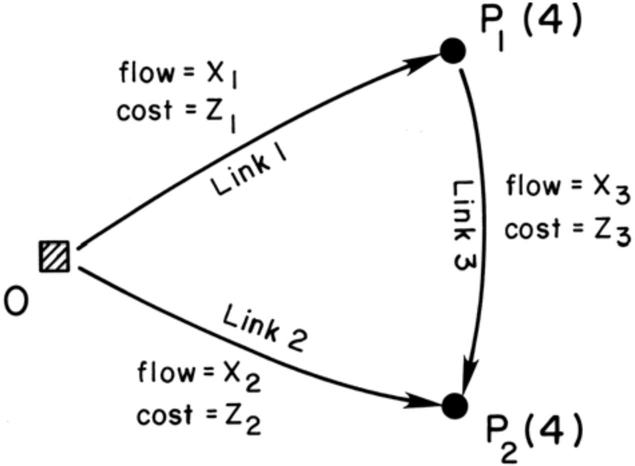
目 录 |Outline

Network Design Issues

In all the scenarios discussed so far, the items followed a predetermined path. Real logistics problems, however, often involve the choice of alternative routes (e.g., alternative ways of shipping) between origins and destinations, in addition to the choice of when and how much to dispatch. In some instances one may even be interested in whether certain routes should be provided at all; or even in the design of an entirely new physical distribution network.

We also found that there were economies of scale in flow; i.e., the optimal cost per item decreased with D'. In the following lectures, we will have to consider logistics problems with multiple destinations, where an item's route is not predetermined and cost decreases with flow. We discuss here some key features of these problems, and conclude this lecture with a comparison of detailed and non-detailed approaches for logistic system design.

目录 |Outline


P1 (4)
flow = X1
cost = Z1
Link 1
Link 3
flow = X3
cost = Z3
O
Link 2
flow = X2
cost = Z2
P2 (4)

- A simple example with one origin and two destinations effectively illustrates the properties of optimal system designs with and without flow economies of scale.
- The origin, O, produces items of type $i(i = 1, 2)$ for destination P_i at a constant rate, given by the parenthetical (括号的) numbers in the figure: $D'_1 = D'_2 = 4$ items per unit time. The combined production rate at the origin is $D'_1 + D'_2 = 8$ items/unit time. The arrows in the figure depict possible shipment trips; these transportation links are numbered 1, 2, 3. While all the items traveling to P_1, must travel directly between O and P_1, the items traveling to P_2 may go either directly or via P_1.
- Let us assume that a fraction (to be decided) x, of the items for P_2 are sent via P_1 and the rest are shipped directly. This establishes a flow $x_1 = 4(1 + x)$ on link 1 (OP_1), a flow $x_3 = 4x$ on link 3 (P_1P_2) and a flow $x_2 = 4(1 - x)$ on link 2 (OP_2).

We also assume that the total cost on the network can be expressed as a sum of link costs, and that these depend only on their own flows. This is a reasonable assumption if no attempt is made to coordinate the shipping schedules on the three links, as then the prorated cost to link should be close to the EOQ expression with demand rate equal to the link flow. Thus, if we let $z_i(x_i)$ denote the cost per item on link i when the flow is x_i, the total system cost per unit time is:
With economies of scale, the functions $x_i z_i(x_i)$ increase at a decreasing rate (are concave) as in

$$TC = \sum_{i=1}^{3} x_i z_i(x_i).$$

Economy of scale

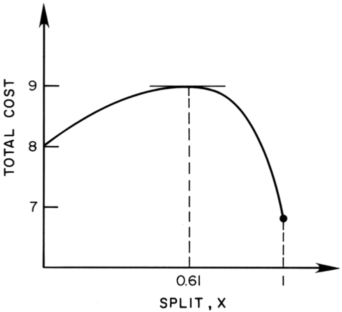
With economies of scale, the functions $x_i z_i(x_i)$ increase at a decreasing rate (are concave). Because the x_i's are linear in the split x, the total cost is a concave function of the split – this (concave) dependence of cost on splits (decision variables) also holds for general networks. Suppose, for example, that

$$z_1 = x_1^{-1/2}, z_2 = 3x_2^{-1/2}, \text{ and } z_3 = 1;$$
$$x_1 z_1 = x_1^{1/2}, x_2 z_2 = 3x_2^{1/2}, \text{ and } x_2 z_3 = x_3;$$

Then, as a function of x, the total cost is calculated as:

$$TC = 2(1+x)^{1/2} + 6(1-x)^{1/2} + 4x$$

The total cost is a concave function of the split, x. For our data the optimal solution is $x^* = 1 \rightarrow$ everything should be shipped through P_1. The total cost is 6.8. Although shipping everything direct may be better for different data, clearly one would never want to split the flow to P_2 among the two routes (OP_2 and OP_1P_2).



- A similar “all-or-nothing” principle holds for networks with multiple ODs if the total cost is a concave function of all the link flows. In that case all the flow from any origin to any destination should be allocated to only one route. This is not difficult to see: one can define a split between any two routes joining an origin and a destination, and since the link flows are linear in that split, the total cost is concave in the split; thus, only one of the routes can carry flow. Networks with diseconomies of scale behave in an opposite manner.
- In that case the total cost function is convex in the splits and there is an incentive to spread out the flow among routes. In fact, if for a one origin and one destination network, there exist several routes with identical cost functions (with dis-economies); it is not difficult to prove that the total flow should be evenly divided among all the routes.

- Networks with flow economies of scale also respond in a different manner to changes in conditions. While, with diseconomies, a small improvement to one of the routes would lead to a small change in the optimal flow distribution, with economies, the optimal flows either stay the same or change by a discrete amount. This can be seen with the example. As long as $z_3 < [2-2^{-1/2}] \approx 1.3$, x^* equals 1, but if z_3 is increased beyond this value ever so slightly, the solution jumps to $x* = 0$.
- This is typical of concave cost problems: minor changes to the input data can induce large changes in the optimal solution. Fortunately, the cost does not behave in such manner; despite the jump in our example the cost is a continuous function of z_3.

$$TC^* = \begin{cases} 8 & \text{if } z_3 \leqslant 2 - 2^{-1/2} \\ 8^{1/2} + 4z_3 & \text{if } z_3 > 2 - 2^{-1/2} \end{cases}$$

目 录 |Outline

The nature of the solution is not the only difference between networks with economies and diseconomies; the way to find it is also different. While networks with diseconomies are well behaved optimization problems without local minima that are not global, networks with economies are not.

Local searches

Except for technical details, all local search algorithms work in the same manner.

1. The total cost is evaluated for an initial feasible solution, described by a set of variables that uniquely identify the decisions; e.g., the set of splits for all origin destination pairs.
2. A small cost-reducing perturbation to the feasible solution (e.g., a differential change to the splits) is then sought. If not found, the search stops because the initial solution is a local minimum; i.e., a solution that cannot be improved without substantial changes. Otherwise, an improved larger perturbation obtained from the original small perturbations is identified, and then used to construct a new improved feasible solution.

The process is then iterated (seeking small cost-reducing perturbations to the new solution, etc.) until no significant improvements result.

Problems of local search algorithms

- Although local search algorithms can be used to find near optimal solutions for large detailed networks with convex costs, the same procedures fail with concave cost networks. The task is then much more complicated, and the network sizes that can be handled by numerical methods much smaller → 对于大网络的凸成本函数，局部搜索算法可以求得近似最优解；但是不适用于凹目标函数
- Local search techniques work acceptably for networks with scale diseconomies, because in those instances any local minimum is a global minimum. Unfortunately, this is not the case with economies of scale. Our simple problem has two local minima: $x = 0$ and $x = 1$. If a local search algorithm is applied to our example, any starting solution with $x < 0.61$ (the maximum in the figure) will converge suboptimally to $x = 0$.
- While for our simple example this can be corrected simply by starting with different x's, the task is daunting for large, highly detailed networks. In that case, the number of potential traps for a local search –all local minima regardless of cost –increases exponentially with the amount of detail.

An example

The items from a large number N of origins are shipped to one destination using two transportation modes (1 and 2). We use x_i to denote the split of production from origin i sent on mode 1, and assume that (to satisfy an agreement with the providers of type-1 transportation) each x_i must satisfy $x_i > h_i$ for some constant $h_i > 0$. Transportation by mode 2 is assumed to be more attractive, but limited in capacity; that is, the sum of the x_i's must exceed a value h.
For a set of splits to be feasible, thus, the following must be true:

$$\sum_{i=1}^{N} x_i \geqslant h, \text{ and } h_i \leqslant x_i \leqslant 1, \forall i$$

We seek the set of feasible splits that minimize the total cost, or equivalently the penalty paid because not all the items can be shipped by mode 2.

Solution of the example

The penalty paid for each origin is assumed to increase with x_i, except at certain values where a fixed amount ϵ_i is saved – perhaps because shipments can then be multiples of a box, requiring less handling*. To simplify the exposition, let us assume that there is only one such value δ_i for every origin, and that away from this value the penalty equals x_i; otherwise the penalty is $x_i - \epsilon_i$. If we define $\epsilon_i(x_i)$ to be: ϵ_i if $x_i = \delta_i$ and 0 otherwise, then the combined penalty for all the origins can be expressed as:

$$\sum_{i=1}^{N}[x_i - \epsilon_i(x_i)]$$

Note that each one of the terms in this summation for which $\delta_i > h_i$ exhibits two local minima in the range of feasibility $[h_i, 1] : x_i = h_i$ and $x_i = \delta_i$.

*假设起点处的惩罚费用随选择第一种方式的比例增大而增大，仅当该比例等于某个特定值时，可节省 ϵ_i 的费用。或因配送可由多个箱子完成，节省了处理费用。

Solution of the example

Any combination of x's, each equaling either h_i or δ_i, and satisfying the constraints is a local minimum, which could stop a search. If the δ_i and the h_i are uniformly distributed between 0 and 1, and h is small, there will be $O(2^{N/2})$ local optima. With so many traps, local search algorithms are doomed to failure for this problem –not because the penalty is discontinuous, but because it is not convex. A different method must be used.

Certainly, one could search exhaustively over all the possible solutions with a combinatorial tool such as branch and bound, but these methods can only handle problems of small size –typically with $O(10^2)$ decision variables or less.

Heuristic method

Alternatively, one could try to exploit the peculiar mathematical structure of the total penalty –or whichever problem is at hand –to develop a suitable algorithm. If successful, the approach would find a solution with all its detail. In our case, the optimization problem can be reduced to a knapsack problem that can be solved easily; in other instances it may be possible to decompose the problem into a collection of small easy problems. Very often, however, a simple solution method cannot be found.

In our case, this would happen if there were more than one (ϵ_i, δ_i) for each origin. Traditionally one then turns to *ad hoc*(为特定目的的) intuitive solution methods (known as heuristics) which one hopes will yield reasonable solutions.

Simplifying the problem

- There is also another approach. If while inspecting the formulation, or even better in the process of formulating the problem, one realizes that certain details are of little importance one should leave them out. Our example illustrates how removing minor details can turn a nightmare into an easy problem.
- If the ϵ_i's are so small that the $\epsilon_i(x_i)$ can be neglected, then the objective function reduces to $\sum_i x_i$. Former sources of difficulty, the ϵ_i and δ_i no longer enter the formulation. With less detail, the problem becomes well behaved (convex), and even admits a closed form solution; e.g., if $\sum_{i=1} h_i \geqslant h$ then the optimal splits are $x_i = h_i$ and the total cost is $\sum_{i=1} h_i = N\bar{h}$.
- Note that the optimal cost is given by an average (there is no need to know precisely each individual h_i in order to estimate the optimal cost), and that the optimal solution can be described with the simple rule "make every split as small as possible", which can be stated without making reference to the h_i's.

目录 |Outline

In the following lectures we will seek solutions to logistic problems using as little details as possible, describing (as in the example) the solution in terms of guidelines which are developed based on broad averages instead of detailed data. We recognize that the solutions obtained from such guidelines may benefit from fine-tuning once detailed data become available; but also note that incorporating all the details into the model early will increase the effort for gathering data and may even get in the way of obtaining a good solution.

Similarity of the logistics system with trees

Observation of mother nature's logistics networks suggests that many logistics systems can be designed in this manner. Trees can be viewed as a logistic system for carrying nutrients from the soil to an above-ground region (the leaves) to meet the sun's rays. While every individual tree of a species is distinct from other individuals, we also see that the members of a species share many common characteristics on average. There is order at the macroscopic level. This is not surprising, since members of the species have adapted to similar environmental conditions, also filling the same niche (发挥相同作用) in the eco-system.

Similarity of the logistics system with trees

The detailed characteristics of an individual tree are (like our logistic systems) developed from two levels of data in two different ways:

- Members of the same species share a genetic code, which has evolved in response to the typical or average conditions that can be expected. This code is analogous to the guidelines of a simple model; e.g., "make each split as small as possible."
- In response to the detailed conditions of its location, a tree develops an individuality within the guidelines of the genetic code, better to exploit the local conditions. This would be analogous to the fine tuning that could have taken place if the ϵ_i, h_i, and δ_i had been given in our example.

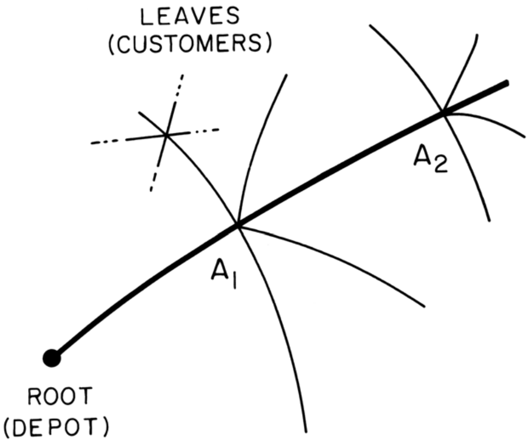
The same could be said for other logistic systems encountered in nature, such as the circulating and nervous systems of the human body.

Similarities

On further inspection we notice that, not only average characteristics, but some specific traits are also the same for all individuals, (e.g., some tree species have always one trunk, all humans have one aorta artery （主动脉）, etc.). It is as if nature had decided that these items of commonality are optimal for almost any conditions that can be encountered; therefore, that part of the design is not open to fine tuning. Perhaps the same can be said of logistics systems.

The logistics systems of nature also have economies of scale. It takes less energy to move a certain flow through one single pipe than through two pipes with one-half the cross section. As in our networks with concave costs, there is an incentive to consolidate flow into single routes that can handle great volumes efficiently. Nature has responded to this challenge by evolving hierarchical systems of conveyance, such as the three hierarchy network.

The ‘logistics system’ of a tree



Scientists have begun to realize that apparently very complex ("fractal") structures, such as a fern leaf (蕨类植物的叶子), can be replicated and/or described with just a few rules and parameters. For the example of the figure, the separations between "nodes" (e.g., A1 and A2) for each hierarchy might be found to be relatively constant, perhaps varying with the distance from the root, as might be the number of branches at every node and the relative size of the main and secondary branches at nodes of the same hierarchy. The latter may also vary with the distance from the "root."*

A physical distribution network should probably be organized in a similar way with the root becoming the depot, the leaves the customers, and the nodes intermediate transshipment centers or terminals. Physical distribution networks that serve similar purposes, just as in nature, should likely share the same hierarchical organization and overall traits even if the specific details differ. As in nature, it should be possible to describe their near optimal configuration with just a few simple rules and parameters.

*两个节点之间的间隔也许恒定或者随着到根部的距离变化。每个节点处的分支数及节点内部相同层次的主次分支的相对大小也呈现相似规律。后者也会随着距离根部的距离而变化。

Any questions?

Readings

- Daganzo. Logistics System Analysis. Ch.2.