Constructal theory of economics

Article *in* Applied Energy · September 2000

DOI: 10.1016/S0306-2619(00)00023-4

Adrian Bejan
Duke University

ELSEVIER

Applied Energy 67 (2000) 37–60

APPLIED ENERGY

www.elsevier.com/locate/apenergy

Constructal theory of economics☆

A. Bejan [a,*], V. Badescu [b], A. De Vos [c]

[a]Department of Mechanical Engineering and Materials Science, Box 90300, Duke University, Durham, NC 27708-0300, USA

[b]Department of Mechanical Engineering, Thermodynamics, Polytechnic University of Bucharest, Spl. Independentei 313, Bucharest 77206, Romania

[c]Department of Electronics, University of Gent, Sint-Pietersnieuwstraat 41, Gent, B-9000, Belgium

Abstract

This paper extends to economics of the constructal theory of generation of shape and structure in natural flow systems that connect one point to a finite size area or volume. By invoking the principle of cost minimization in the transport of goods between a point and an area, it is possible to anticipate the dendritic pattern of transport routes that cover the area, and the shapes and numbers of the interstitial areas of the dendrite. It is also shown that by maximizing the revenue in transactions between a point and an area, it is possible to derive not only the dendritic pattern of routes and their interstices, but also the optimal size of the smallest (elemental) interstitial area. Every geometric detail of the dendritic structures is the result of a single (deterministic) generating principle. The refining of the performance of a rough design (e.g. rectangles-in-a-rectangle) pushes the design towards a structure that resembles a theoretically fractal structure (triangle-in-triangle). The concluding section shows that the law of optimal refraction of transport routes is a manifestation of the same principle and can be used to optimize further the dendritic patterns. The chief conclusion is that the constructal law of physics has a powerful and established analogue in economics: the law of parsimony. The constructal theory, as extended in this paper, unites the naturally-organized flow structures that occur spontaneously over a vast territory, from geophysics to biology and economics. © 2000 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Constructal theory; Spatial economics; Tree networks; *p*-Median problem

* Corresponding author. Tel.: +1-919-660-5310; fax: +1-919-660-8963.

E-mail address: abejan@duke.edu (A. Bejan).

☆This paper is based on a version previously published in Energy Convertion and Management Vol. 41(13) pp 1429–1451. Authors wishing to cite this article should cite the original source.

PII: S0306-2619(00)00006-4

Nomenclature

A_i	Area (m^2).
B_i	Base (m).
c_i	Unit cost (\$/unit).
C_i	Cost (\$/s).
d_i	Distance (m).
f_i	Goods flow rate (units/s).
g	Price paid by consumer (\$/unit).
H_i	Transversal dimension (m).
k_i	Interaction parameter (\$/s·m).
K_i	Cost (\$/m·unit).
L_i	Length (m).
m_i	Goods flow rate (units/s).
n_i	Number of area constituents in a construct.
R_i	Revenue (\$/s).
x, y	Cartesian coordinates (m).

Greek

γ	Rate of production or consumption per unit area (units/s·m).

Subscripts

c	Critical.
i	Order of construct.
max	Maximum
min	Minimum
opt	Optimum

1. Introduction to constructal theory

The analytical structure of equilibrium (classical) thermodynamics has been used in the past to construct analogies and models in economic theory. The objective of this paper is to bring to the attention of economists an important transformation that has occurred in thermodynamics in the wake of the energy crisis. These developments can be used not only as new leads toward predictive models in economics, but also as a more general theory that unites economics ideas with laws that are known to govern other physical and biological processes.

The thermodynamics transformation of the past two decades is represented by the development of methods of modeling and optimization of flow systems, that is, systems that are in states of internal non-equilibrium. Internal gradients of pressure and temperature drive internal currents, and consequently and necessarily, each system is characterized by irreversibility and spatial structure (i.e. shapes, channels, architecture). The models of such systems combine from the start the laws of thermodynamics with the laws that govern irreversible (rate, transport) processes such as fluid flow, heat transfer and mass transfer. The models also account for the global constraints that must be faced by the system, for example, finite sizes, properties of materials and finite times of operation. This approach to modeling allows us to determine our objective (e.g. power output from a power plant) as a function of the physical characteristics and global constraints of the system. Such models are useful only for orientation; in real life, the design of a physical installation (e.g. a power plant) is based on cost minimization subject to constraints, according to the well established methods of thermo-economics (e.g. Ref. [1], chapters 7–9). Thermo-economics accounts not only for the costs of transportation recognized in this paper but also for the capital costs associated with the means of transport.

One of the new methods of thermodynamic modeling and optimization has already been brought to the attention of the economics community: endoreversible thermo-economics [2–4]. In this paper, we draw attention to a recent development known as constructal theory [5]. It was found that (i) when a system houses a flow that proceeds from one point on the boundary to every point inside the system's volume, or vice versa, and (ii) the system is heterogeneous in that the flow may exist in at least two dissimilar regimes (high and low resistances), then the optimal internal architecture consists of low resistance channels that form a tree network and high resistance interstices that cover every point of the given volume. This approach is completely deterministic: every geometric detail is a result of minimizing (subject to global constraints) the overall resistance to the point-to-volume or volume-to-point flow.

The main contribution of constructal theory is that it identifies a single principle — the constrained minimization of global resistance — as the generating mechanism of shape and structure in a vast number of natural-flow systems [6]. Examples are all the dendritic systems (bronchial trees, vascularized tissues, river drainage basins and deltas, lightning, neural dendrites, dendritic crystals), all the natural ducts with nearly-round cross-sections (bronchial passages, blood vessels, underground rivers) and all the open channels (rivers) that display a universal scaling between width and depth. Constructal theory was also stated in terms of minimizing the time of travel between a finite-size area and one point [6], which led to the prediction of street patterns and urban growth.

In this paper, constructal theory is extended into the realm of economics. The economic activity is assessed via the flow-access optimization principle, as identified in Refs. [5,6], and the economic and business structure is the result. To see how constructal theory accounts for the origin of structure in economics and business, we will consider a stream of goods that proceeds from one point (the producer or distributor) to every point of a finite-size territory (i.e. the consumers). The flow may also proceed in the opposite direction, from a finite area (e.g. grain, or carpets woven by individuals) to one point of collection.

The fundamental contribution of this extension is that the "law of parsimony" emerges as the economics analogue of the resistance minimization principle recognized in physics and engineering, and that constructal theory emerges as an even more general theory of parsimony in nature.

2. Shape and structure to achieve the minimization of cost

Consider a stream of goods that proceeds from one point (the producer or distributor) to every point of a finite size territory (containing the consumers). The objective is to minimize the total cost associated with the given point-to-area or area-to-point stream.

The economics principle of economies of scale tells us that the cost is lower when the goods move in the aggregate; that is, when they are organized into thicker streams. The cost is also proportional to the distance traveled. Clearly, the cost plays the same role as the local thermal resistance in heat trees, the inverse of the travel speed in street trees, or the local fluid flow resistance in fluid trees. The given business territory is covered naturally by trees; that is, links of decreasing cost, starting from the highest unit cost that is allocated to the smallest area scale (the individual) and continuing in a sequence of intermediaries (distributors and collectors) who handle increasingly larger fractions of the total stream of goods.

We illustrate this deterministic process (see Fig. 1) by considering the area A in which goods must arrive at (or depart from) every point at the uniform rate γ [units/(m^2 s)]. The total stream of goods $m = \gamma A$ flows between A and a single point M (the producer or collection point). For a given total stream m, it is advantageous to have higher rates near the collection point M and lower rates in distant areas dA. Such a non-uniform source of goods was studied in the framework of solar cells [7], where the 'good' to be collected is the photocurrent. However, in many cases, a uniform production rate density $\gamma = m/A$ is given, and this distribution cannot be changed. Therefore, we restrict ourselves to this case.

Several means of transport are available, and they are represented by the sequence of costs K_i [\$/(m·unit), $i = 0, 1, 2 \ldots$, such that $K_0 > K_1 > K_2 \ldots$. Each K_i represents

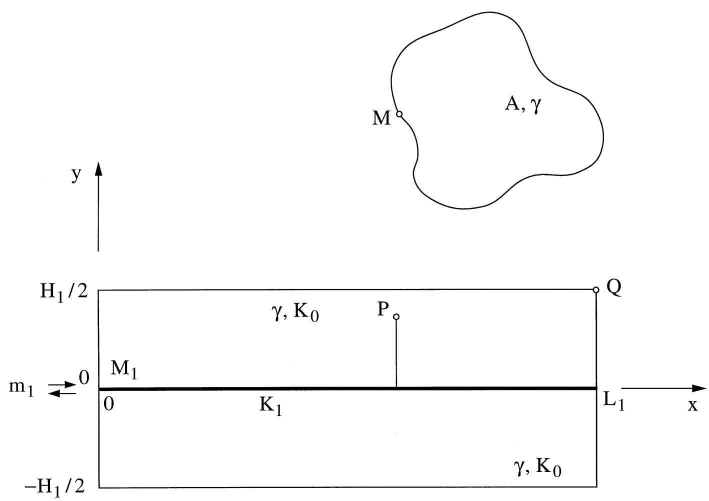

Fig. 1. The area-to-point transport configuration (top), and the elemental area A_1 (bottom); the double arrow means that the flow of goods may proceed in either direction.

the cost per unit of goods transported and per unit of distance traveled. The problem consists of covering A with trajectories of various unit costs such that the total cost required by the m stream is a minimum. One way to approach this problem is by allocating an area element to each K_i link and connecting the links such that the entire area A is connected to M.

Let us assume that the smallest area element — the elemental area — over which the stream is delivered to (or collected from) every point is A_1 (see Fig. 1). In this section, the size of A_1 is assumed known and fixed. The area A_1 is "elemental" because the stream associated with it ($m_1 = \gamma A_1$) is channeled along a single link (K_1). Every point (P) of the elemental area has access to the K_1 link along a direct (perpendicular) path of the highest unit cost, K_0. In other words, goods travel between A_1 and M_1 by two means: first, the most expensive mode of transport touches every point of the infinity of points found in A_1, and secondly, the less-expensive mode of transport (K_1) makes the final connection with the exit point M_1.

The minimization of cost suggests that we shape A_1 as a rectangle (H_1L_1) that "surrounds" the K_1 link, such that K_1 is placed on the longer of the two axes of the rectangle. The shape parameter H_1/L_1 is free to vary. The per-unit-time cost associated with the transport of one unit from P to M_1 is $K_0y + K_1x$. The cost of transporting the total stream between A_1 and M_1 is

$$C_1 = 2\gamma \int_0^{H_1/2} \int_0^{L_1} (K_0 y + K_1 x)\mathrm{d}x\mathrm{d}y = \gamma\left(\frac{1}{4}K_0 H_1^2 L_1 + \frac{1}{2}K_1 H_1 L_1^2\right) \tag{1}$$

Because the elemental area and its total stream are fixed, we may express the cost C_1 as

$$\frac{C_1}{m_1} = \frac{K_0 A_1}{4L_1} + \frac{1}{2}K_1 L_1 \tag{2}$$

We reach the important conclusion that the shape (of optimal form) can be deduced from the cost minimization. The optimal shape of the elemental area is independent of its size,

$$\left(\frac{H_1}{L_1}\right)_{\mathrm{opt}} = 2\frac{K_1}{K_0} < 1 \tag{3}$$

Associated features of this geometric optimum are $H_{1,\mathrm{opt}} = (2K_1A_1/K_0)^{1/2}$, $L_{1,\mathrm{opt}} = (K_0A_1/2K_1)^{1/2}$ and the total minimum cost associated with the transport of the m_1 stream:

$$C_{1,\min} = m_1(A_1K_0K_1/2)^{1/2} \tag{4}$$

Two observations define the course of the geometric optimization that will follow. First, instead of minimizing the cost integrated over A_1 [Eq. (1)], we could have minimized the unit cost faced by the most distant producer or consumer (point Q in

Fig. 1). That cost is $K_0H_1/2 + K_1L_1$. Having been minimized subject to the constant elemental area $A_1 = H_1L_1$, this cost leads to exactly the same geometric optimum as in Eq. (3). When the elemental shape is optimised, the cost from Q to the bend in the path ($x = L_1$) equals the cost from the bend to M_1. Note that the unit cost faced by individual points (P) situated closer than Q relative to M_1 is always smaller than the unit cost from Q to M_1. This means that what is good for the individual, especially for the individual with the worst geographic location, is good for the entire group of producers or consumers that inhabits the elemental area. This sheds light on why the spatial organization of the flow of goods (the first route K_1) can occur spontaneously, that is without a discussion, plan and agreement for the entire group. In the following analysis, we will continue with calculations of total costs integrated over finite areas, as in Eq. (1).

The second observation is based on the minimum cost determined via Eq. (4). What happens when the goods must flow between the point M_1 and an increasingly larger territory A_1? Both the flow rate m_1 and the cost $C_{1,\min}$ increase, but the rate of increase in the total cost ($C_{1,\min}/m_1$) increases as well. It increases as $A_1^{1/2}$. If global cost minimization is the driving force behind the generation of spatial economics structure, then the population of producers or consumers (γ) will search for a geometric way of slowing down the cost increase that comes with territorial expansion.

The population is already organized into optimally-shaped area elements of type A_1. The alternative to covering a larger territory (i.e. the alternative to increasing A_1) is to assemble a number of A_1 elements into a larger construct A_2, Fig. 2. The number of constituents $n_2 = A_2/A_1$, or the external shape H_2/L_2, is the unknown. The total stream of goods flowing from A_2 to the new single point M_2 (the larger producer, collection point or user of goods) is $m_2 = \gamma A_2 = n_2m_1$. The K_1 links of the A_1 elements are connected to M_2 through a new link of unit cost K_2. Because the new link (perhaps, an advanced mode of transportation) facilitates the transport of

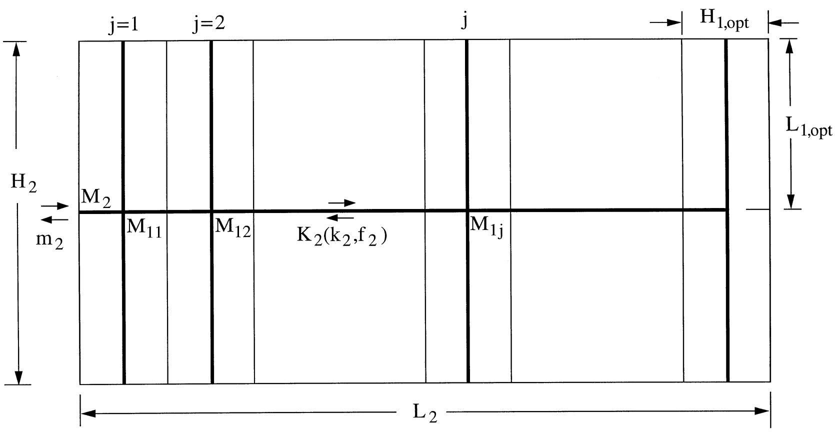

Fig. 2. The second area A_2, as a construct of n_2 elemental areas; the double arrow means that the flow of goods may proceed in either direction.

a stream that is larger than in each elemental area, its unit cost is lower than in the elements, provided $K_2 < K_1$.

The total cost (C_2) required for the transport of m_2 between A_2 and M_2 is obtained by adding the $C_{1,\min}$ shares contributed by the n_2 constituents to the cost of transport along the central K_2 link. The latter is less than $m_2\ K_2\ L_2$ because the actual flow rate via this route is less than m_2: the flow rate varies from the full value m_2 at the root point (M_2), all the way down to zero at the opposite end of the K_2 link. This variation is approximately linear when n_2 is sufficiently large. In this limit, the total cost contributed by the K_2 link is $\frac{1}{2}m_2K_2L_2$ because the average flow rate along the K_2 route is $m_2/2$.

So the cost associated with the area-to-point configuration $A_2 \rightarrow M_2$ is

$$C_2 = n_2C_{1,\min} + \frac{1}{2}m_2K_2L_2 \tag{5}$$

This result can be rearranged using Eq. (4), $A_1 = A_2/n_2$ and $L_2 = (n_2/2)H_1$ to reveal

$$\frac{C_2}{m_2} = \left(\frac{K_0K_1A_2}{2n_2}\right)^{1/2} + \frac{K_2}{2}\left(\frac{K_1A_2n_2}{2K_0}\right)^{1/2} \tag{6}$$

An optimal elemental area size ($A_{1,\mathrm{opt}} = A_2/n_{2,\mathrm{opt}}$), such that the total cost C_2 is a minimum, exists, i.e.:

$$n_{2,\mathrm{opt}} = 2\frac{K_0}{K_2} > 1 \tag{7}$$

$$C_{2,\min} = m_2(K_1K_2A_2)^{1/2} \tag{8}$$

The optimal external shape is described by $(H_2/L_2)_{\mathrm{opt}} = K_2/K_1$, $H_{2,\mathrm{opt}} = (A_2K_2/K_2)^{1/2}$ and $L_{2,\mathrm{opt}} = (K_1A_2/K_2)^{1/2}$. What we have optimized is the shape of A_2, not the size of A_2. The optimal shape is fixed when K_2 and K_1 are specified. The A_2 area is not known because it is proportional to the elemental area A_1, which is not specified. Only when we maximize the revenue derived from the flow of goods (Section 4) will we optimize the shape and size of each area construct.

It is useful to compare Eq. (8) with Eq. (4) to see that in the second area structure (Fig. 2), the ratio $C_{2,\min}/m_2$ continues to be proportional to the total area (A_2), but that the rate of increase has been reduced. The factor $(K_0K_1/2)^{1/2}$ of Eq. (4) has been replaced by $(K_1K_2)^{1/2}$ in Eq. (8): the latter is smaller when $K_2 < K_0/2$, which is true in the present case because of Eq. (7).

The more basic message of the geometric optimization of the transport configurations of Figs. 1 and 2 is that cost minimization can be achieved (i) by optimizing the area and the structure, and (ii) by increasing the internal complexity of the structure. This method of cost minimization can be pursued further, towards more complex internal-structures, as long as new and less costly modes of transport become available. For example, if the new mode has the unit cost K_3 that is sensibly smaller than K_2, then it makes sense to cover a larger territory, not by increasing A_2

Table 1
The optimized structure of point-to-area or area-to-point transport for minimum cost

I	$n_{i,\text{opt}}$	$(H_i/L_i)_{\text{opt}}$	$C_{i,\min}/m_i$
1	—	$2K_1/K_0$	$(A_1K_0K_1/2)^{1/2}$
2	$2K_0/K_2$	K_2/K_1	$(A_2K_1K_2)^{1/2}$
$\geqslant 3$	$4K_{i-2}/K_i$	K_i/K_{i-1}	$(A_iK_{i-1}K_i)^{1/2}$

indefinitely using the design of Fig. 2 and Eqs. (3) and (7), but by combining n_3 constructs of type A_2 into the new construct A_3 shown in Fig. 3. The n_3 streams of size m_2 are delivered or collected by a central route of length L_3 and unit price K_3. The total stream that reaches every point of the area A_3 is $m_3 = n_3m_2$. When n_3 is sufficiently greater than unity, the total cost associated with the transport between A_3 and the root point M_3 is

$$C_3 = n_3C_{2,\min} + \frac{1}{2}m_3K_3L_3 \tag{9}$$

The optimal values of n_3 and H_3/L_3 that minimize the ratio C_2/m_3 are listed in Table 1. This optimization can be continued toward constructs of higher order: when $i \geqslant 3$, the results fall into the pattern summarized by the recurrence relations listed in Table 1.

3. Alternate methods

The preceding analysis was based on two simplifying assumptions, adopted from the earlier work on heat-and-fluid tree networks [5,6]. These assumptions allowed us to show most clearly how the flow structure is deduced consistently from a single principle. One assumption is that the optimization results (the architecture) determined at one level of assembly are preserved for reuse as internal features of the

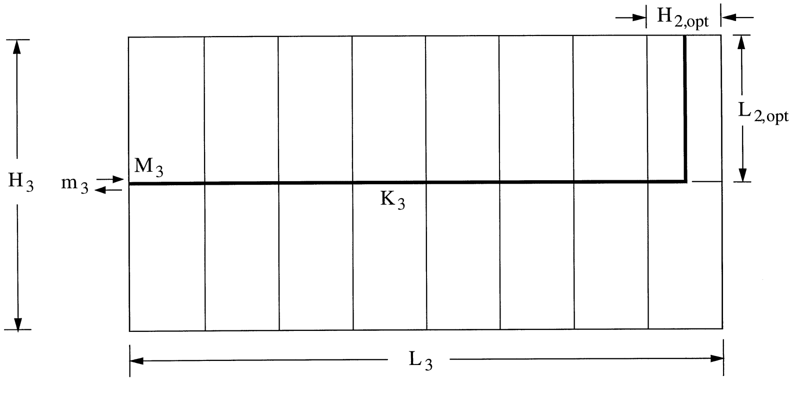

Fig. 3. The third area A_3, as a construct of n_3 areas of type A_2; the double arrow means that the flow of goods may proceed in either direction.

next, larger area construct. The other assumption is that smaller areas assembled into a larger area communicate with each other only through the root points M_i, while the rest of their rectangular perimeters are impermeable. Additional, less pivotal simplifying assumptions, such as the rectangular shape of area elements and the angles between successive transport paths, are discussed in Sections 6 and 7.

These assumptions simplified the analysis because in each new construct they kept the number of degrees-of-freedom to a minimum. Their impact on the goodness of the optimized architecture was evaluated systematically in several studies dealing with tree-shaped heat flows [8–10]. The evaluation consisted of repeating the same work without making these assumptions, by coping with the large number of degrees of freedom and solving the problem numerically in a continuum. A similar numerical approach is being used in the simulation and optimization of river basins and vascularized tissues [11,12]. The numerical work on heat trees showed that the questioned assumptions have a relatively minor impact on the optimized shape and structure, and practically no effect on the optimized performance (the minimized global resistance). The work on river basins and vascularization [11,12] also showed in fully-numerical simulations, that begin with postulating an initial, non-optimal tree network, the final architecture is also influenced by the initial features of the postulated network. If the initial network is chosen randomly, the final optimized architecture is random, i.e., not reproducible.

In summary, the constructal method provides a most transparent route — a short-cut — to the optimal performance, i.e. the optimal access (for minimum resistance, time or cost) between an area and one point. The constructal tree is not the mathematically optimal design: that structure can only be determined numerically based on the most refined formulation, involving the largest possible number of degrees of freedom. The constructal solution is the simplest structure that (i) is a tree, (ii) is deterministic, i.e. is deduced by invoking a single principle and (iii) performs practically as well as the much more refined (and more "natural" looking) tree structures optimized using black-box numerical simulations.

This is also a good place to begin the discussion of how constructal theory fits in the research that is being conducted today in the field of spatial economics. This field is highly developed because of the computer age and the limitless applications of the access optimization problem. Despite such a busy and important background, it may seem that the ideas of this paper were developed in complete isolation from what goes on in the mainstream. This is true in the real sense: the ideas originate from a completely separate field (thermodynamics). Our work is to communicate these ideas to the spatial economics community and to explain their purely theoretical origin. They were not derived from, or built on, developments that any of us might have found by reading the spatial economics literature.

For example, today, considerable work is being devoted in spatial economics to the so-called *p*-median problem. Some of the most representative and recent references in this area are Refs. [13–19]. According to Church and Sorensen's recent review [20], "the *p*-median problem involves the location of a fixed number of facilities in such a manner that the total weighted distance of all users assigned to their closest facility is minimized. The *p*-median problem was first posed by Hakimi [21] on a network of nodes and arcs. Each node was considered a place of demand as well as a potential facility location. The arcs represented transportation or accessibility linkages, and could be used for facility locations as well".

The present ideas differ fundamentally from the p-median framework. First, we do not assume (postulate) a network: our objective is to deduce the structure, i.e. to predict the very existence of a "structure" on the basis of a single principle. Secondly, we do not simulate a territory by using a finite number of points (nodes): we optimize the access to a finite-size area, i.e. to an infinity of points. Thirdly, we do not assume "links" between nodes, so that we may neglect the white area that separates two neighboring nodes. In constructal theory, the transport over a white area (K_0) is exactly as important as the transport along the route (K_1) that belongs to the area. Fourthly, there is only one "facility" in the constructal problem: the point that serves as a sink or source for the flow that reaches completely the given area.

In agreement with other area-point or volume-point constructal examples in physics and engineering [5,6], this paper leads us theoretically to tree shaped paths as optimal routes for transport. The tree image does not have loops and seems to contradict the "grids" — the networks with loops — that characterize urban areas, commerce and communications. However, there is no contradiction because the constructal tree is *the structure of the flow*: the path, or paths followed by the area-point *stream*. The constructal tree is not a collection of immobile pieces of hardware (ducts, cables, or paved roads): it is the flow that happens to inhabit the necessary links when the area and point are specified. Take any urban grid, pick a single source or sink point and try to reach the infinity of points of the given area, and then you will visualize mentally the tree shaped stream that flows through some portions of the grid.

Grids of streets develop because, in real life, the many inhabitants of an area must have access to several facilities, not one [6]. Each facility point (source, sink) is the root of its own tree. When all these trees are superimposed on the same area, they form a complicated net (with loops), which, incidentally, serves as the empirical starting point in the p-median model. Constructal theory starts from the small and simple, not from the large and complicated. It is an atomistic theory about the origin of the first building block. The elemental flow structure (area-point tree) may be used to reconstruct the origin of more complicated patterns (e.g. grids), in the same way that atoms are used to explain molecules, crystals and planets.

4. Shape and structure as deduced by the maximization of revenue

In this section, we solve a different problem of geometric optimization of point-to-area or area-to-point transport. We show that if, instead of minimizing cost, the objective is to maximize revenue, it is possible to predict not only the tree-shaped network of the transport routes, but also the size of the elemental area A_1.

Two features distinguish the new approach from the model used earlier. First, the unit cost associated with a certain link, c_i (\$/unit), is assumed to be proportional to the distance traveled, d_i (m), and inversely proportional to the flow rate of goods along that link, f_i (units/s),

$$c_i = k_i \frac{d_i}{f_i} \tag{10}$$

where k_i (\$/(s m)) is a parameter that describes the interaction — the match — between producer and consumer. The ratio k_i/f_i is equivalent to the cost parameter K_i used in the preceding section. The advantage of using k_i/f_i is one of increased flexibility: on the same route and at the same flow rate f_i, two different producers may face different costs (i.e. different k_i values). The cost analysis for the elemental area A_1 (Fig. 1) proceeds according to the steps outlined in Eqs. (1) and (2). In the first integrand of Eq. (1), we use $(k_0/f_0)y + (k_1/f_1)x$ in place of $K_{0y} + K_{1x}$ and arrive at the elemental cost formula

$$C_1 = \frac{\gamma}{2} A_1 \left(\frac{k_0 H_1}{2 f_0} + \frac{k_1 L_1}{f_1} \right) \tag{11}$$

The second new feature is the revenue obtained by the producer who distributes and sells a stream of goods over an area. If g (\$/unit) is the price paid by the individual consumer who resides at a point (x, y) of the elemental area A_1, then the revenue cannot exceed $m_1 g$. From this ceiling value, we subtract the transportation cost C_1, and obtain the net revenue

$$R_1 = m_1 \left(g \frac{m_1 k_0}{4 \gamma L_1 f_0} \frac{L_1 k_1}{2 f_1} \right) \tag{12}$$

This expression shows that the net revenue vanishes in two extremes, i.e. when the flow of goods stops ($m_1 = 0$) and when the flow rate reaches the critical value

$$m_{1,\mathrm{c}} = 4 \gamma L_1 \frac{f_0}{k_0} \left(g - \frac{L_1 k_1}{2 f_1} \right) \tag{13}$$

This means that R_1 has a maximum with respect to m_1. The net revenue has an additional maximum with respect to L_1, because R_1 tends to minus infinity for both limits, $L_1 \to 0$ and $L_1 \to \infty$. The maximum net revenue is obtained by solving simultaneously $\partial R_1/\partial m_1 = 0$ and $\partial R_1/\partial L_1 = 0$, and the results are

$$m_{1,\mathrm{opt}} = \frac{8}{9} \gamma g^2 \frac{f_0 f_1}{k_0 k_1} \tag{14}$$

$$L_{1,\mathrm{opt}} = \frac{2}{3} g \frac{f_1}{k_1} \tag{15}$$

Because $m_1 = \gamma A_1$, Eq. (14) prescribes the optimal size of the elemental territory that should be serviced by the producer,

$$A_{1,\mathrm{opt}} = \frac{8}{9} g^2 \frac{f_0 f_1}{k_0 k_1} \tag{16}$$

The optimal shape of the elemental area is obtained by combining $H_1/L_1 = A_1/L_1^2$ with Eqs. (15) and (16): the resulting expression is equivalent to Eq. (3), namely $(H_1/L_1)_{\rm opt} = 2(k_1/f_1)/(k_0/f_0)$. The twice maximized net revenue that corresponds to the two optimal parameters determined in Eqs. (14,15) is

$$R_{1,\max,\max} = \frac{8}{27}\gamma g^3 \frac{f_0 f_1}{k_0 k_1} \tag{17}$$

To summarize, the geometric maximization of net revenue at the elemental level has two degrees of freedom, the size and shape of A_1. New, relative to the results of the cost minimization (Section 2), is the optimal size of the elemental area. According to Eq. (16), the elemental area is large when the product is expensive (i.e. large g) and the transport is inexpensive (small k_0/k_0 and k_1/f_1). This optimal area size is independent of the surface density of the flow of goods γ because it is essentially a balance between the revenue generated by the stream of goods and the cost of transporting the same stream. The twice-maximized revenue per unit area is equal to $\gamma g/3$.

Consider next the second-order area element A_2 of Fig. 2, where $A_2 = n_2 A_1$. The root (storage) points of the elemental areas $\left(M_{1j},\ j = 1, 2, \ldots, n_2/2\right)$ are connected to the global root point M_2 by a new route of cost factor k_2 and flow rate f_2. The conservation of goods requires $m_2 = n_2 m_1$. Each elemental area contributes to the net revenue maximized in accordance with Eq. (17). The question is how much of the total revenue $n_2 R_{1,\max,\max}$ is used to offset the cost associated with the transport along the second-order route.

Along the short stretch between points M_2 and M_{11}, the flow rate is equal to the total flow rate for the second-order area, m_2. The unit cost is $(k_2/f_2)H_{1,\rm opt}/2$, and this means that the cost associated with the segment $M_2 M_{11}$ is

$$C_{2,1} = m_2 \frac{k_2}{f_2} \frac{H_{1,\rm opt}}{2} \tag{18}$$

Along the next segment ($M_{11}M_{12}$), the flow rate is $m_2 - 2m_{1,\rm opt}$, and the distance traveled is $H_{1,\rm opt}$. Together, these quantities pinpoint the total cost contributed by the segment,

$$C_{2,2} = (m_2 2m_{1,\rm opt}) \frac{k_2}{f_2} H_{1,\rm opt} \tag{19}$$

The remaining segments of the k_2/f_2 path are analyzed similarly, because the length of each is equal to $H_{1,\rm opt}$. For example, the cost associated with the segment $M_{1,j-1}M_{1j}$ is

$$C_{2,j} = [m_2 - 2(j-1)m_{1,\rm opt}] \frac{k_2}{f_2} H_{1,\rm opt} \tag{20}$$

In conclusion, the total cost due to transport along the central link that leads to M_2 is

$$C_2 = \sum_{j=1}^{n_2/2} C_{2,j} \tag{21}$$

The net revenue collected over A_2 is equal to $R_2 = n_2 R_{1,\max,\max} - C_2$, which becomes

$$R_2 = n_2 m_{1,\mathrm{opt}} \left(g - \frac{n_2 k_2}{4 f_2} H_{1,\mathrm{opt}} \right) \tag{22}$$

This quantity has a maximum with respect to n_2, because it vanishes at $n_2 = 0$ and at $n_2 = 4 f_2 g/(k_2 H_{1,\mathrm{opt}})$. The optimal number of elements in the second-order area is obtained by solving $\partial R_2/\partial n_2 = 0$, i.e.

$$n_{2,\mathrm{opt}} = \frac{3k_0/f_0}{2k_2/f_2} \tag{23}$$

This solution is valid if $n_{2,\mathrm{opt}} \geqslant 2$. In Eq. (23), we reach the important conclusion that optimal spatial growth (organization or assembly) can be deduced on the basis of revenue maximization. A similar conclusion was reached based on cost minimization in Eq. (7). It is worth noting that Eqs. (23) and (7) do not prescribe the same number of elements in the A_2 assembly: if we recall the change in parameters made at the start of this section (namely, $K_0 = k_0/f_0$, $K_2 = k_2/f_2$), we see that revenue maximization recommends a smaller number of elements than cost minimization.

The remaining parameters of the optimized second-order area can be determined from $m_{2,\mathrm{opt}} = n_{2,\mathrm{opt}} m_{1,\mathrm{opt}}$ and the geometric relations $H_{2,\mathrm{opt}} = 2L_{1,\mathrm{opt}}$ and $L_{2,\mathrm{opt}} = (n_{2,\mathrm{opt}}/2) H_{1,\mathrm{opt}}$. The resulting expressions are $H_{2,\mathrm{opt}} = (4/3) g f_1/k_1$ and $L_{2,\mathrm{opt}} = g f_2/k_2$, as shown in Table 2. At this optimum, the net revenue expression (22) becomes

$$R_{2,\max} = \frac{2}{3} \gamma g^3 \frac{f_1 f_2}{k_1 k_2} \tag{24}$$

which reminds us that, at the second-level of assembly (A_2), the net revenue was maximized with respect to a single variable. The corresponding area of the assembly is

$$A_{2,\mathrm{opt}} = \frac{4}{3} g^2 \frac{f_1 f_2}{k_1 k_2} \tag{25}$$

which means that the revenue per unit area at this level is $R_{2,\max}/A_{2,\mathrm{opt}} = \gamma g/2$. This ratio is higher than the corresponding ratio at the elemental-area level ($\gamma g/3$). The stepwise increase in revenue per unit area is an additional incentive to continue the geometric optimization sequence towards larger constructs. The other incentive is the increase in revenue, which proceeds in the direction of a larger area formed by connecting the streams of optimized smaller areas. We will examine this trend in the next section and Fig. 4.

The analytical results are listed in Table 2 and are valid provided $n_i \geqslant 2$. The results settle into a pattern when $i \geqslant 4$: this feature differs somewhat from the establishment of a repetitive pattern based on cost minimization, which occurred when $i \geqslant 3$ (Table 1). In the present solution, the routes (k_i, f_i) form a tree network that is completely deterministic. Every geometric detail is the result of invoking a single principle — the maximization of revenue.

If the recurrence formulae listed for $i \geqslant 4$ in Table 2 were to be repeated ad infinitum in the opposite direction, until the area scale of size zero is reached, then the resulting tree structure would be a fractal. The present structure is not a fractal — it is an Euclidean figure — because the number of levels of assembly is finite [6]. Access to the infinity of points of the given area is made via the costliest transport mode (k_0, f_0), which is placed over all the scales that are smaller than the elemental area $A_{1,\mathrm{opt}}$. Unlike in fractal tree networks, where the smallest scale (the inner cut-off) is arbitrary and chosen for graphic impact, in the present construction, $A_{1,\mathrm{opt}}$ is finite and predictable, based on the same principle of revenue maximization that generates the rest of the network structure.

5. Development of the economics structure in time

We should not read too much into the suggestion that the sequence outlined in Table 2 continues indefinitely. The passing of time, or better, the development of transport technology in time, dictates how far the structure spreads in two dimensions. In Table 2, the index i could be associated with the increase in time. For example, if the modes of transport (in order of decreasing cost per unit) are hand delivery (k_0, f_0), light auto transport (k_1, f_1), heavy trucks (k_2, f_2) and rail transport (k_3, f_3), then the final area-to-point connection is, at the most, an assembly of the third order. This does not mean that the area covered by the largest assembly $(A_{3,\mathrm{opt}})$ is small. The step changes in area size, from one assembly to the next, are described approximately by

$$\left(\frac{A_i}{A_{i-1}}\right)_{\mathrm{opt}} \sim \frac{k_{i-2}/f_{i-2}}{k_i/f_i} >> 1 \tag{26}$$

The new assembly is much larger than its predecessor when its cost parameter (k_i/f_i) is much smaller than the cost parameter of the assembly formed two steps earlier.

The transition from one area assembly to the assembly of the next higher order is abrupt: size and complexity change stepwise. Why should a producer or distributor opt for such abrupt changes? What drives these “transitions” in structure growth and development?

These questions are addressed by the rightmost column of Table 2. As we commented earlier, the transition from the elemental area ($A_{1,\text{opt}}$) to the first assembly ($A_{2,\text{opt}}$) is recommended by the promised increase in revenue per unit area. The ratio $\gamma g/3$ means that in a design of the elemental-area type, the producer receives only one third of the money paid by all the consumers, while the transporter receives the remaining two thirds. In an area design of the first assembly type, the producer and the transporter receive equal shares of the revenue generated per unit area. This is another example of the principle of equipartition, which is a frequent feature in engineering, geophysical and biophysical applications of constructal theory [6]. Equipartition of revenue between producer and transporter is preserved when $i \geqslant 2$, as shown in Table 2.

Another way to anticipate the stepwise increase in area size and complexity is to invoke the maximization of net revenue in time, as shown in Fig. 4. The producer begins servicing one or more areas of the elemental type, A_1, which has one central route (k_1, f_1) and a diffuse mode of transport (k_0, f_0) that reaches every point. The elemental structure starts from $t = 0$, when $A = 0$, and expands in time until it reaches the optimal size and external shape represented by the point ($A_{1,\text{opt}}$, $R_{1,\max,\max}$). If the elemental structure would continue to expand beyond this optimum, the net revenue would drop.

The route to a higher revenue consists of assembling two or more structures of type $A_{1,\text{opt}}$ into a first assembly A_2, assuming that a new and less costly mode of transport is available (k_2, f_2). The resulting structure (point b in Fig. 4) is not an optimum. However, it produces a revenue that is greater than the sum of the maximized revenues of the components of type $A_{1,\text{opt}}$. The A_2 structure grows from point b until it reaches its peak of revenue production ($R_{2,\max}$, $A_{2,\text{opt}}$). The process repeats itself beyond this second peak: sudden coalescence and a jump in revenue are followed by a gradual increase in revenue until a new maximum is reached. The horizontal distance between $A_{2,\text{opt}}$ and the abscissa of point c is at least as large as $A_{2,\text{opt}}$: in Fig. 4, this distance was drawn shorter because of space limitations.

Table 2
The optimized spatial structure of point-to-area or area-to-point transactions for maximum revenue

i	$n_{i,\text{opt}}$	$(H_i/L_i)_{\text{opt}}$	$A_{i,\text{opt}}$	$R_{i,\max}$ [a]	$R_{i,\max}/A_{i,\text{opt}}$
1	—	$\frac{2k_1/f_1}{k_0/f_0}$	$\frac{8}{9}g^2\frac{f_0f_1}{k_0k_1}$	$\frac{8}{27}\gamma g^3\frac{f_0f_1}{k_0k_1}$	$\frac{1}{3}\gamma g$
2	$\frac{3k_0/f_0}{2k_2/f_2}$	$\frac{4k_2/f_2}{3k_1/f_1}$	$\frac{4}{3}g^2\frac{f_1f_2}{k_1k_2}$	$\frac{2}{3}\gamma g^3\frac{f_1f_2}{k_1k_2}$	$\frac{1}{2}\gamma g$
3	$\frac{3k_1/f_1}{2k_3/f_3}$	$2\frac{k_3/f_3}{k_2/f_2}$	$2g^2\frac{f_2f_3}{k_2k_3}$	$\gamma g^3\frac{f_2f_3}{k_2k_3}$	$\frac{1}{2}\gamma g$
$\geqslant 4$	$\frac{k_{i-2}/f_{i-2}}{2k_i/f_i}$	$\frac{2k_i/f_i}{k_{i-1}/f_{i-1}}$	$2g^2\frac{f_{i-1}f_i}{k_{i-1}k_i}$	$\gamma g^3\frac{f_{i-1}f_i}{k_{i-1}k_i}$	$\frac{1}{2}\gamma g$

[a] For $i = 1$, this parameter was maximized with respect to two variables, $R_{1,\max,\max}$.

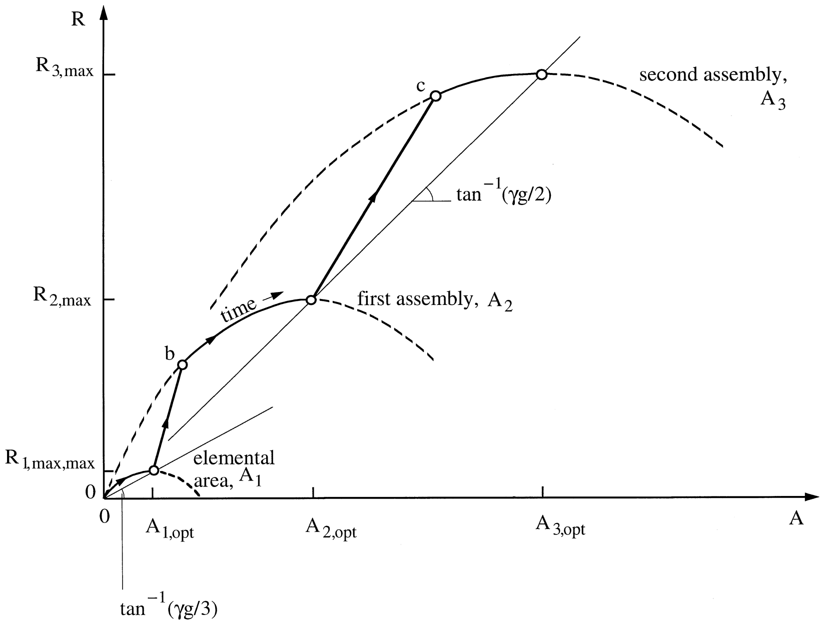

Fig. 4. Increase in the size and complexity of the area-point flow, as a result of the maximization of net revenue in time.

6. Optimally-shaped triangular areas

In this section, we show that costs can be reduced still further by optimizing the shapes of the areas that make up the constructal transport structure. In Section 2, all the area elements were assumed rectangular: that assumption allowed us to assemble area elements into a larger area, and to cover the larger area completely. The downside of the rectangular shape is that it enforces a non-uniform distribution of unit cost, especially when we compare the points situated on its perimeter. In Fig. 1, for example, the largest unit cost is "concentrated" in the two corners of type Q, because these corners are the points P situated the furthest from the common destination, or origin, M_1.

We reconsider the area-point cost minimization problem and begin the constructal sequence from the triangular area element shown in Fig. 5. The area $A_1 = B_1 L_1/2$ is fixed and is covered in the y direction by high cost transport (K_0). The path of relatively low cost transport coincides with the axis of symmetry x. The area's shape-parameter B_1/L_1 is not fixed. The total cost required to transport the stream of goods γA_1 between A_1 and M_1 is

$$C_1 = 2\gamma \int_0^{L_1} \left\{ \int_0^{H_1/2} [K_0 y + K_1(L_1 - x)] \mathrm{d}y \right\} \mathrm{d}x$$
$$= 2\gamma \left(\frac{1}{24} K_0 B_1^2 L_1 + \frac{1}{12} K_1 B_1 L_1^2 \right) \quad (27)$$

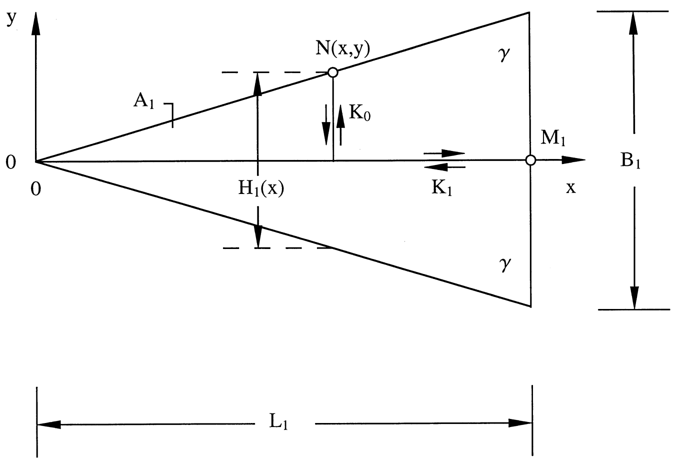

Fig. 5. Triangular elemental area A_1 for area-point transport; the double arrow means that the flow of goods may proceed in either direction.

where $H_1(x) = B_1x/L_1$ is the local transversal dimension of the triangular area. The C_1 expression can be minimized with respect to the base/length ratio of the triangle, with the following results:

$$B_{1,\text{opt}} = 2\left(\frac{K_1A_1}{K_0}\right)^{1/2} \quad L_{1,\text{opt}} = \left(\frac{K_0A_1}{K_1}\right)^{1/2} \tag{28}$$

$$\left(\frac{B_1}{L_1}\right)_{\text{opt}} = 2\frac{K_1}{K_0} \quad C_{1,\min} = \frac{2^{3/2}}{3}m_1(A_1K_0K_1/2)^{1/2} \tag{29}$$

Eq. (29) shows, first, that the optimal slenderness ratio of the triangle is exactly the same as the optimal slenderness ratio for a rectangular element of the same area — see Eq. (3). Secondly, the cost $C_{1,\min}$ is smaller than the corresponding cost in the rectangular design — see Eq. (4). The cost reduction factor associated with optimizing elemental triangles instead of elemental rectangles is $2^{3/2}/3 = 0.94$.

A special quality of the optimized triangle becomes apparent when we calculate the unit cost for transport between M_1 and an arbitrary point on its periphery, $N(x, H_1/2)$,

$$c_1 = K_0H_1/2 + K_1(L_1 - x) \tag{30}$$

Noting that $H_1 = B_1x/L_1$ and using the optimized aspect ratio $(B_1/L_1)_{\text{opt}}$ of Eq. (29), we conclude that the unit cost is constant, i.e. independent of the position x of the peripheral point N:

$$c_1 = K_1L_1 = \frac{1}{2}K_0B_1 = (A_1K_0K_1)^{1/2} \tag{31}$$

In this way we have deduced that the stretching of the points of largest unit cost into a continuous line (as opposed to one or two points) is a mechanism that generates not only minimal total cost for A_1 but also for the geometric form: i.e. the complete architecture of A_1.

Because of the constant-c_1 property of the symmetric portion of the perimeter of area A_1, the best geometric figure for A_1 is the isosceles triangle, and the best of all such triangles is the one with the slenderness $(B_1/L_1)_{\text{opt}}$. When the triangle is more slender than the best, $B_1/L_1 < (B_1/L_1)_{\text{opt}}$, the peripheral point of highest unit cost is the sharp tip of the triangle. When the triangular shape is less slender than the optimal triangle, the points of highest unit cost are the two corners that define the base B_1. When the slenderness is optimal, the sharp tip and the two base corners (and the lines that connect them) have the highest unit cost.

The interior of the optimally-shaped triangle is covered by a family of V-shaped lines of constant unit cost, which are terminated by the same base line ($x = L_1$). The sides of each V-shaped figure are parallel to the $y = \pm H_1(x)/2$ sides of the A_1 triangle. The point M_1 is the internal "line" $c_1 = 0$, i.e. the "center" of this family of nested and geometrically similar triangles.

The cost reduction that was achieved by switching from elemental rectangles (Fig. 1) to elemental triangles (Fig. 5) is even more attractive for larger area scales. Consider the construct of area A_2 shown in Fig. 6, which is a substitute for Fig. 2. The shape of A_2, or the function $H_2(x)$, is not specified. The A_2 territory is covered incompletely by an infinity of geometrically similar triangular elements, $A_1(x)$, where $L_1(x) = H_2(x)/2$. Each elemental triangle has the slenderness ratio $B_1(x)/L_1(x) = 2K_1/K_0$ [cf. Eq. (29)]; each has a constant-c_1 periphery. However, the c_1 value

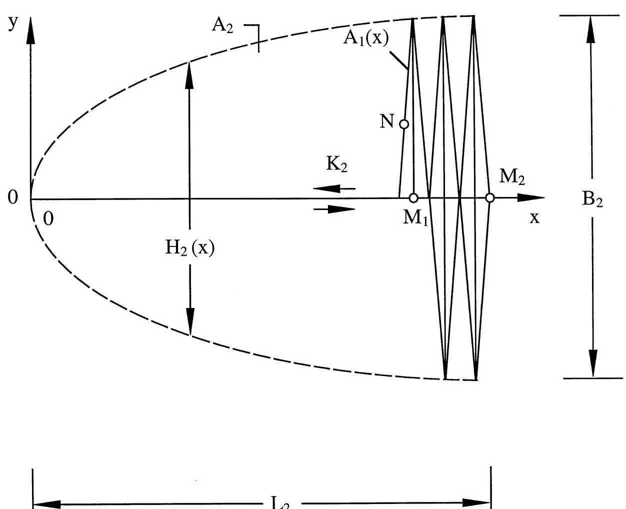

Fig. 6. Second area A_2, with unspecified shape and covered incompletely by triangular elements; the double arrow means that the flow of goods may proceed in either direction.

depends on the position x on the symmetry axis (K_2), namely $c_1(x) = K_1L_1(x)$, cf. Eq. (31).

The optimal shape of A_2 follows from the requirement that the unit cost (c_2) between a peripheral point (N) and new destination point M_2 be the same for every peripheral point, i.e. all along the contour of the territory covered by the transport:

$$c_2 = c_1(x) + K_2(L_2 - x), \tag{32}$$

Recalling that $c_1 = K_1L_1(x) = K_1H_2(x)/2$, we conclude from Eq. (32) that H_2 is linear in x, i.e. the optimal A_2 shape is an isosceles triangle of base B_2 and length L_2. We also find that $c_2 = K_1B_2/2 = K_2L_2$, and consequently

$$\left(\frac{B_2}{L_2}\right)_{\text{opt}} = 2\frac{K_2}{K_1} \quad L_2 = \left(A_2\frac{K_1}{K_2}\right)^{1/2} \quad B_2 = 2\left(A_2\frac{K_2}{K_1}\right)^{1/2} \tag{33}$$

If $A_{1,\max}$ is the largest elemental triangle used in the A_2 construct, namely $A_{1,\max} = A_1(x = L_2)$, then Eq. (33) also means that $A_2/A_{1,\max} = K_0/K_2 >> 1$.

To calculate the total cost associated with the transport between the A_2 construct and the point M_2, we analyze the stream of goods that arrives at the location x, from above and below the K_2 axis. There are two elemental triangles of size $A_1(x)$ that share the same base $B_1(x)$. They produce the stream $2m_1$, and, per unit length ($\mathrm{d}x$), the stream $m_1' = 2m_1/B_1$, where $m_1 = \gamma A_1$. The total stream that flows between the A_2 construct and the point M_2 is

$$m_2 = \int_0^{L_2} m_1' \mathrm{d}x = \frac{\gamma}{2}A_2 \tag{34}$$

Note that the size of m_2 is only half of γA_2 because the triangular elements cover exactly half of the area A_2. The other half of A_2 is the sum of all the triangular interstices situated between the adjacent $A_1(x)$ elements.

The cost associated with the stream $m_1'\,\mathrm{d}x$ has two components. Along the K_2 axis, the cost component is $K_2(L_2 - x)m_1'\mathrm{d}x$. Off the K_2 axis, over the vertical stripe of width $\mathrm{d}x$, the cost is $C_1'\,\mathrm{d}x$, where $C_1' = 2C_{1,\min}/B_1$, and $C_{1,\min}$ is given by Eq. (29). Thus the total cost is delivered by the integral

$$C_{2,\min} = \int_0^{L_2} \left[C_1' + m_1'K_2(L_2 - x)\right]\mathrm{d}x \tag{35}$$

which, after some algebra and the use of Eq. (34), yields

$$C_{2,\min} = \frac{7}{9}m_2(A_2K_1K_2)^{1/2} \tag{36}$$

Next to Eq. (8), this result shows that the constant-cost shaping of the periphery of the transport territory inscribed in A_2 produces a significant reduction in the total cost. The cost reduction factor (7/9 = 0.78) is more significant than the reduction registered at the elemental level.

The analysis and optimization of larger constructs follow the steps that we just outlined for the A_2 construct. We find that each new (larger) area must be an isosceles triangle of a certain slenderness ratio, so that each point on its toothy periphery is characterized by the same unit cost. This conclusion was drawn directly on Fig. 7, which shows the A_3 triangle covered by an infinite number of geometrically similar $A_2(x)$ triangles. The stream of goods that arrives on the K_3 axis at the location x is $m'_2 = 2m_2(x)/B_2(x)$. The cost associated with this stream in regions off the K_3 axis is $C'_2 = 2C_{2,\min}(x)/B_1(x)$. We obtain the following results, in this order:

$$\left(\frac{B_3}{L_3}\right)_{\text{opt}} = 2\frac{K_3}{K_2} \quad L_3 = \left(A_3\frac{K_2}{K_3}\right)^{1/2} \quad B_3 = 2\left(A_3\frac{K_3}{K_2}\right)^{1/2} \tag{37}$$

$$m_3 = \frac{\gamma}{2^2}A_3 \quad C_{3,\min} = \frac{23}{27}m_3(A_3K_2K_3)^{1/2} \tag{38}$$

The second part of Eq. (38) shows that relative to the corresponding rectangular design ($i = 3$, Table 1), the total cost has been reduced by the factor 23/27 = 0.85. This reduction is comparable with the reduction obtained at the A_2 level, see Eq. (36). The relative size of the A_3 construct is $A_3/A_{2,\max} = K_1/K_3 > 1$, where $A_{2,\max}$ is the largest $A_2(x)$ construct used in the A_3 internal structure, namely $A_{2,\max} = A_2(x = L_3)$.

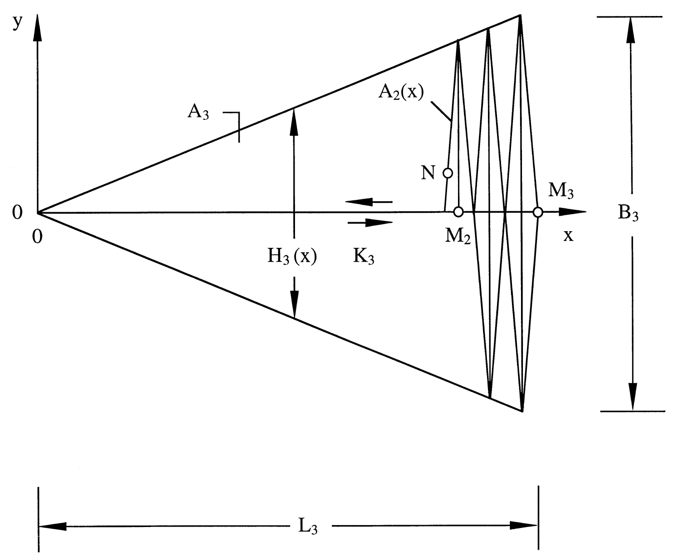

Fig. 7. Triangular third area A_3, covered incompletely by triangular constructs A_2; the double arrow means that the flow of goods may proceed in either direction.

The fourth construct is another triangle (A_4), the shape of which is described by relations that fit the pattern visible already in Eqs. (33) and (37). The total flow rate of goods is $m_4 = \gamma A_4/2^3$, and the associated minimized cost is $C_{4,\min} = (73/81)\ m_4\ (A_4 K_3 K_4)^{1/2}$.

The paragraph written at the start of Section 5 applies here as well. Although the sequence analyzed in this section can be continued to areas larger and more complicated than A_4, what we have presented is sufficient for the deduction of several important conclusions. The newest is that the triangle-in-triangle constructs (Figs. 6 and 7) look more "fractal", unlike their rectangle-based counterparts (Fig. 2). The reason is that, in principle, in each larger triangle, we could fit an infinite number of slender and geometrically-similar triangles. On the other hand, in the present theory and the earlier extensions of constructal theory [5,6], the smallest element size (volume, area) is finite. In the elemental unit (the "atom"), the flow is ruled by laws that differ from those that govern the flows at larger scales. Constructal theory is an atomistic theory.

Constructs with triangular components happen to look more "natural", i.e. more like the dendritic patterns found in nature (e.g. leaves, young urban growth, fingers, dendritic crystals). The constructs based on rectangular components fill the allotted space better (completely). This observation suggests that "the urge to optimize" is why natural structures tend to look more and more fractal-like. In other words, it is the refining of the performance of a rough design (e.g. the Euclidian structure of Fig. 2) that pushes the design towards a fractal structure. This tendency has more general implications, not just in spatial economics (see Section 8). It applies to all natural tree-shaped structures, which are always imperfect and incomplete (Euclidian), and in which the trend toward consistent refinements (and fractal-looking structures) is evident.

7. Law of refraction in constructal theory

In the preceding constructions, we made the simplifying assumption that the angle between two successive paths of transport is 90°. For example, in the cost minimization analysis based on Fig. 1, we started with the assumption that the K_0 path is perpendicular to the K_1 path. In Fig. 2, K_2 was perpendicular to K_1, and so on. These angles of confluence too can be optimized to decrease further the cost per assembly, or to maximize the revenue per assembly.

To illustrate the angle optimization opportunity, consider again the elemental area A_1 of Fig. 1, but this time allow the most distant corner (Q) to be reached along a K_0 path that makes an unspecified angle β with the line perpendicular to K_1. This more general situation is shown in Fig. 8. The per-unit-time cost required to transport one unit of goods from point Q to M_1 via the turning point T is

$$c = \frac{K_0 A_1}{2L_1}\left(\frac{1}{\cos\beta}\frac{K_1}{K_0}\tan\beta\right) + K_1 L_1 \tag{39}$$

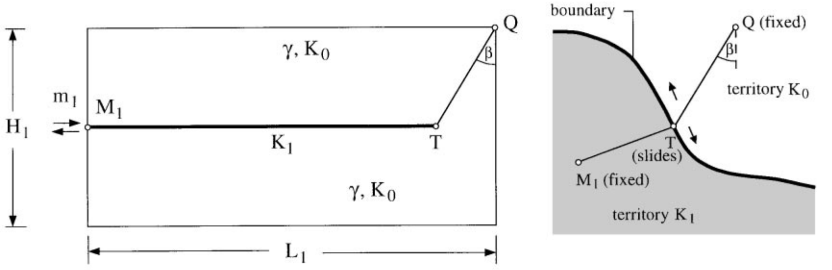

Fig. 8. (a) Elemental area with variable angle of confluence between the K_0 and K_1 paths. (b) The refraction principle of minimizing the cost of transport between two fixed points.

This cost can be minimized not only by choosing the external shape H_1/L_1 [as in Eq. (3)], but also by selecting the angle β. The results of this two-variable minimization procedure are

$$\left(\frac{H_1}{L_1}\right)_{\text{opt}} = 2\tan\beta_{\text{opt}} \tag{40}$$

$$\beta_{\text{opt}} = \sin^{-1}(K_1/K_0) \tag{41}$$

When K_1 is sensibly smaller than K_0, β_{opt} is negligibly small, the K_0 and K_1 paths become perpendicular, and Eq. (40) approaches Eq. (3).

In general, there is an optimal angle of confluence, or an optimal angle of "refraction" if we liken the broken line QTM$_1$ as a ray of light that passes from a low speed medium (K_0) into a high speed medium (K_1). The analogy between this angle optimization and Fermat's principle of light refraction has been noted independently in two fields [6,22]. In constructal theory, it was shown that the confluence angle can be optimized at every subsequent level of assembly, as in the minimum-time constructs for point-to-area travel reported in Ref. [6]. In economics, beginning with Lösch's treatise [22], the refraction principle is a recognized deterministic method of transport route maximization [23,24]. Lösch's angle optimization is different from the constructal version because, first, it does not recognize the opportunity to optimize the shape of the territory in which refraction occurs, and, secondly, the break point T is constrained to slide along a given curve that serves as the boundary between territory K_0 and territory K_1 (see the right-hand side of Fig. 8).

In summary, Lösch's angle optimization is simply an analog of Fermat's, because it is about point-to-point travel across a given boundary, whereas in constructal theory, angle optimization is one feature in a more complex geometric construction for point-to-area or area-to-point transport. In this paper, angle optimization was not emphasized because its impact is minor. Routes are nearly perpendicular when the unit cost sequence is steep ($K_{i-1} >> K_i$), and the other degree of freedom — the optimization of external area shape and numbers of constituents in new assemblies — is solely responsible for the formation and growth of the structure in space and time.

8. The constructal law, or the law of parsimony

In this paper we showed that by minimizing the cost in point-to-area or area-to-point transport, it is possible to anticipate the formation and growth of dendritic routes over a growing territory. The generation of structure is a reflection of the optimization of area at each area scale. We also showed that by maximizing revenue in point-to-area or area-to-point, transactions it is possible to anticipate not only the expanding (compounding) dendrites of transport routes, but also the size of the smallest area element that is accessed via the highest unit cost available (K_0). Every geometric detail of the structure is deterministic. It is the result of invoking only one principle.

The principle is known in economics as the law of parsimony [22]. In physics, biology and engineering, constructal theory unveils this principle as a law of access optimization for internal currents. We have shown in this paper that the same law that generates structure in natural flow systems far from internal equilibrium also generates structure in economics. The wide applicability of the law of parsimony in economics has been acclaimed [22–24], but its manifestations outside the realm of economics were not emphasized. On the contrary, they were disclaimed. For example, Haggett and Chorley wrote (Ref. [24], p. v):

> We are too conscious of the dangers of easy analogy and strained metaphor to claim that, for example, stream systems and transport systems are geographically 'the same'; to do so would force us to ignore aspects of network structure and evolution that are intrinsically important to physical and human geographers respectively.

Our conclusion is to claim precisely what Haggett and Chorley have rejected. The similarities between the spatial structures of physical and economic flows are not mathematical coincidences. These structures are deterministic and are generated by the same principle. The constructal law or its earlier statements (the law of parsimony, Fermat's and Heron of Alexandria's principle [6]) is the universal law of nature that accounts for the generation of shape and structure in heterogeneous flow systems subjected to constraints.

Acknowledgements

The authors thank the reviewers for their insightful and constructive comments. Professors Badescu and De Vos acknowledge with gratitude the support received from the Commission of the European Communities through the Inco-Copernicus keep-in-touch action Carnet 2. Professor Bejan's work was sponsored by the National Science Foundation.

References

[1] Bejan A, Tsatsaronis G, Moran M. Thermal design and optimization. New York: Wiley, 1996.
[2] De Vos A. Energy Conversion and Management 1995;36:1.
[3] De Vos A. Energy Conversion and Management 1997;38:311.
[4] De Vos A. Energy Conversion and Management 1999;40:1009.
[5] Bejan A. Int J Heat Mass Transfer 1997;40:799.
[6] Bejan A. Advanced engineering thermodynamics. 2nd Ed. New York: Wiley, 1997.
[7] Benítez P, Mohedano R, Minano J. Conversion efficiency increase of concentration solar cells by means of non-uniform illumination. In: Proceedings of the 14th European Photovoltaic Solar Energy Conference, Barcelona, 1997. p. 2378.
[8] Ledezma GA, Bejan A, Errera MR. J Appl Phys 1997;82:89.
[9] Dan N, Bejan A. J Appl Phys 1998;84:3042.
[10] Bejan A, Dan N. J Heat Transfer 1999;121:6.
[11] Rodriguez-Iturbe I, Rinaldo A. Fractal river basins. Cambridge (UK): Cambridge University Press, 1997.
[12] Meakin P. Fractals, scaling and growth far from equilibrium. Cambridge (UK): Cambridge University Press, 1998.
[13] Tansel BC, Francis RL, Lowe TJ. Management Science 1983;29:482.
[14] Hakimi SL. Annals Operations Research 1986;6:77.
[15] Brandeau ML, Chiu SS, Batta R. Annals Operations Research 1986;6:223.
[16] Tamura H, Sengoku M, Shinoda S, Abe T. IEICE Trans 1990;E73:1989.
[17] Chaudhry SS, Choi I-C, Smith DK. Int J Operations & Production Management 1995;15:75.
[18] Watanabe K, Tamura H, Sengoku M. IEICE Trans Fundamentals 1996;E79-A:1495.
[19] Longley P, Batty M. Spatial analysis: modelling in a GIS environment. Cambridge (UK): Geo Information International, 1996.
[20] Church RL, Sorensen P. Integrating normative location models into gis: problems and prospects with the p-median problem. In: Spatial analysis: modelling in a GIS environment. Cambridge (UK): GeoInformation International, 1996 (chapter 9).
[21] Hakimi SL. Operations Res 1964;11:450.
[22] Lösch A. The economics of location. New Haven: Yale University Press, 1954.
[23] Haggett P. Locational analysis in human geography. London: Edward Arnold, 1965.
[24] Haggett P, Chorley RJ. Network analysis in geography. New York: St. Martin's Press, 1969.