Geometrics as an Approach to Macroscopic Theories of Traffic Flow

ROBERT BRENNER, Engineer-Statistician, Institute of Transportation and Traffic Engineering, University of California, Los Angeles

• GEOMETRICALLY depicting families of relationships is one method for storing large amounts of information in very compact form, as, for example, the thermodynamic properties of steam in the Mollier chart. Geometric portrayals can also suggest and clarify otherwise obscure, or at least not easily recognized, theoretical considerations. From relatively abstract geometrical surfaces, Gibbs (2) deducts fundamental thermodynamic concepts. As suggested by this classic thermodynamic approach, this paper presents similar geometrical treatments of transportation relationships in order to provide a convenient method for compact storage of large amounts of information and a basis from which new theories may be identified.

The geometrical schema are a part of a "macroscopic" approach to traffic flow. Several concepts characterize this approach. One is that the effectiveness of any given transportation link is to be deduced from measurements made at inbound and outbound cordon lines, referred to as boundaries, without making any intervening measurements. This boundary-to-boundary treatment is similar to, and in fact was suggested by, classical thermodynamic problem solving, wherein energy and mass fluxes are measured at boundaries of some abstract system and from these measurements state changes of the system are deduced. This particular mode of thermodynamic analysis is often referred to as macroscopic; hence, the same term is used to describe the analogous approach to traffic flow. As demonstrated presently, the purposes of this macroscopic formulation are to provide bases for making traffic flow theory more readily applicable to operational problems and to link traffic flow theory to the larger picture of urban and regional planning.

SOME CONCEPTS RELATED TO MACROSCOPIC THEORY OF TRAFFIC FLOW

The Moving Unit

It is convenient to consider the transportation function in an urbanized region as the movement of a collection of units from node to node. A given transportation problem might optimally have the moving unit defined in one way; another problem might require a different definition. In traffic analyses, the vehicle would seemingly be the most appropriate unit. In transit studies, it might be the person being moved. In the movement of goods, it might be some amount of mass, or some amount value of volume. For purposes of developing the conceptual framework, the moving unit will first be limited to a single vehicle. Later, the moving unit will be groups of vehicles.

The System

Transportation links interconnect the nodes in the urbanized region. With the vehicle as the moving unit, the transportation link of interest will be a segment of road, street, or freeway over which vehicles normally travel. Each of these "real" transportation links is to be mapped into one or more abstract systems. Each system (SYS) is defined by at least one input boundary (INB) and by at least one output boundary (OUB). The boundaries may be arbitrarily located to coincide with any point in the real network.

Paper sponsored by Committee on Theory of Traffic Flow.

A moving unit is "in" a link or system when it crosses an input boundary; it is "out of" a system when it crosses an output boundary. Every boundary is considered to be infinitely thin in the dimension parallel to the direction of travel. The moving unit is also considered to be infinitely thin in this dimension. Thus, at any instant, a unit is either "in" or "out of" a given system. Also, it must be in one, and only one, system at any given instant.

A system may have more than one input or output boundary, but it must have at least one of each. A given combination of an input and an output boundary will define a path (PTH). Thus, a system having multiple input or output boundaries will have multiple paths. This multiple path formulation characterizes real-world situations wherein a moving unit at a given point in a network may move in any of several alternative directions. These cases are shown in Figure 1 by a unit at INB 3A which can proceed either to OUB 3A or OUB 3B, or by a unit at OUB 3B which could have come from INB 3A or INB 3B.

To simplify the presentation, the path whereby a moving unit is to negotiate a system is considered fixed once the unit crosses an INB; it will not change its path through a system because of conditions inside the system. Stated another way, the intrasystem feedback to the moving unit is not treated in this paper. Another simplification is that each moving unit that enters a system continues through and clears the system. The variation of a moving unit terminating its travel in the system being analyzed is not considered here.

Properties of Systems

A fundamental concept advanced here is that each system has properties related to, and measurable by, the performance of moving units traversing it. These newly-defined properties are in addition to conventional geometric and materials properties such as width, length, curvature, superelevation, and surfacing. Travel time (i.e., delay) exemplifies a class of the new properties that are evaluated solely by measurements at boundaries. For the Ith moving unit traversing the Kth system,

$$\text{TRT(I.K)} = \text{CLT(I.OUB K)} - \text{CLT(I.INB K)} \tag{1}$$

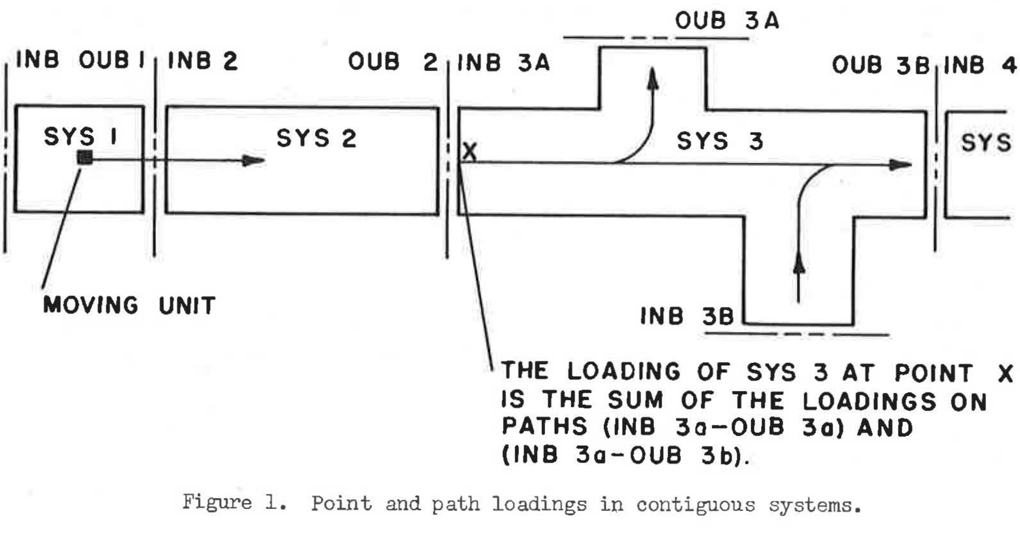


Figure 1. Point and path loadings in contiguous systems.

in which

TRT(I. K) = travel time of Ith moving unit traversing Kth system;
CLT(I. INB K) = clock time at which Ith unit arrived at input boundary of Kth system; and
CLT(I. OUB K) = clock time at which Ith unit arrived at output boundary of Kth system.

Smog, on the other hand, exemplifies another class of the new properties that would be measured best by integrating the performance of the moving unit over the entire length of the system.

$$\mathrm{SMG(I.\,K)} = \sum_{k} \mathrm{SMG(I.\,k)} \tag{2}$$

in which

SMG(I. K) = smog performance of Ith moving unit traversing Kth system;
SMG(I. k) = smog performance of Ith moving unit traversing subsystem k which is very small relative to overall Kth system, but very large in relation to domain over which SMG measurement is made; and
$K = \sum k$.

The performance measures may be summed over all moving units, which is the case of interest now. In a later work, the summation will be over all systems, which make up a path, thus permitting the value of the property for the path to be synthesized from the properties of the systems that comprise the path. For TRT and SMG, respectively,

$$\sum_{I} \mathrm{TRT(I.\,K)} = \mathrm{TRT(ALL\ I.\,K)} \tag{3}$$

$$\sum_{I} \mathrm{SMG(I.\,K)} = \mathrm{SMG(ALL\ I.\,K)} \tag{4}$$

Although the geometric and materials properties of a system are deterministic, the properties of the system due to the performance of moving units are probabilistic because, for one thing, the performance of each unit in the set ALL I need not be the same with regard to a given measure. Therefore, the system property might better be described by the statistical distribution of the values associated with the separate I's. This distribution might be described by any of several measures of central tendency, such as its mean, median, and mode. Thus, the smog property of a system might be the mean smog output of all moving units traversing the system, if the mean is selected as the convenient parameter for describing the property. Possibly some measure of dispersion, such as the standard deviation or range, might be used.

To simplify discussion, the ALL I will usually be dropped from expressions such as Eqs. 3 and 4, leaving SMG(K) and TRT(K), respectively. This latter usage emphasizes the fundamental concept of assigning to a system properties whose measure arises out of the performance of moving units in traversing the system.

Macroscopic Moving Units

Also fundamentally related to the proposed macroscopic approach to traffic flow is the performance of groups of vehicles as distinct from individual vehicles. Two types of vehicle groups are defined here: the ensemble and the loading.

Ensemble.—At the instant that the Ith car crosses the input boundary of the system [CLT(I. INB], there exists in the system some set of vehicles. As the Ith car proceeds through the system, the set of vehicles ahead of it changes; some vehicles in the original set leave the system at the output boundary or at intervening off-ramps, and others join the set from intervening on-ramps. The hypothetical summation of the set of vehicles ahead of the Ith car at all instants between CLT(I. INB) and CLT(I. OUB) is defined here as its ensemble [ENS(I)]. Although a random variable, the ENS(I) is not treated in this work as a performance measure per se of the system, but rather as a covariate that must somehow be taken into acount in interpreting any measurement of a performance variable of interest, such as travel time.

Loading. —During a specified interval of clock time, a group of vehicles will arrive at and enter a system at each input boundary. There will be a separate group for each input boundary, but all groups defined by the same clock time interval will be collectively referred to as the loading (LDG).

The clock time boundaries can be varied at will to reflect any given operational problem or situation. For example, the period from 4 to 6 PM on a weekday generally coincides with the afternoon travel peak, and thus can be referred to as the afternoon peak loading. Another reason for formulating the loading concept around clock time boundaries is the belief that loadings defined in this manner are closely correlated with community activities, and because community activities are highly predictable with time of day, the loadings would also be highly predictable. For example, the loading on an input boundary described around the parking lot of a company will be closely linked with the quitting time of the company. Further, the loading on some nearby system will have a high auto-correlation with the loading at the parking lot.

Symbolically, the loading for the interval CLT(INB) = a and CLT(INB) = b is comprised of all vehicles meeting the requirement that

$$a \leq \mathrm{CLT(I.\,INB)} \leq b \tag{5}$$

If there are, in all, m input boundaries to a system, the number of units in the system loading for the a to b interval is

$$N\left[\mathrm{LDG}\,\{\mathrm{SYS}\ a\text{-}b\}\right] = \sum_{i=1}^{m} N\left[\mathrm{LDG}\,\{(\mathrm{INB} = i) \cdot (a\text{-}b)\}\right] \tag{6}$$

The loading at each INB can be factored according to OUB, which would be identically a path description. For n possible output boundaries for vehicles entering the system at INB 1,

$$N\left[\mathrm{LDG}\,\{(\mathrm{INB} = 1) \cdot (a\text{-}b)\}\right] = \sum_{j=1}^{n} N\left[\mathrm{LDG}\,\{(\mathrm{INB} = 1) \cdot (a\text{-}b) \cdot (\mathrm{OUB} = j)\}\right] \tag{7}$$

The loading may be described according to criteria other than path; for example, wheelbase classification, nodal point of origin in the region, and nodal destination in the region. Double, triple, and even higher order summations are possible on various combinations of these and other appropriate descriptive criteria.

The loading and the ensemble represent groups of vehicles that are closely related, and which can be mapped into each other, on appropriate adjustment for phase relationships in clock time. Nevertheless, there are basic differences between them. The ensemble is a group of vehicles that is an impedance to travel of the moving unit; studies of these groups appear to be particularly well suited for examining unusual "within system" costs of movement. The loading, on the other hand, is a group of vehicles that comprise the demand made on the system for service; studies of these groups appear to be particularly well suited for examining the interaction of transportation with contiguous land usage.

In the same sense that a (performance) property of a system is to be quantified by measuring the performance of a single moving unit traversing the system, the quantification may be based on the performance of some group of moving units (e.g., the loading) traversing the system. How the measurement should be made of performance of the group is a matter of choice—summation of performance of each unit in the group vs some mathematical manipulation of the performance of the first and last units of the group. In the following presentation of the geometrical schema, to a system performance measure based on the performance of a single moving unit is considered the lower bound of a group measure, that is; it is the measure on a "group of 1."

GEOMETRICAL SCHEMA

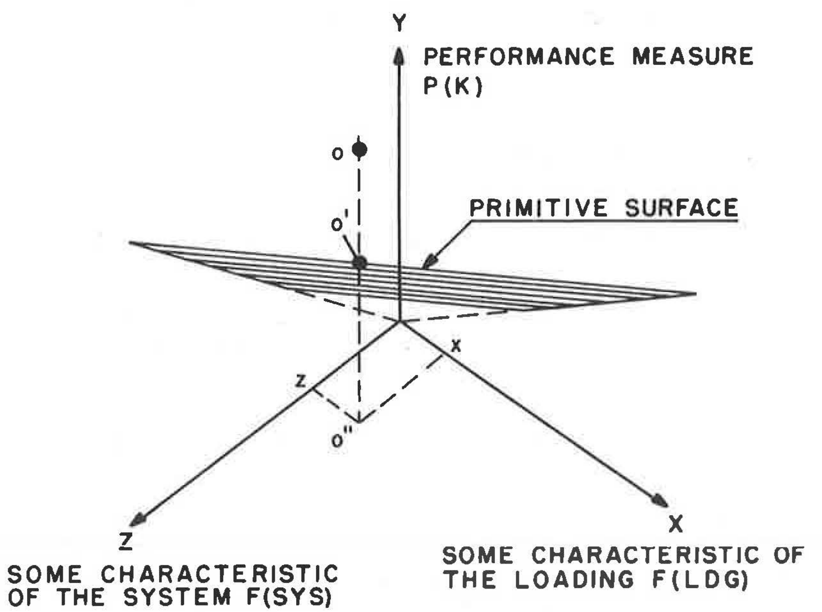
The Fundamental Space

It is proposed that a fundamental space for transport phenomena be defined by the following cartesian coordinates, referred to as the fundamental coordinates:

Coordinate X = some characteristic of the loading F(LDG).
Coordinate Y = some property P(K) which is a performance measure for the Kth system.
Coordinate Z = some characteristic of the system.

The space is shown in Figure 2; the projections on the ZY and XY planes are shown in Figures 3 and 4, respectively.

The X-coordinate may be any characteristic of the loading, some examples being the number of moving units arriving at a given INB in unit time, the make-up of this



Coordinate X: Some characteristic of the loading upon the K^{th} system.
Example:
F(LDG) = N(CLT 1 - CLT 2)

Coordinate Y: Some performance measure expressed as property of the K^{th} system--P(K).

Coordinate Z: Some characteristic of the system F(SYS).
Example:
Length of the system

Figure 2. Proposed fundamental space.

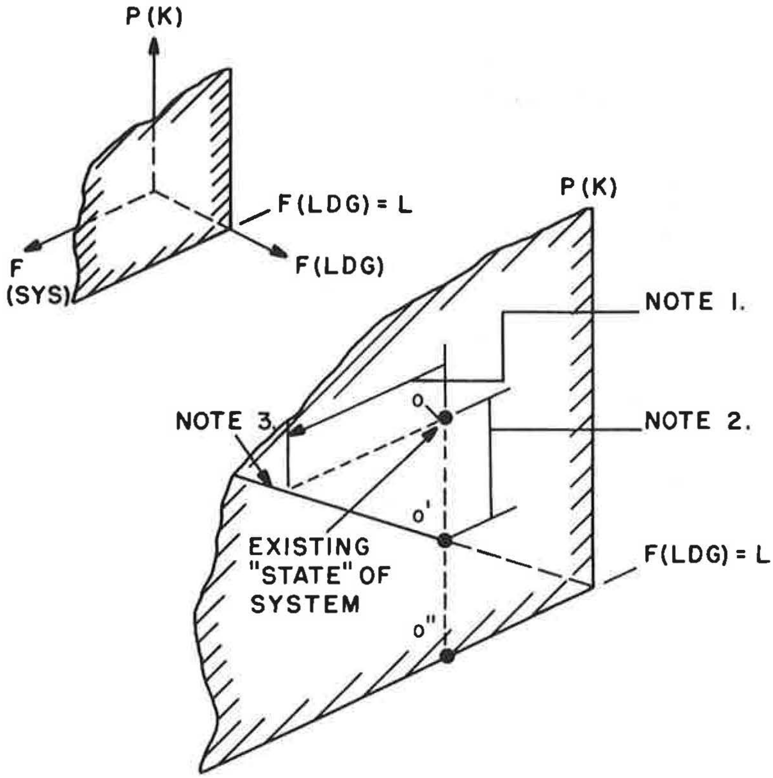


Figure 3. Fixed loading argument for fundamental space.

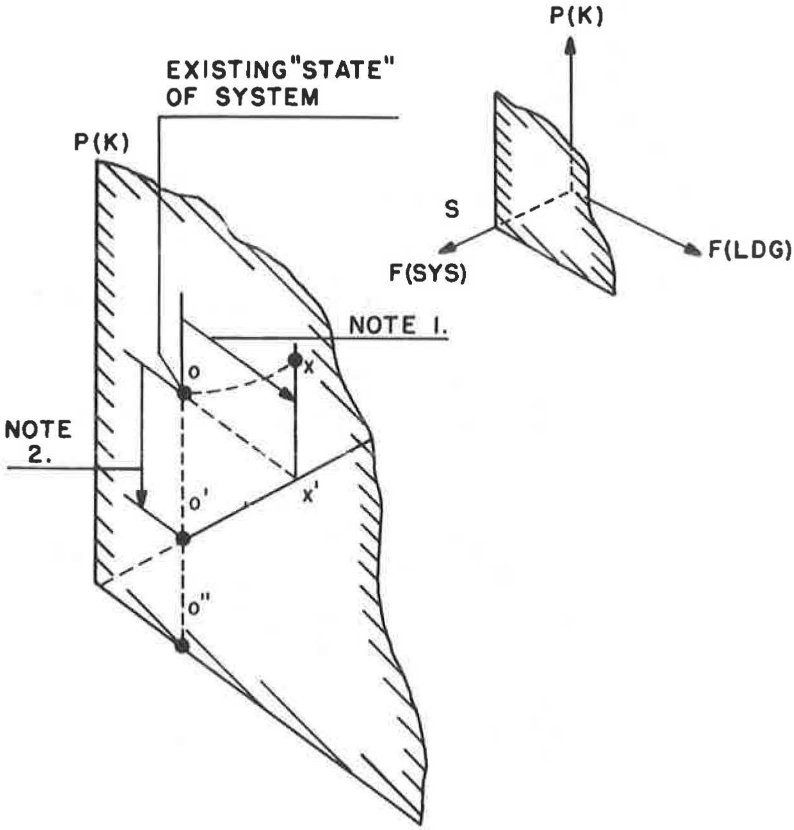
collection of vehicles according to type of vehicle (private cars, trucks, buses, etc.), and the OUB to which the moving unit is heading.

The Y-coordinate is to be any newly-defined system property that is a performance measure of the system. In this work, travel time is used. Other measures such as smog, dollars, and accidents might also be used. The expression P(K) is used as a generic description of any performance property, and the concern is to develop a geometry that portrays the fluctuation of this measure in the defined space.

The Z-coordinate may be any characteristic of the system. The distance between INB and OUB might be used. Other geometric dimensions also might be used, such as curvature, superelevation, number of lanes, and number of intervening on-ramps in the case of a freeway network. The dimension, traffic signals per mile, used by Irwin, Dodd, and von Cube (3) is used similarly in this context.

Primitive Surface and State Function

The existence of some optimum value P′(K) of the performance measure P(K) is now posited for a given loading on a given system. The primitive surface is defined as the locus of this optimum value for all combinations of loading and system; that is, all admissible pairings of X, Z values. The optimum value might not be capable of ever being realized. For example, with travel time as the performance measure, the criterion for the primitive surface might be that the moving unit be traveling at the speed of light, and if the characteristic of the system was its length, the primitive surface would describe the time for light to cover the distance Z. A more practical criterion for the primitive surface might be some defined maximum operating speed, or, as presently discussed in more detail, the primitive surface might be the lower limit of some confidence interval derived from purely empirical considerations. It suffices now to assert the existence of some such optimum surface.



NOTE 1. MAXIMUM INCREASE IN F(LDG) AT F(SYS) = S WITHOUT CHANGING P(K).

NOTE 2. MAXIMUM REDUCTION IN P(K) AT F(SYS) = S WITHOUT CHANGING F(LDG).

Figure 4. Fixed system argument for fundamental space.

The system, its loading, and its performance value are represented by a single point o in the X, Y, Z space, called the state point. The line through the state point and perpendicular to the XZ plane will intersect the primitive surface at o′, which, by definition must be the optimum performance value for the given system-loading combination (X, Z). In Figure 2, the state point approaches its optimum o′ from above and would be the case for a performance measure such as travel time having some minimum value for its optimum. Other performance measures (e.g., some speed function) might have a maximum for the optimum, in which case the state point would approach its optimum from below. The former case appears to be more widely applicable and is used in presenting the concepts.

For some performance measures, the primitive surface might be defined negatively; that is, the state point might have a zero optimum and the primitive surface might be some undesired performance level. Number of accidents or the likelihood of accidents exemplifies a performance measure of this sort. Of course, the state point would approach such a primitive surface from below. The proximity of the state point to the surface would measure the relative hazard associated with negotiating the particular hypersystem.

There are basic differences in the ways in which the state point can shift in the fundamental space. The system change $\Delta F(SYS)$ is an independent argument which is deterministic in that the system boundaries can be varied at will. The loading change $\Delta F(LDG)$, and any performance change $\Delta P(K)$, would be stochastic. Any of these changes in a fundamental coordinate can take place with one, both, or neither of the other coordinates remaining fixed. Several interesting cases are briefly presented here.

The fixed loading argument for the fundamental space is shown in Figure 3. The line o'z' represents the intersection of the loading plane F(LDG) = L with the primitive surface. The existing state of the system is at o, and the line oz is everywhere in plane F(LDG) = L and has dP(K) = 0. If the primitive surface is a minimum, the line oz' represents the maximum increase possible in F(SYS) at F(LDG) = L without changing performance P(K). If, for example, F(SYS) is the distance between INB and OUB of SYS(K), then oz' represents the maximum theoretically possible increase in length of system without changing the P(K) for fixed loading. In actuality, the state point would not move along oz', but rather along some curve oz.

The point o' is the intersection of the locus of the state point with the primitive surface. The line oo' is the maximum theoretically possible reduction in P(K) at F(LDG) = L for F(SYS) = S.

The fixed system argument is shown in Figure 4. The line o'x' is the intersection of the system plane F(SYS) = S with the primitive surface. The line ox' is everywhere in plane F(SYS) = S, and has dP(K) = 0. Then, ox' represents the maximum theoretically possible increase in F(LDG) without changing the P(K) for the fixed system F(SYS) = S. In actuality, the state point would not move along ox', but rather along some curve ox.

The actual state function is shown as the surface zox in Figure 5, and every point on it is above its vertical projection on the primitive surface. For most applications, the state function should prove to be useful by itself; that is, without reference to the primitive surface. The primitive surface is interesting insofar as it can be, by definition, the boundary of the state function, and from it the maximum possible coordinate changes can be deduced for the state point.

There is additional motivation for the primitive surface. The actual state function shown in Figure 5 is surface zox. But the value of P(K), defined as a parameter of the Kth system, will have some probability distribution, and its variance will have both F(SYS) and F(LDG) variance components. Thus, the surface itself will be shifting with respect to the fixed primitive surface z'o'x'. The length z" represents the actual performance P(K) of the Kth system at F(SYS) = z" and F (LDG) = o". This length can be divided into two segments, z"z' and z'z. By definition, z"z' is a constant at F(SYS) = z", and F(LDG) = o", leaving z'z as a stochastic segment reflecting the random fluctuation of P(K) at F(SYS) = z" and F(LDG) = o". The primitive surface thus serves as a base from which the nonrandom contributions can be partialed out of the measured performance of the system.

The vertical distance from the primitive surface to the actual performance surface, typified by segment z'z, can be used in operational decision making involving several different performance measures. If the performance measure P(K) in Figure 5 is travel time and Q(K) is another performance measure such as "smog" output, for system F(SYS) = z", F(LDG) = o", there would be a value Q(K) describing the performance of the system on the Q-dimension. Similarly, each other system could be described by a set of performance measures P, Q, R, ... For each such measure, there would be its unique primitive surface (similar to z'o'x'), its actual surface (similar to zox), and finally a stochastic segment (similar to z'z). The expected values of these segments provide a basis for comparing the consequences of one operational decision on all performance measures.

To illustrate this comparison, if operation decision A causes SYS(z'''o") to prevail instead of SYS(z"o") with the accompanying changes A(P), A(Q), A(R), . . . , with A(P) being the change in the stochastic segment z'z in the zxP space, A(R) in the zxR space, etc., and further, if operational decision B produces changes B(P), B(Q), B(R), . . . , the criteria for selecting decision A over B, or vice versa, are as given in Table 1.

The dimensions of these various decision consequences need not be the same. The measure P, being travel time, would be in minutes; Q, representing "smog," might be

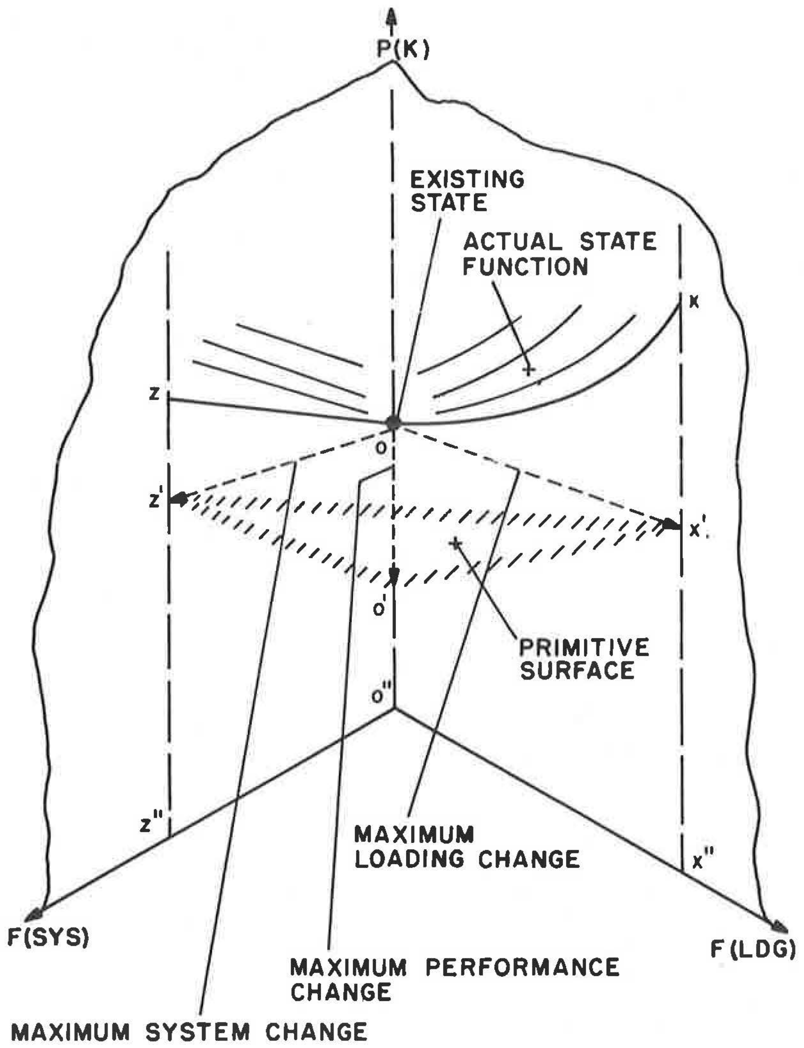


Figure 5. Maximum changes in state from actual state function to primitive surface.

TABLE 1

DECISION CONSEQUENCES

Decision A	Decision B	Criteria
A(P)	B(P)	A(P) - B(P)
A(Q)	B(Q)	A(Q) - B(Q)
A(R)	B(R)	A(R) - B(R)

expressed in mass of hydrocarbon; and R, as a measure of safety, might be expressed in probability units. Nor would the signs of the various consequences have to be the same within the same decision. For example, a reduction in travel time might be accompanied by an increase in accident likelihood.

The analysis might be continued into the subjective realm by assigning value to the different performance measures. If unit values of performance measures P(K), Q(K), R(K), ..., are p, q, r, ..., respectively, the decision consequences are [A(P)] [p], [A(Q)] [q], ... Although such application of relative values might be required for nontechnical reasons (e.g., political pressure or public relations), objective analyses seemingly should be limited to identifying A(P), A(Q), ... B(R), but not to treating them in functional combinations.

Critical Surface Concept

In the preceding section, some function of loading F(LDG) is treated as a fundamental coordinate, and the performance coordinate P(K) relates to the performance of this loading in negotiating the Kth system. In this section, the loading coordinate will be replaced with an ensemble coordinate F(ENS), and the performance coordinate will become a variable related to the Ith vehicle, P(I. K). The Ith vehicle might well be some parameter of the loading, and the F(ENS) would no doubt have some functional relationship to F(LDG). But, as stated before, the limitations of graphic representation on a two-dimensional drawing of a multidimensional space necessitate the separate expanded treatment.

The expression, ensemble of the Ith car ENS(I), describes the hypothetical summation of the set vehicles ahead of the Ith car at all instants between CLT(I. INB) and CLT (I. OUB). One way of classifying the vehicles in the ensemble is according to path. For example, the N(INB. OUB) and the N(LON. OUB) vehicles are in the ENS(I), but the one group takes one path (INB. OUB) in getting through the system, whereas the other group takes the path LON. OUB. Another classifying scheme is by wheelbase; for example, N(WBC 4) represents the number of vehicles in wheelbase category 4.

The formulation here treats the effects that the ensemble and two of its component subsets have on the value of P(I. K). The subsets considered are the number of commercial vehicles; i.e., trucks in the ensemble, and the number of on-ramp vehicles in the ensemble. The same methods can be used for other subsets as well.

Were there no vehicles in its ensemble, the Ith vehicle would perform at some optimum of its own selection. As the number of vehicles in each subset of the ensemble becomes larger, performance of the Ith vehicle is affected, presumably unfavorably. Different decrements in performance are to be expected with differently-constituted ensembles. The characteristic decrement pattern proposed here is that for any subset of the ensemble, the performance varies monotonically as the number of vehicles in the subset increases. For all sizes up to a given size of subset, the second derivative of performance with respect to size of subset is considered to be zero. At this point, it becomes positive, and may continue positive therefrom with increasing size of subset. This pattern is shown in Figure 6, in which the argument is the percentage of the ensemble that is on-ramp traffic. In Figure 7, the argument is the commercial vehicle percentage. In each case, the locus of the state point is shown for constant size of ensemble, and is referred to as a constant volume curve or simply "isovol."

Point B in Figure 6 represents the point for ISOVOL N(ENS) = N at which $\partial^2 P(I.K)/\partial(ONR)^2$ changes from zero to positive. The locus of point B over all isovols is defined as the "critical on-ramp curve." The significance of this curve is that all points to the left of it represent the states in which performance is changing relatively slowly, with increasing on-ramp traffic. Beyond the critical curve, the change is much more rapid and possibly nonlinear as well. The set of points to the left of the curve might be conveniently described as being in the "stable" region, and those to the right in the "unstable" region.

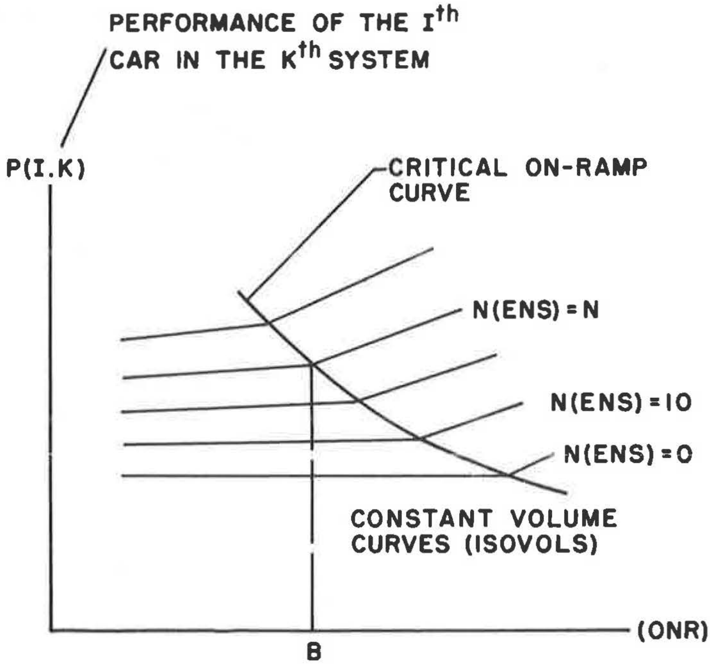
This pattern is offered largely on intuitive bases, although there has been some supporting empirical evidence in the work of others as well. Irwin et al. (3) present a family of curves of this nature. In each curve, travel time is depicted according to a basic capacity function:

$$T_V = t_i + \left[\frac{d\left(F_V - F_c\right)}{F_c}\right] L \tag{8}$$

in which

F_V = link volume (cars per hour per lane);

F_C = critical volume = practical capacity of link above which flow becomes unstable and travel time rises rapidly (cars per hour per lane);



$$(\mathrm{ONR}) = \text{On-Ramp Percentage} = \frac{\mathrm{N(ENS.\,ONR)}}{\mathrm{N(ENS)}}$$

N(ENS. ONR) = Number of on-ramp cars in the ensemble
N(ENS) = Number of all vehicles in the ensemble

The critical on-ramp curve is defined as the locus of point B, where:

$$\left[(\mathrm{ONR}) \leqq \mathrm{B}\right] \longrightarrow \left.\frac{\partial^2 \mathrm{P(I.\,K)}}{\partial(\mathrm{ONR})^2}\right)_{\mathrm{N}} = 0$$

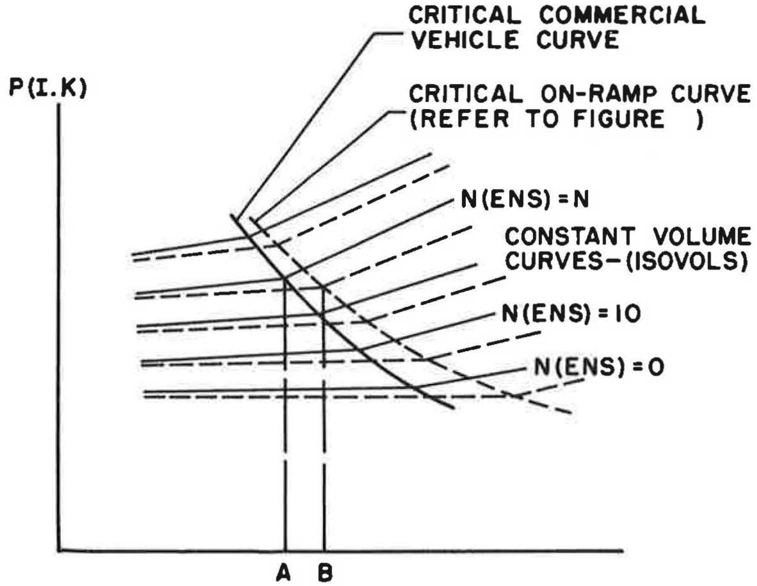
$$\left[(\mathrm{ONR}) > \mathrm{B}\right] \longrightarrow \left.\frac{\partial^2 \mathrm{P(I.\,K)}}{\partial(\mathrm{ONR})^2}\right)_{\mathrm{N}} > 0$$

Figure 6. Effect of on-ramp traffic on performance of Ith car in Kth system.

T_V = link travel time at volume F_V (minutes);
t_i = link travel time per mile at critical volume F_C (minutes per mile);
L = link length (miles); and
d = delay parameter (minutes per mile) = d_1 for $F_V \leq F_C$ and d_2 for $F_V > F_C$.

Irwin et al., point out that different values may be assigned for each category of link, and further report that their observations could be adequately fitted with the values d_1 = 0.5 and d_2 = 10.0. This capacity function coincides in form with that proposed here. Up to the critical volume F_C, the travel time curve rises comparatively slowly; beyond F_C, the slope is much greater.

The idea of Irwin et al. of unique delay parameters for each of their ten categories of links closely parallels the concept implicit in the linear model of travel time as given in the background work (1). It also closely corresponds to the concept of a stable array



(COM) = Commercial Vehicle Percentage

$$= \frac{N(ENS, COM)}{N(ENS)}$$

N(ENS, COM) = Number of commercial vehicles in the ensemble

N(ENS) = Number of all vehicles in the ensemble

The critical commercial vehicle curve is defined as the locus of point A, where:

$$\left[(COM) \leq A\right] \longrightarrow \left.\frac{\partial^2 P(I, K)}{\partial (COM)^2}\right)_N = 0$$

$$\left[(COM) > A\right] \longrightarrow \left.\frac{\partial^2 P(I, K)}{\partial (COM)^2}\right)_N > 0$$

Figure 7. Commercial vehicle effect on performance of Ith car in Kth system.

of performance constants for describing the network of systems. Their link categories are defined according to speed limit and number of signalized intersections per mile. Somewhat more general categories of systems are proposed here: tangent sections, curves, intersections, freeway lengths, etc.

The same pattern is found by Lehman (4) in a context that is even more relevant to this work. During the course of testing Wingo's (5) ingression treatment of traffic, Lehman obtained data from which the regression of TRT(I) on N(ENS) was obtained (Fig. 8). In this figure, travel time of the Ith car over the Kth system is plotted as a function of its ensemble ENS(I). The Kth system, in this case, is a 4-mi stretch of two-lane street used very heavily by commuters between the San Fernando Valley and West Los Angeles. The input boundary was at a point beyond which no additional traffic entered the system. The output boundary was at the top of a very steep hill (Roscomare Road and Mulholland Drive). All traffic of interest was northbound of Roscomare Road, and had to come to a complete stop at the intersection with Mulholland Drive.

The results show that, up to an ensemble of approximately 155 vehicles, the travel time function is increasing fairly slowly, with increasing N(ENS). At the "critical" N(ENS) = 155, the slope changes very sharply.

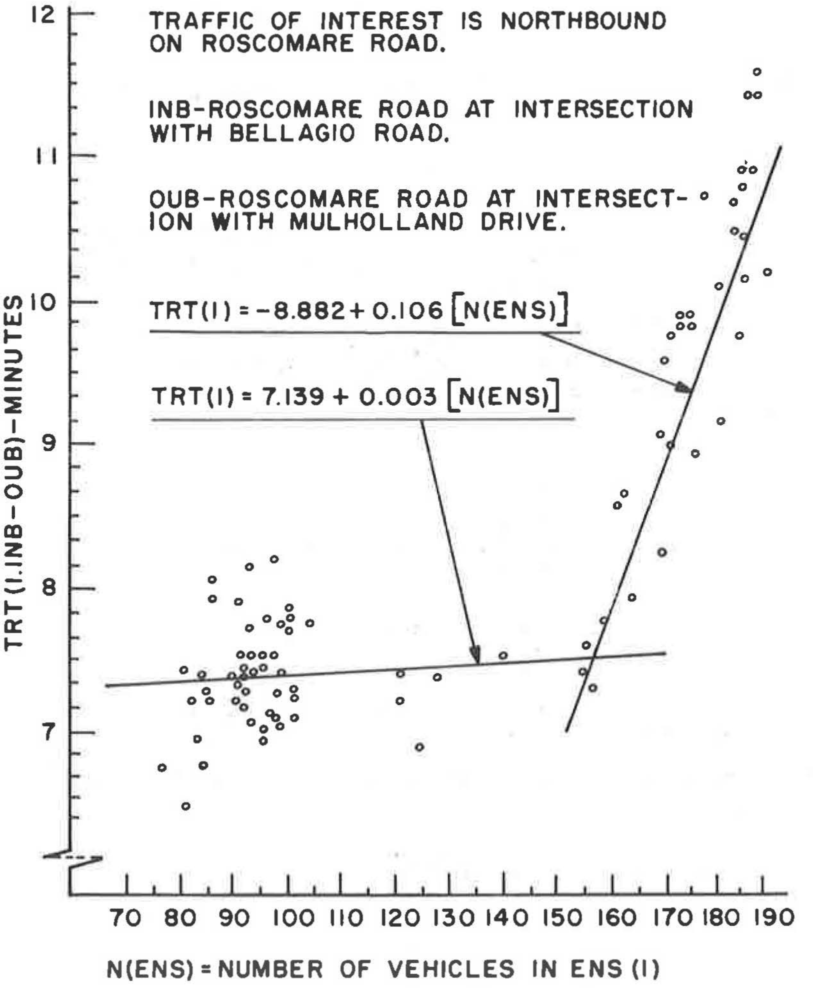


Figure 8. Regression of travel time on size of ensemble for travel over 4-mi length of surface street with severe choke at output boundary.

Wingo's expression for the ingression loss due to a choke (the choke here was the stop sign) is of the form:

$$T = T_o + \frac{N}{2C_c} \tag{9}$$

in which

T = travel time;
T_o = travel time of vehicle in system when that vehicle is only one making a demand on system;
N = number of vehicles making demand on system [essentially, N(ENS)]; and
C_c = choke capacity.

Lehman suggests a modification of Wingo's expression, proposing "that there is a number x such that x vehicles may be accommodated by the system without there being any instantaneous demand on the system resulting in an ingression loss." The resulting equation is

$$T = T_o + A\left(\frac{N - X}{C_c}\right) \tag{10}$$

in which A = 0 if $N \leq X$ and A = 1 if $N > X$.

Lehman's data were randomly split into two samples. From examination, no points occurred in the range $140 < N(ENS) < 150$. Consequently, the value N(ENS) = 145 was selected as the critical N(ENS), [the point x in (9. 2)]. The regression functions were found to be

For N(ENS) > 145 vehicles:

Sample A:

$$TRT(I) = -8.882 + 0.1058\,[N(ENS)] \tag{11}$$

Sample B:

$$TRT(I) = -\,8.197 + 0.1024\,[N(ENS)] \tag{12}$$

For N(ENS) < 145 vehicles:

Sample A:

$$TRT(I) = 7.139 + 0.0026\,[N(ENS)] \tag{13}$$

Sample B:

$$TRT(I) = 6.832 + 0.0057\,[N(ENS)] \tag{14}$$

The two regression lines for Sample A of the data are shown in Figure 8, and support the belief that there is a critical size of ensemble beyond which travel time rises rapidly.

Many other reported works show this pattern of a critical point up to which traffic performance changes gradually, and beyond which it deteriorates very rapidly. Essentially, this pattern is also being proposed here. Up to a given on-ramp percentage, travel time over the system for a given size ensemble is not seriously affected; beyond this point, travel time rises very rapidly. Up to a given percentage of commercial vehicles, the effect of commercial vehicles on other traffic is not serious; beyond this point, the performance of other traffic is sharply affected.

It is doubtful that, in actual traffic, the characteristic pattern is maintained insofar as the sharp discontinuity at which the second derivative goes from zero to positive is concerned. A more precise treatment would describe the transition as taking place over some domain rather than at a point. Instead of the critical curve, there would be a critical bandwidth which probably would change from one isovol to the next. Instead of being divided solely into a stable and an unstable region, the space would also have this bandwidth as a metastable region.

However, the concept of the space being divided into a stable and an unstable region holds, regardless of whether the demarcation is by the line or bandwidth. The line is used here to simplify the presentation.

Identification of the stable, unstable, and metastable regions of a space would provide (among other benefits) macroscopic warrants for operational decisions. A state point in the stable region would signify that not much of a performance betterment was to be achieved as contrasted with the potential achievement if the point is in the unstable region. The metastable region would be an inconclusive domain. In this way, the proposed geometrical schema offer one approach to a type of sensitivity analysis pursuant to operational decision making.

CONCLUSION

One of the several objectives of traffic flow theorists is to discover and validate analytical equations that describe traffic flow phenomena. Such equations lead to deeper insight into the problems and, hopefully, solutions that will be of value to practitioners.

Similarly, analytical equations have been widely used to describe thermodynamic properties of matter, and how these properties change with various thermodynamic processes. In the treatment of gas problems, for example, the perfect gas equation, PV = RT, is almost always used for first approximations. It can be sufficiently accurate for some applications, but not for others. One form of first-order correction is to use the van der Waals equation; a higher order correction is attained through use of virial equations.

Regardless of which equation is used, there is always the question of how well it represents the data. Tribus (6) states that from the point of view of the scientist the important question relates to the theoretical significance of the equation of state, (for example, the van der Waals equation purportedly corrects for the effective volumes of the molecules themselves and incremental pressure due to intermolecular forces), whereas from the point of view of the engineer, the question is how convenient the equation is to use and how badly it can be in error.

From Tribus' remarks, it can be inferred that although empirical data are not necessarily a beginning point from which theories are developed, sooner or later theoretical work must be validated against such data. Thus, some of the most important thermodynamic knowledge is empirical information, carefully organized in various tabular or geometric forms.

Similar arrangements of empirical data in geometrical spaces are proposed here to portray transportation relationships among selected coordinates. Seemingly, if empirical methods produce the standards by which analytical treatments are appraised in as rigorous a discipline as thermodynamics, similar empirical methods should be at least as important for evaluating analytical results in a less rigorous, newly-emerging discipline such as traffic flow theory.

Most traffic flow theorists would quickly grant the importance of empirical methods to their work, but would also point out, and with reason, that most existing traffic data are not suitable for their purposes. Although numerous traffic studies are published every year (and untold thousands of additional studies remain unpublished in the files of operational traffic departments), there is no evidence that the vast amount of data was collected under anything approaching similar field conditions. The data in transportation literature, save for isolated cases, can only be regarded as a very large collection of anecdotal information, independently meaningful, but certainly not suitable for testing any general hypotheses.

Yet the existence of these numerous data is evidence of a vast traffic measurement capability which, if organized, could provide for the needed empirical validation of theoretical findings. The proposed fundamental space and related geometrical schema might well provide a formalism for organizing these empirical data, and for standardizing the conditions under which such data would be collected in the future.

The benefits in having empirical traffic data in an organized fashion are readily recognized. They immediately provide for putting theoretical findings to Tribus' test of how badly they can be in error and how well they portray the real world. There is also the possibility that when organized in geometrical schema, new avenues of theoretical study will become apparent which are not as yet recognized from abstract reasoning.

Finally, there is the possibility that the geometrical schema will increase the usefulness of empirical data to the applied practitioner. The usefulness of Mollier charts for applied purposes is obvious. Similar representations should be equally valuable in traffic applications. More specifically, the geometrical schema advanced here are deliberately constructed to reduce the variance of estimates of traffic parameters. In the examples given in this paper, the schema eliminate two major variance components of the performance measure; namely, system characteristics and the loading on the system. Thus, measurements of performance of a curve are not to be grouped with those of a tangent section or with those of an intersection. Nor are measurements of performance of a system having a light loading to be grouped with those for a heavy loading. It seems perfectly natural from the standpoint of practical operations to keep dissimilar elements of a road network separated, and this is, in fact, the approach of most practitioners in the field. In this sense, the geometrical schema may be formally defining practices that engineers are already following informally.

REFERENCES

1. Brenner, R., "Analysis of Traffic Flow Using Time as the Depending Variable." Doctoral dissertation, Department of Engineering, University of California, Los Angeles (Feb. 1962).
2. Gibbs, J. W., "A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces." Collected Works of J. Willard Gibbs, Yale University Press, 1:33-54 (1956).
3. Irwin, V. A., Dodd, N., and von Cube, H. G., "Capacity Restraint in Assignment Programs." HRB Bull. 297, 109-127 (1961).
4. Lehman, G. V., "A Study of Ingression Loss in a Flow System." Department of Engineering, University of California, Los Angeles, Senior Project Report No. 1866 (May 1961).
5. Wingo, L., "Measurement of Congestion in Transportation Systems." HRB Bull. 221, 1-28 (1959).
6. Tribus, M., "Thermostatics and Thermodynamics." Van Nostrand (1959).