CAMBRIDGE

Cambridge University Press
978-0-521-87100-6 - Networks: Optimisation and Evolution
Peter Whittle
Excerpt
More information

Tour d'Horizon

Whither? Why?

The contents list gives a fair impression of the coverage attempted. Networks, both deterministic and stochastic, have emerged as objects of intense interest over recent decades. They occur in communication, traffic, computer, manufacturing and operational research contexts, and as models in almost any of the natural, economic and social sciences. Even engineering frame structures can be seen as networks that communicate stress from load to foundation.

We are concerned in this book with the characterisation of networks that are optimal for their purpose. This is a very natural ambition; so many structures in nature have been optimised by long adaptation, a suggestive mechanism in itself. It is an ambition with an inbuilt hurdle, however: one cannot consider optimisation of design without first considering optimisation of function, of the rules by which the network is to be operated. On the other hand, optimisation of function should find a more natural setting if it is coupled with the optimisation of design.

The mention of communication and computer networks raises examples of areas where theory barely keeps breathless pace with advancing technology. That is a degree of topicality we do not attempt, beyond setting up some basic links in the final chapters. It is well recognised that networks of commodity flow, electrical flow, traffic flow and even mechanical frame structures and biological bone structures have unifying features, and Part I is devoted to this important subclass of cases.

In Chapter 1 we consider a rather general model of commodity distribution for which flow is determined by an extremal principle. This may be a natural physical principle (e.g. the minimal dissipation principle of electrical flow) or an imposed economic principle (e.g. cost minimisation in the classic transport problem). The principle minimises a convex cost function, subject to balance constraints; classic Lagrangian methods are then applicable. These lead to the concept of a potential, and find their most pleasing development in the Michell theory of optimal structures, for which the optimal design is characterised beautifully in terms of the relevant potential field.

All this material is classic and well known, at least as far as flow determination is concerned – design optimisation may be another matter. However, we do find one class of cases for which the analysis proceeds particularly naturally. These are the models with *scale-seminvariant costs*, for which the cost of operating a link is the volume of the link (its length times some notion of a cross-sectional area) times a convex function of the flow density. The corresponding dual cost is volume times the dual function of potential gradient. Let us refer to these simply as 'seminvariant-cost models' or, even more briefly, as 'SC models'. They constitute quite a general class of cases, which indeed includes most of the standard examples. However, they turn out to have a property that is disconcertingly dramatic.

The SC models for simple flow all reduce under a free design optimisation to the net that solves the classic transport problem: an array of direct straight-line links from sources to sinks. If the commodity is not destination-specific, then these routes will never cross in the plane, and there is a corresponding property in higher dimensions.

This is a gratifying reduction, and one that is acceptable in some cases. In Chapter 2 we solve a continuum problem by these means, and the corresponding solutions of the structural problems in Chapters 7–9 are less naive. However, in most cases this simple solution is totally unrealistic. A practical road net could never be of that pattern, and to formalise the reasons why is an instructive exercise. One would expect 'trunking': that traffic of different types should share a common route over some distance, as manifested by local traffic's feeding into major arteries and then leaving them as the destination is approached.

If the network is subject to a variable load, then this indeed has the effect of inducing some trunking in the optimal design; see Chapter 4. The reason is that there is then an incentive for links to pay their way by carrying as steady a load as possible. Links will then be installed which can carry traffic under a variety of loads, usually making route-compromises in order to do so. However, once this steadiness of flow has been achieved, there is no incentive for further trunking.

The real incentive to trunking must be that a link of whatever capacity carries a base cost. By its simple existence it has harmed some amenity, and incurred an environmental cost. More generally, one can say that environmental invasion implies that the installation cost of a link is a concave rather than a linear function of its capacity. For SC models it turns out then that the capacity-optimised sum of installation and operating costs is itself now a concave rather than a convex function of the flow rate. Otherwise expressed, the effect is that a single high-capacity link over a route section is to be preferred to a bundle of low-capacity links. The criterion then favours trunking, and indeed has a dramatic effect; see Chapter 5. The optimal network is inclined to a tree pattern, with traffic from neighbouring source nodes converging together to what is indeed a trunk; this trunk then branching successively as it approaches a cluster of sink nodes.

Trunking can be seen as inducing a hierarchical structure, in which local traffic is carried locally; traffic to distant but clustered destinations is gathered together on to a trunk route, and there can be such gatherings at several levels. Such a structure can be seen at its clearest in telephone networks, in which there can be exchanges at several levels, and a call will be taken up to just that level of exchange from which there is a route down to the receiver. This hierarchical structure, differing radically from the array of direct connections derived on a naive criterion, is the pattern for optimal nets under an environmentally conscious criterion. In Chapter 5 we consider the optimal choice of exchange levels in an isotropic environment. It turns out that ideally the trunking rate should be constant up to a certain level, although the continuous gradation of trunking implied by this assertion is not practical. The investigation throws up some questions in what is essentially geometric probability, treated in Appendix 1.

Road networks (Chapter 6) also aim to be hierarchical, but labour under the difficulty that roads actually do take a great deal more physical space than do telephone links, for example. They are particularly constrained by the fact that they are largely confined to the two-dimensional plane, from which they can depart only at great expense. They suffer also from the statistical effect of congestion, and it is a moot point whether or not this is trunk-inducing. Congestion in queueing networks is relatively well understood; congestion on the continuum of a multi-lane highway is a much subtler matter, discussed briefly in Chapter 6.

The study of engineering structures in Chapters 7–9 takes us back to the 'naive' case, when the cost of building to a certain capacity is proportional to that capacity. However, solution for the optimal structure is now considerably less naive, because of the vector nature of the potential. The theory of Michell structures is not only one of the most beautiful examples of optimisation but, published in 1904, one of the earliest. It generalises to the SC case, as we demonstrate. Evolutionary methods really do make a computational appearance here, as the only way of performing the optimisation in all but the simplest cases. We review the striking results of Bendsøe and Sigmund in this context, also of Xie and Stevens.

However, natural evolutionary optimisation (on the scale of years) is evident in a continuum version of the problem: the formation of bone. Bone is continually subject to a general wasting, but also to reinforcement in those locations and directions where there is stress. The effect is then to achieve a structure optimal for the conditions. Load-bearing elements, such as the femur (thighbone), demonstrate not only a shape but also an anisotropic structure of the form that the Michell theory would predict. Moreover, they demonstrate their response to variability of load in the replacement of interior solid bone by cancellous (spongy) bone; see Chapter 9.

Stochastic variation plays a greater role in succeeding chapters. Part II is devoted to the topic of artificial neural networks (ANNs), whose study constitutes the most concerted attempt yet to both mimic and understand the logical performance of natural neural networks, and to perhaps approach it by adaptive rules. By comparison with the distributional nets of Part I, adaptability to variable loading is of the essence; it is required of the net that it should make an appropriate response to each of a great variety of significant inputs. In contrast to the models of Part I, the nodes of the network now incorporate a nonlinear response element, a partial analogue of the animal neuron, seemingly necessary for the degree of versatility demanded. Lagrangian methods survive in the technique of 'back propagation' for the iterative improvement of input/output response.

The ANN literature is now an enormous and diffuse one, and we treat only a few specific cases that make definite structural points. In a reversal of the historical view, many standard statistical procedures can be seen as a consequence of ANN evolution. In a back-reversal, statistical analysis demonstrates that a simple neuronal assembly (a McCulloch–Pitts net) is insufficient for the functions demanded; the net has also to be able to rescale signals and to choose between competing alternatives. To achieve a dynamic version of the Hamming net, in fact. When such a net is elaborated to deal with compound signals one begins to see structures identifiable (and in no facile sense) with natural neural systems.

Actual biological systems have to operate over a very wide range of input intensities, and absolute levels of signals mean very little. Such systems achieve internal communication by oscillatory patterns, and W. Freeman has clarified the basic neuronal oscillator by which these are generated and sensed. When one couples the oscillator with the normalising mechanism for signal strength one finds an explanation for 'neuronal bursting'; the regular bursts of oscillation which are so pronounced an observational feature.

The treatment of processing networks in Part III does not in fact have a large network content, as manufacturing requirements generally specify the sequence of operations to be performed fairly closely. There can be allocation problems, however, when congestion occurs and different job streams are competing for attention. The classic Jackson network, which serves so well under appropriate conditions, does not address this point, and it is now realised that the introduction of local precedence rules into queueing networks can do more harm than good to performance. We then leap straight from the exact and amenable treatment of the Jackson network to the exact and fairly amenable treatment of the Klimov index. This certainly solves the problem in principle when processing resources can be freely redeployed, and gives indications of how to proceed when they cannot. Appendix 2 gives some background treatment of the Klimov index and related matters.

When we come to communication networks in Part IV we mostly take for granted that networks will be hierarchical, and that the aim is to extract as much performance from the net as possible under varying load, which can take the form of both short-term statistical variation and longer-term secular variation. The treatment of loss networks owes a great deal to F. P. Kelly, who passed through this topic on his long trek through evolving technology and methodology. In Chapter 15 we consider the fluid limit, for which the relationships of the various shadow prices are evident and the optimal admission/routing policy easily derivable. These features survive in a first stochasticisation of the problem. Chapter 16 considers also the second stochasticisation: that in which one uses state feedback to regulate the system. As Kelly has shown, the ideal instrument for real-time control of admissions and routing is that of trunk reservation. By this, reaction to small-scale features (the numbers of free circuits) achieves effective control of the large-scale features (the make-up of the traffic that occupies the busy circuits).

However, control rules of this type have their limits when there may be sources of congestion deep in the system that are subject neither to direct observation nor to direct control. This is the case for the Internet, which also by its nature operates in a very decentralised fashion. Some account is given in Chapter 17 of the protocols which have proved so successful in controlling this remarkable giant organism, and of the more relaxed optimality concepts which have guided thinking.

While the state of the net for either the Internet or the Worldwide Web may be inaccessible to observation at a given time, the nature of the network that develops can be discerned, and has awoken much interest. One particular feature remarked is that the distribution of node degree (the number of immediate links made with a given Web page) follows an inverse power law. This is often spoken of as the network's having a 'scale-free' character (not to be confused with the scale seminvariance of costs defined earlier). Such a law could just be explained on the classical theory of random graphs, although not particularly naturally. A more plausible mechanism has been suggested by

Barabási and Albert: that of ‘preferential attachment’. However, any conjecture that the mechanism, whatever it is, might prove self-optimising is weakened by evidence that nets generated by a random mechanism show extremely poor performance by comparison with a properly designed hierarchical system. These matters are discussed in Chapter 18. The supplementary Appendix 3 describes some aspects of random graph theory of which people working in the Web context seem to be unaware.

I

Distributional networks

By ‘distributional networks’ we mean networks that carry the flow of some commodity or entity, using a routing rule that is intended to be effective and even optimal. The chapter titles give examples. Operational research abounds in such problems (see Ball *et al.*, 1995a,b and Bertsekas, 1998), but the versions we consider are both easier and harder than these. In operational research problems one is usually optimising flow or operations upon a given network, whereas we aim also to optimise the network itself. Even the Bertsekas text, although entitled *Network Optimization*, is concerned with operation rather than design. The design problem is more difficult, if for no other reason than that it cannot be settled until the operational rules are clarified. On the other hand, there may be a simplification, in that the optimal network is not arbitrary, but has special properties in its class.

The ‘flow’ may not be a material one – see the Michell structures of Chapters 7–9, for which the entity that is communicated through the network is stress. The communication networks of Part IV are also distributional networks, but ones that have their own particular dynamic and stochastic structure.

Cambridge University Press
978-0-521-87100-6 - Networks: Optimisation and Evolution
Peter Whittle
Excerpt
More information

1

Simple flows

1.1 The setting

By ‘simple flow’ we mean those cases in which a single divisible commodity is to be transferred from the *source nodes* of a network to the *sink nodes*, and the routing of this transfer is determined by some extremal principle. For example, one might be sending oil from production fields to refineries in different countries, and would wish to achieve this transfer at minimal cost. (For the ideas of this chapter to be applicable one would have to assume that all oil is the same – if different grades of crude oil are to be distinguished then the more general models of Chapter 3 would be needed.) This is the classical single-commodity *transportation problem* of operations research (see e.g. Luenberger, 1989). However, we mean to take it further: to optimise the network as well as the routing. Further, in Chapters 4 and 5 we consider the radical effect when the design must be optimised to cope with several alternative loading patterns (i.e. patterns of supply and demand) or with environmental pressures.

Another example would concern the flow of electrical current through a purely resistive network. This has virtually nothing to do with the practical distribution of electrical power, which is achieved by sophisticated alternating-current networks, but the model is a very natural one, having practical implications. For given input/output specifications the flow through the network is determined physically: by balance relations and by Ohm’s law. However, Ohm’s law can be seen as a consequence of a ‘minimal dissipation’ criterion, so that one again has an extremal principle, this time a natural physical principle rather than an imposed economic one.

This example is the simplest of a whole class of models, for which the flow is characterised by an extremal principle, and which lead to the classic and fruitful concept of complementary variational principles. Within this class we shall find a realistic subclass, that of seminvariantly scaled costs, for which design optimisation turns out to be particularly simple.

Yet another example we shall come to in Chapters 7–9 is that of optimal structural frameworks, for which the entity that is transferred is *stress*. This is the complete analogue of flow; it is transferred from the points at which load is applied through the framework to the *foundation*: the backstop that accepts all load.

The simple flow model is no longer adequate for the road networks of Chapter 6. This is because stochastic effects begin to make themselves felt, and also because there are several classes of traffic – classification by destination alone is already enough to change the situation radically. Other issues also arise as we develop the theme. If we are to optimise networks freely then we are forced to consider a much more general class of models: those allowing any pattern of flow on the continuum of physical space; see Chapter 2. We are also forced to recognise environmental constraints, which completely change the character of the optimal solution; see Chapter 5.

1.2 Flow optimisation

Denote the nodes of the network by j, taking values in the set $\{1, 2, \ldots, N\}$. Let f_j be the prescribed constant rate at which the commodity is supplied to node j from the external world, so that f_j is positive for source nodes and negative for sink nodes. Balance then requires that $\sum_j f_j = 0$, although this is a point we return to. Let x_{jk} be the rate of flow of the commodity from node j to node k. This can be nonzero only if there is a direct link between the nodes. If one thinks of the network as a graph, then the link would be termed an arc – we shall use the two terms interchangeably. In this chapter we shall assume the link to be undirected, in that flow can be in either direction, and x_{jk} can be of either sign. We must then adopt the convention that $x_{kj} = -x_{jk}$. When we come to road or communication traffic, then flows in opposite directions must be distinguished, and given separate directed links.

The flow x must obey the balance relation

$$\sum_k x_{jk} = f_j \tag{1.1}$$

at all nodes of the network. This is not in general sufficient to determine the flow, and so room is left for optimisation. Let us take as criterion that the flow is required to minimise the expression

$$C(x) = \sum_{j,k} c_{jk}(x_{jk}) \tag{1.2}$$

subject to the balance conditions (1.1). Here $c_{jk}(x_{jk})$ is to be regarded as the 'cost' of carrying flow x_{jk} along the jk link. We shall assume that direction is immaterial, so that the value of $c_{jk}(x_{jk})$ is unchanged if we reverse the direction of flow or the order of j and k. The symbol $\sum_{j,k}$ denotes a summation over the range $1 \leq j < k \leq N$, so that each undirected link is counted just once. If we really wished to sum over all ordered combinations of j and k we would sum with respect to the two variables separately.

The cost function $c_{jk}(x_{jk})$ can represent some economic criterion on which one bases optimisation of the flow, or it can arise as expression of a physical extremal principle; we shall see examples of both. Let us denote a specimen such cost function simply by $c(x)$. Then the basic properties we shall demand of this function are that it be convex and nondecreasing for positive x with $c(0) = 0$. The second and third assumptions are natural, but so is the first. It corresponds to the idea that the marginal cost of an increase in flow increases with flow. If the assumption fails, then one has a novel and significant physical phenomenon.

Let us denote the class of such functions by $\mathcal{C}$. There are occasions when it is useful to suppose c differentiable, with a one-to-one relationship between its argument x and its gradient $c'(x)$. Let us therefore define the class $\mathcal{C}_s$ of *strictly convex* cost functions, a subclass of $\mathcal{C}$ for which the differential of $c(x)$ exists and is strictly increasing for positive x, with

$$\lim_{x \downarrow 0} \frac{c(x)}{x} = 0, \qquad \lim_{x \uparrow +\infty} \frac{c(x)}{x} = +\infty. \tag{1.3}$$

If we consider directed links then we need consider only nonnegative x. If we consider undirected links, as we do for the moment, then we simply add the assumption of evenness: $c(x) = c(-x)$.

We have then to minimise the convex function (1.2) of flow pattern x subject to the linear constraints (1.1). This is the *primal problem*, which can be solved by Lagrangian methods in the strong form of convex programming (see e.g. Bertsekas *et al.*, 2003). Define the Lagrangian form

$$L(x, y) = \sum_{j,k} c_{jk}(x_{jk}) + \sum_j y_j(f_j - \sum_k x_{jk}), \tag{1.4}$$

where y_j is the Lagrangian multiplier associated with constraint (1.1). Denote the unconstrained minimum of L with respect to x by $D(y)$. The *dual problem* is that of finding the value of y that maximises $D(y)$. Denote this maximising value by $\bar{y}$. Then one key assertion is: that the value of x minimising the Lagrangian form $L(x, \bar{y})$ solves the primal problem.

The other key conclusion is the interpretation of $\bar{y}_j$ as a marginal price: the rate of change of minimal cost incurred with change in f_j. Let $M(f)$ be the minimal value of total transport cost $C(x)$ for prescribed f. Then one can loosely state that

$$\bar{y}_j = \frac{\partial M(f)}{\partial f_j}. \tag{1.5}$$

This is an identification which must be stated more exactly if it is to be true, and there are caveats when f is subject to constraints such as the balance constraint $\sum_j f_j = 0$. We cover these points in Section 1.7. In the meanwhile, it is helpful to be aware that (1.5) holds in some sense.

One advantage of the Lagrangian approach is the reduction in dimensionality. The vector y is of dimension N, the number of nodes, whereas the dimensionality of x equals the number of links, which could be as high as $N(N-1)/2$. However, the question is of course whether these calculations can be performed at all. A useful concept is that of the *Fenchel transform*

$$c^*(y) = \max_x [xy - c(x)] \tag{1.6}$$

of a cost function $c(x)$, a stronger (if more narrowly applicable) version of the classic Legendre transform. The square bracket can be regarded as the net profit one would make if one incurred a cost $c(x)$ by accepting flow x, but received a subsidy at rate y for doing so. Expression (1.6) is then the maximal profit one could make by choosing the value of x. Relation (1.6) defines a transform, converting a function c of flow to a function c^* of subsidy rate. If c is convex then so is c^*, and one has in fact that $c^{**} = c$; relation (1.6) still holds if one switches the roles of c and c^* (see e.g. Bertsekas *et al.*, 2003).

Note then the evaluation

$$D(y) = \sum_j y_j f_j - \sum_{j,k} c^*_{jk}(y_j - y_k). \tag{1.7}$$

The relevant value $\bar{y}$ of the multiplier vector y is that maximising this expression.

A case that allows very explicit treatment is that of electric current flowing through a network of resistors. For this the assumption is that

$$c_{jk}(x_{jk}) = \frac{1}{2} R_{jk} x_{jk}^2, \tag{1.8}$$

where R_{jk} is the resistance of link jk. Expression (1.8) is the rate of energy dissipation in the link, and we are appealing to the principle that the actual flow minimises dissipation. In this special case the Lagrangian form (1.4) is minimal with respect to x at

$$x_{jk} = \frac{y_j - y_k}{R_{jk}}. \tag{1.9}$$

Relation (1.9) expresses Ohm's law if y_j is interpreted as the potential (voltage) at node j. In fact, Ohm's law and the minimal dissipation principle are equivalent, in that each implies the other. One has still to determine the potentials y_j by appeal to the constraints (1.1) or by maximisation of the dual form (1.7), which now becomes

$$D(y) = \sum_j y_j f_j - \sum_{j,k} \frac{(y_j - y_k)^2}{2R_{jk}}.$$

As always, the case of quadratic costs and linear constraints holds a special place, both as being relatively amenable and as showing unexpected connections (see e.g. Doyle and Snell, 1984).

1.3 Seminvariantly scaled costs

When it comes to optimisation of the network, we would like to know how the cost $c_{jk}(x_{jk})$ of flow x_{jk} on link jk depends upon the physical size of the link. Let us suppose that the link has 'length' d_{jk} and 'cross-section' or 'rating' a_{jk}. We put these terms in quotation marks because they are not yet well-defined; interpretation is best kept elastic for the moment. Then we shall suppose that the cost function has the form

$$c_{jk}(x_{jk}) = a_{jk} d_{jk} \phi(x_{jk}/a_{jk}), \tag{1.10}$$

where ϕ is a convex function. The physical argument for this is that (dropping the jk subscript for the moment) $p = x/a$ is a flow density, so that $\phi(p)$ can be regarded as the cost per unit conductor volume of maintaining a flow density p. The factor ad is then the total volume of the link. The variable p is certainly meaningful; in the electrical context it is a a flow density, and it has equally meaningful roles in other contexts (see Section 1.6).