STUDIES IN
REGIONAL SCIENCE AND URBAN ECONOMICS 14

SPATIAL ECONOMICS: DENSITY, POTENTIAL, AND FLOW

M. J. BECKMANN
T. PUU

North-Holland

SPATIAL ECONOMICS: DENSITY, POTENTIAL, AND FLOW

Studies in Regional Science and Urban Economics

Series Editors

ÅKE E. ANDERSSON
WALTER ISARD
PETER NIJKAMP

Volume 14

NORTH-HOLLAND – AMSTERDAM • NEW YORK • OXFORD

Spatial Economics: Density, Potential, and Flow

MARTIN BECKMANN
Brown University, U.S.A.
and
Technische Universität,
München, F.R.G.

TÖNU PUU
University of Umeå, Sweden

1985

NORTH-HOLLAND – AMSTERDAM • NEW YORK • OXFORD

ISBN: 0 444 87771 1

Publisher:
ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

PRINTED IN THE NETHERLANDS

INTRODUCTION TO THE SERIES

Regional Science and Urban Economics are two interrelated fields of research that have developed very rapidly in the last three decades. The main theoretical foundation of these fields comes from economics but in recent years the interdisciplinary character has become more pronounced. The editors desire to have the interdisciplinary character of regional sciences as well as the development of spatial aspects of theoretical economics fully reflected in this book series. Material presented in this book series will fall in three different groups:

- interdisciplinary textbooks at the advanced level,
- monographs reflecting theoretical or applied work in spatial analysis,
- proceedings reflecting advancement of the frontiers of regional science and urban economics.

In order to ensure homogeneity in this interdisciplinary field, books published in this series will:

- be theoretically oriented, i.e. analyse problems with a large degree of generality,
- employ formal methods from mathematics, econometrics, operations research and related fields, and
- focus on immediate or potential uses for regional and urban forecasting, planning and policy.

Åke E. Andersson
Walter Isard
Peter Nijkamp

THE INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS

is a nongovernmental research institution, bringing together scientists from around the world to work on problems of common concern. Situated in Laxenburg, Austria, IIASA was founded in October 1972 by the academies of science and equivalent organizations of twelve countries. Its founders gave IIASA a unique position outside national, disciplinary, and institutional boundaries so that it might take the broadest possible view in pursuing its objectives:

To promote international cooperation in solving problems arising from social, economic, technological, and environmental change

To create a network of institutions in the national member organization countries and elsewhere for joint scientific research

To develop and formalize systems analysis and the sciences contributing to it, and promote the use of analytical techniques needed to evaluate and address complex problems

To inform policy makers and decision makers of how to apply the Institute's methods to such problems

The Institute now has national member organizations in the following countries:

Austria
The Austrian Academy of Sciences

Bulgaria
The National Committee for Applied Systems Analysis and Management

Canada
The Canadian Committee for IIASA

Czechoslovakia
The Committee for IIASA of the Czechoslovak Socialist Republic

Finland
The Finnish Committee for IIASA

France
The French Association for the Development of Systems Analysis

German Democratic Republic
The Academy of Sciences of the German Democratic Republic

Federal Republic of Germany
The Max Planck Society for the Advancement of Sciences

Hungary
The Hungarian Committee for Applied Systems Analysis

Italy
The National Research Council

Japan
The Japan Committee for IIASA

Netherlands
The Foundation IIASA–Netherlands

Poland
The Polish Academy of Sciences

Sweden
The Swedish Council for Planning and Coordination of Research

Union of Soviet Socialist Republics
The Academy of Sciences of the Union of Soviet Socialist Republics

United States of America
The American Academy of Arts and Sciences

TORD PALANDER, 1902–1972

IN MEMORIAM

Πάντὰ ρει

(All things flow – *Herakleitos*)

Acknowledgments

This book is the result of happy collaboration over almost a decade, and we gladly share the blame for the final product. In particular, however, Martin Beckmann is responsible for most of chapters 1, 2, 3, and 6 and Tönu Puu for chapters 4, 5, 7, and 8 as well as sections 2.1.8 and 3.4.3.

The research documented in this volume has benefited from the hospitality of several scientific institutes and from the sponsorship of one research financing fund. The authors wish to express their deep gratitude to all these institutions and the people working there.

The work started during our stay at IIASA, Laxenburg (Austria) in August and September 1979. Since then we have benefited from the creative atmosphere at IIASA during several further visits in 1982, 1983, and 1985. A most productive period was spent in a friendly atmosphere at SPUR, Louvain-la-Neuve (Belgium) in January 1982. In January 1983 our collaborative work was continued during Martin Beckmann's tenure as a visiting professor at Umeå University (Sweden).

Among those who have made a special effort to see this project through, our friend Åke E. Andersson stands out; we gratefully acknowledge his invaluable help on many occasions.

Last, but not least, the work of one of us (T.P.) as well as the production of his previous book in the same series, was financed by the Swedish Council for Research in the Humanities and Social Sciences. For this we wish to express our deep gratitude.

MARTIN BECKMANN TÖNU PUU

Contents

1 *Introduction*

1.1 THE VON THÜNEN AND WEBER SCHOOLS OF LOCATION THEORY

Throughout location theory and spatial economics there runs a dichotomy between what might be called the von Thünen and the Weber approaches to modeling economic space. The first considers economic activities that are extended in space and hence use land explicitly. The second considers that activities are localized at points in space but are spaceless in themselves (we will say more about this in Section 5.3). Both are justified in their proper context. The approach to spatial economics developed in this book is solidly in the von Thünen tradition: activities are space consuming. They are described by their spatial densities. The spatial relationships are made as explicit and graphic as possible. This is in contrast to recent developments in regional economics where the spatial structure has been suppressed and replaced by mere matrices of abstract distance.

By contrast, classical location theory has treated space in the same way as would ordinary intuition or Newtonian physics: as a continuum. Thus, in von Thünen's classic opening paragraph:

> "Consider a very large town in the center of a fertile plain which does not contain any navigable rivers or canals. The soil of the plain is assumed to be of uniform fertility which allows cultivation everywhere. At a great distance the plain ends in uncultivated wilderness by which this state is absolutely cut off from the rest of the world. This plain is assumed to contain no other cities but the central town and in this all manufacturing products must be produced. The city depends entirely on the surrounding country for its supply of agricultural products . . . The question is now, how under these circumstances agriculture will develop and how distance from the city affects agricultural methods when these are chosen in the most rational manner?" (von Thünen 1921, pp. 11–12, our translation)

The "homogeneous plain" in von Thünen's model is the plane of Euclidean geometry. This approach to space can also be found in the work of Launhardt (1869, 1882, 1885), Palander (1935), and Lösch (1954).

It was *linear programming* and its extensions, nonlinear and integer programming, that made fashionable, nay required, a discrete realization of space as a set of points whose interrelations are defined by distances or transportation cost matrices. Henceforth no geometric realization of space was required, all operations being performed on systems of equations and inequalities into which space entered only through measures of distance. The images of lines of movement, boundaries, areas, lines of equal price or "isotimes," or of "transportation surfaces" (Palander 1935) evaporated, to be replaced by activity variables and dual variables or efficiency prices.

While the results of these calculations usually permit a spatial interpretation and may be transformed back into spatial images, this is rarely done. The integration of spatial economics into the matrix of economic theory has been achieved by transforming away the spatial variables.

1.2 THE METRIC OF ECONOMIC SPACE

1.2.1 The Least-Cost Principle

What distinguishes location theory and spatial economics generally from the remainder of economic theory is the *explicit recognition of distance in the form of transportation cost*. Transportation cost should be interpreted broadly as a cost required to move

- Persons
- Commodities
- Information

all of which are generally referred to here as objects.

The spatial structure of transportation cost gives rise to a metric (in fact, to alternative metrics) based on the following principle. The economic distance $d(\mathrm{A}, \mathrm{B})$ from point A to point B is defined as the least cost of moving an object from A to B. It defines a metric since

$d(\mathrm{A}, \mathrm{B}) > 0$	if $\mathrm{A} \neq \mathrm{B}$	transportation cost between distinct points is positive
$d(\mathrm{A}, \mathrm{A}) = 0$		transportation cost between identical points is zero
$d(\mathrm{A}, \mathrm{C}) \leqslant d(\mathrm{A}, \mathrm{B}) + \mathrm{d}(\mathrm{B}, \mathrm{C})$		the triangle inequality holds by definition of "least cost"

If the transportation cost from A to B is not identically equal to the transportation cost from B to A, then there are in fact two metrics, based respectively on transportation cost "to" and transportation cost "from." In the case of personnel, the relevant transportation cost is usually the round-trip cost, defining yet another metric. Moreover, transportation cost may depend on the state of the transportation system, which, in turn, may vary systematically and periodically over time. This multiplicity of metrics is a complicating feature that sometimes cannot be ignored in locational analysis.

Notice that the least-cost principle in transportation always defines a metric in economic space, regardless of whether its representation is in continuous or discrete terms.

In the following sections we give some illustrations of the least-cost principle.

1.2.2 Metric of a Transportation Network

Consider first a discrete transportation network. It may be described abstractly by its matrix of arcs (edges) ij whose lengths are a_{ij}

$$A = ((a_{ij}))$$

with

$$\begin{aligned} a_{ii} &= 0 \\ a_{ij} &> 0 \qquad i \neq j \qquad \text{if a directed arc exists from } i \text{ to } j \\ a_{ij} &= \infty \qquad\qquad\quad \text{if no directed arc exists from } i \text{ to } j \end{aligned} \tag{1}$$

For the arc lengths we require neither symmetry

$$a_{ij} = a_{ji} \tag{2}$$

nor the triangle inequality

$$a_{ij} + a_{jk} \geqslant a_{ik} \tag{3}$$

To obtain the distances d_{ik} from any point of origin i to any point of destination k one must find the shortest path. Such a path always exists if the network is strongly connected. To obtain this path one may proceed iteratively as follows. Let d_{ik}^n denote the distance obtained in the nth round. If no path from i to k has been found then

$$d_{ik}^n = \infty$$

In the $(n + 1)$th round one calculates

$$d_{ik}^{n+1} = \min_j \,(a_{ij} + d_{jk}^n)$$

In a finite, strongly connected graph the d_{ik}^n, which decrease monotonically in n, converge to limits d_{ik} in finitely many steps. Thus the metric d_{ik} is the unique solution of the equation system

$$\begin{aligned} d_{ik} &= \min_j \,(a_{ij} + d_{jk}) \\ d_{ii} &= 0 \end{aligned} \tag{4}$$

Keeping the origin fixed (say) one obtains a tree of shortest paths from that particular origin, and for each vertex of the network a number, the distance of this vertex from the chosen point of origin. Similarly, keeping a destination fixed one obtains a tree of shortest paths to this destination. On this tree one can mark off, for each vertex, the remaining distance to the fixed destination. These trees coincide in the symmetric case (2).

1.2.3 The Continuous Plane

To extend this to the two-dimensional continuous plane we consider first the case that is analogous to the symmetric case in a network. Let the spatial coordinates be

$\underline{x} = (x_1, x_2)$ and let local transportation cost depend only on location but not on direction

$$k = k(\underline{x}) = k(x_1, x_2)$$

The distance between two points $\underline{x}^0 = (x_1^0, x_2^0)$ and $\underline{x} = (x_1, x_2)$ is then obtained by finding a path that minimizes the transportation cost integral, a "shortest path"

$$\int_{\underline{x}^\circ}^{\underline{x}} k(x_1(s), x_2(s))\,\mathrm{d}s$$

Keeping the origin fixed one now obtains two sets of curves:

- Trajectories of shortest paths, i.e. of least cost movement, and
- Curves connecting points of equal distance from the given point of origin.

In classical location theory (Palander 1935) these two sets of curves are known respectively as

- Transport lines, and
- Isovectures.

Formally these may be obtained as follows:

Let $\lambda(\underline{x})$ denote the value of minimum transportation cost to $\underline{x}$ from a given origin $\underline{x}^0$

$$\lambda(\underline{x}) = \min_{\underline{x}(s)} \int_{\underline{x}^\circ}^{\underline{x}} k(x_1(s), x_2(s))\,\mathrm{d}s \tag{5}$$

or "economic distance" of $\underline{x}$ from $\underline{x}^0$; and let $\underline{x} = \underline{x}(s)$ be the shortest path. Now write

$$\frac{\mathrm{d}\underline{x}}{\mathrm{d}s} = \left(\frac{\mathrm{d}x_1}{\mathrm{d}s}, \frac{\mathrm{d}x_2}{\mathrm{d}s}\right) = (\dot{x}_1, \dot{x}_2) = \dot{\underline{x}}$$

Then

$$\operatorname{grad} \lambda = k \cdot \frac{\dot{\underline{x}}}{|\dot{\underline{x}}|} \tag{6}$$

Here the vector $\dot{\underline{x}}/|\dot{\underline{x}}|$ has unit length. If ϕ denotes a flow vector of arbitrary (but positive) strength that traces out a shortest path, then (6) assumes the form

$$\operatorname{grad} \lambda = k\,\frac{\phi}{|\phi|} \tag{7}$$

This is the so-called gradient law. It plays a major part in the analysis of efficiency and equilibrium in spatial markets (cf. Chapter 2).

The meaning of (6) and (7) is that shortest paths and isovectures are orthogonal to each other. This is always true when transportation cost does not depend on direction, in other words when it is "isotropic" (for the anisotropic case cf. Section 2.4). It does not matter that transportation cost $k(\underline{x})$ depends on location.

It follows that one set of curves determines the other: the isovectures may be obtained from the shortest paths that originate in $\underline{x}^0$ by marking off equal distances or integrating along these paths. Conversely, the transport lines may be constructed as orthogonal trajectories to the given isovectures. As a by-product, the local rate of transportation cost $k(x)$ is obtained as the absolute value of the gradient

$$k(\underline{x}) = |\text{grad}\, \lambda(\underline{x})|$$

1.2.4 Euclidean Metric

The simplest metric is clearly that where $k(x)$ is uniform. It may then be set equal to unity

$$k(\underline{x}) \equiv 1$$

This is none other than the familiar Euclidean metric that underlies the "naive" or common intuitive notion of two- and three-dimensional space. In this case the shortest paths are straight lines and the isovectures are circles. The transportation cost may be integrated in closed form

$$\lambda(\underline{x}) = \int_{\underline{x}^0}^{\underline{x}} ds = \int_{\underline{x}^0}^{\underline{x}} (dx_1^2 + dx_2^2)^{1/2} = [(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2]^{1/2}$$

1.2.5 Refraction

The next case is well-known in theoretical optics. It was discussed in an economic context by Palander (1935), von Stackelberg (1938), and Lösch (1940).

Let there be a discontinuous change across a line from one level of transportation cost k_0 to another level k_1. An example is the crossing of a national boundary. Then the transport lines exhibit refraction, as shown in Figure 1.1.

Another special case is that where the actual transportation cost is a monotone transform of the integral in (5)

$$\begin{aligned} \lambda(x) &= m\,[\min_{\underline{x}(s)} \int_{\underline{x}^0}^{x} k(\underline{x}(s))\,ds] \\ &= \min_{\underline{x}(s)} m \left[\int_{\underline{x}_0}^{x} k(\underline{x}(s))\,ds \right] \end{aligned}$$

The isovectures are the same, but their labeling changes. It follows that the trajectories are also the same. The following is an example of this.

1.2.6 Alternative Modes of Travel

In personal transportation, time is often the overriding factor. Consider the minimum time required to cover various Euclidean distances, by foot, by bus, or by airplane.

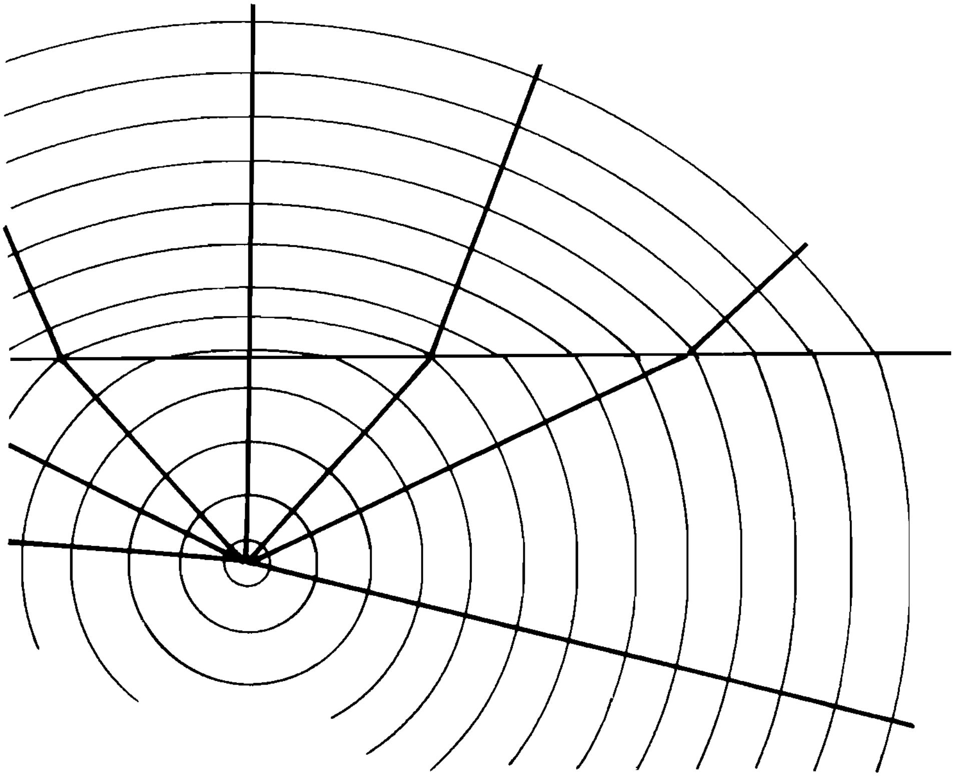

Figure 1.1. Refraction of flow lines.

Each mode is optimal in a certain range of distances (see Figure 1.2). The same principle applies to the transportation costs of goods when fixed costs of loading are involved. From the figure one can see that economic distance is a monotone, increasing, concave function of Euclidean distance when alternative modes of transportation are available. Recall that the isovectures and trajectories are the same under either Euclidean or economic distance; only the numbers denoting the distances change.

1.2.7 Realization of an Isotropic Metric in Two-Dimensional Space

The type of metric described in Section 1.2.4 is realized at its simplest on essentially two-dimensional surfaces, for example an ocean with shipping or a snow-covered plain on which dog-team transportation is used. It is also the appropriate metric for planning a future transportation route.

Consider now the more common case of transportation within a network that is already in place. Most have been planned along regular lines (triangular, rectangular, hexagonal networks), which makes transportation in networks anisotropic. But in practice the presence of many small deviations distorts this ideal plan. The result is once more

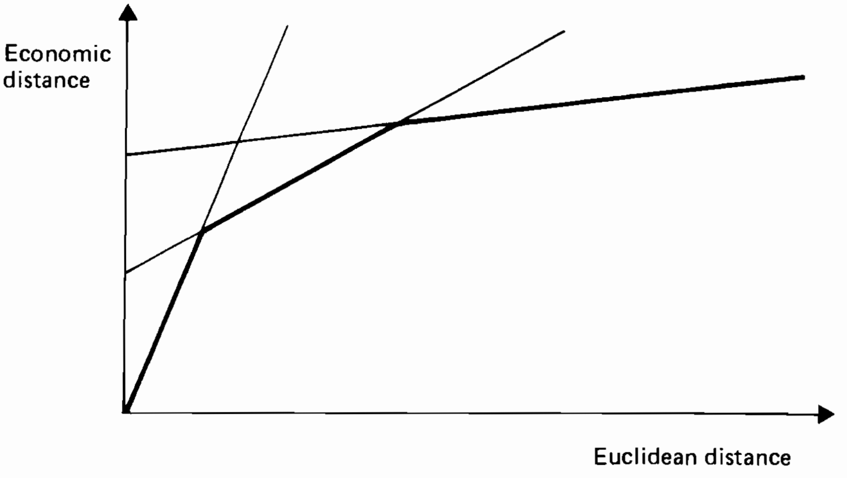

Figure 1.2. Economic versus Euclidean distance.

an almost isotropic metric, at least over large distances (cf. Puu 1979b). This serves to justify, at least to a first approximation, our treatment of transportation cost metrics in two-dimensional space as isotropic. They are not assumed to be uniform, however, since local conditions may vary both in regard to the supply of transportation (density and capacity of the network) and demand (traffic load).

We therefore postpone for the moment discussion of those cases in which local transportation cost is dependent on direction. The relationship between the directions of transport lines and isovectures in such cases is more complicated and is governed by so-called transversality conditions. Some typical cases, including

- Manhattan metric,
- Hexagonal isovectures,
- Minkowski metric, and
- Riemann metric

are discussed later in Section 2.4.2.

1.3 THE BACKGROUND TO THIS STUDY

The two-dimensional, continuous spatial framework of classical location theory in the von Thünen tradition implied that the underlying commodity shipments should constitute a field of flows. However, this was not brought out explicitly. The flow field notion came into economics from a different side: namely, through the study of the continuous analog to the linear programming problem of minimizing the cost of carrying out a given transportation program. This was first done by Kantorovich (1942) in his classic paper "On the Translocation of Masses."

Martin Beckmann, unaware at that time of the Kantorovich paper, but having studied

Koopmans' classical article on the "Optimum Utilization of the Transportation System" (Koopmans 1949), wrote a Cowles Commission discussion paper soon after arriving at the University of Chicago in October 1950, entitled "A Formal Approach to Localization Theory." He then studied its implications for the analysis of spatial market equilibrium in subsequent discussion papers. After substantial revisions, Beckmann's "A Continuous Model of Transportation" was published in *Econometrica* in 1952. An expository version emphasizing the spatial market interpretation appeared in the *Weltwirtschaftliches Archiv* in 1953.

During the Econometric Society Meeting in Uppsala in 1954, Beckmann had the opportunity to discuss this approach with Tord Palander, still the leading location theorist at that time, even though Palander's own interests had shifted somewhat by then. It was through this contact that Tönu Puu became aware of the continuous flow approach to spatial economics and was stimulated to develop it further. In turn, Puu's lively interest rekindled Beckmann's concern with the continuous flow model. Thereafter Puu developed the structural stability approach in continuous flow analysis.

Besides the work reported in this monograph, there has been considerable interest in the continuous modeling of space by theoretical geographers. In this connection, the names of Waldo Tobler and Leslie Curry are perhaps of special significance. A short bibliography can be found at the end of the book.

1.4 THE PLAN OF THIS BOOK

The continuous flow approach is presented first for the problem of spatial equilibrium in a single-commodity market. This paradigm is fully developed in Chapter 2. It brings into sharp focus the differences between economic and physical flow fields. Although economic relationships must manifest themselves in physical terms, the driving forces turn out to be quite different. Profit maximization is unlike pressure, gravity, or other simply-structured physical forces. It is only when we turn to spatial interactions of other types that the analogy to physical laws becomes closer, as demonstrated in Section 3.4 (Dynamics) of Chapter 3. In Chapter 3 we move to applications of the continuous commodity flow model. The first section examines spatial pricing under competitive and monopolistic conditions, and we then turn to land use. Both of these applications involve partial equilibrium models in which the commodity or group of commodities under consideration is only a subset of all commodities. Land use in general equilibrium is studied next and a new interpretation is given to the classical von Thünen model.

Chapter 4 outlines a long-term model of spatial economic equilibrium. The model is of general equilibrium character as the decisions of firms and households are analyzed. There are two interdependent markets: the labor market and the market for one produced commodity. This commodity is assumed to be perfectly malleable: it satisfies all consumption needs, including housing, it can be used to provide transportation services, and it can even be invested as capital.

The decisions of the firms and households concern not only the quantities of labor and the commodity currently demanded and supplied. They also involve such long-term issues as optimum capital stock and optimum population size. Trade emerges in the model despite the fact that there is only one produced commodity. This is possible because land and the located enterprises may be the property of households other than those living at the specified location. There thus arises an interregional transfer of dividends (profits = land rent), which upsets Walras' law locally.

The main contribution of this chapter is the use of topological dynamics to characterize the qualitative features of flows and spatial organization under the assumption of structural stability. It is shown that the stable configurations of spatial economic organization are quadratic, and that, in particular, the hexagonal configurations of Christaller and Lösch become structurally unstable once the homogeneous space (where all communication is along straight lines) is abandoned.

Chapter 5 then deals with models in planning format. However, the social welfare function only occurs in one optimality condition in each model, whereas the rest (optimality conditions for production and trade) are of Paretian character. The latter have obvious interpretations in terms of individual optimization by producers and transportation enterprises. It is even demonstrated that any social optimum can be obtained with consumer autonomy, provided an appropriate interregional income-transfer policy is designed. This is shown by a set of aggregate identities among the various monetary expenditures.

The main outcome of this chapter is a theorem on specialization and trade. This is nothing more than the spatial subdivision in the von Thünen rings, adjusted to our more general setting. We thus recall the von Thünen theory for trade and specialization, which does not depend on comparative advantages or trapped resources. On the contrary, we assume all land to be of homogeneous quality and all the other inputs to be perfectly mobile. (To begin with, we even disregard the initial relocation costs for capital and labor, assuming that we are dealing with very long-term phenomena. Later, we introduce the relocation costs explicitly.)

The planning format is also used for a comparison of the von Thünen and Weber location principles (Section 5.3). It is seen that von Thünen's principle ("What to produce at each location?") ensures optimal use of resources, whereas Weber's principle ("Where to locate each production process?") results in a strictly weaker condition that does not ensure optimal use of resources.

Chapter 6 reconsiders some classical location problems in the light of the continuous flow approach. These include not only labor, resource, and market orientation but once again a restatement of the classical Weber problem. Some recent developments are sketched out, such as exhaustible resources, indivisibilities and increasing returns, and commuting and urban structure.

In Chapter 7 we leave those trade models that deal with a finite set of flow fields, and discuss interaction models. With continuously dispersed populations and an interaction principle according to which each location must communicate with all other locations, we deal with a (nondenumerable) infinity of vector fields, from which we somehow have to derive a measure of traffic. Along with traffic we consider congestion in transportation, crowding in housing, optimal urban land use, and the optimal distribution of population between center and periphery.

Chapter 8 examines business-cycle and growth models built on the multiplier and accelerator principles, augmented by an interregional trade multiplier, and adjusted to continuous two-dimensional space. It is seen that this procedure gives rise to "wave" or "heat-diffusion" type equations. The main results are that even the simplest models, with only local action in space and time, generate irregular cycles and variable periodicity. In spaceless models this only happens with complex distributed lag systems, or when nonlinearities are present.

Chapter 9 recapitulates the main conclusions reached and the book closes with an appendix and a short bibliography.

2 *The Continuous Transportation Model*

2.1 ECONOMIC THEORY

2.1.1 Introduction

Figures 2.1–2.6 show examples of commodity markets extending over large areas. Prices vary between locations, but in an orderly way. Commodity shipments are not shown but may be inferred: presumably shipments occur from low-price areas to high-price areas. But can an exact relationship be established between commodity movement and the price distribution? To be precise, suppose the quantities produced and the quantities consumed of a commodity are given for every location. Can the equilibrium prices, the exports and imports, and the size and direction of the shipments be determined as solutions to equilibrium conditions in a competitive but spatially extended market?

One way to attack this problem is to divide the spatial market into discrete units, for each of which supply and demand are specified, and to consider the flows between adjacent units. This discretization ignores, however, the spatial arrangement, unless one imposes a net of cells defined by a coordinate system (see Section 2.5.4 below). Flows are then broken down into an East–West (or horizontal) and a North–South (or vertical) component, and two coordinates are assigned to each variable.

But having gone this far, it is both natural and convenient to choose a representation in terms of continuous spatial coordinates, for example x_1 and x_2. In this way the spatial arrangement of the problem is preserved while the advantages of continuous analysis can be utilized at the same time.

2.1.2 The Divergence Law

To begin with, assume then that a commodity's supply and demand are given for each location regardless of price. This does not mean that they are independent of price but rather that the current prices are already substituted in the supply and demand function. Supply and demand are given here in terms of areal densities, i.e. physical quantities of the commodity per unit area.

Among those commodities that are produced and consumed over widely dispersed areas are the following: agricultural products, notably vegetable foodstuffs, and cattle, game, and fish, as well as materials requiring further processing such as wool, cotton, and

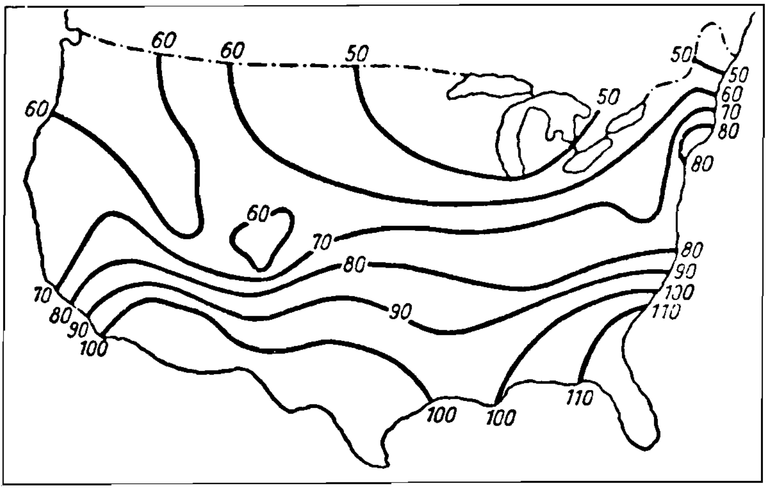

Figure 2.1. Producer prices for potatoes in the United States, in cents per bushel on December 1, average for 1906–15. (From H. Working, *Factors Determining the Price of Potatoes in St Paul and Minneapolis*, Minneapolis, 1922.)

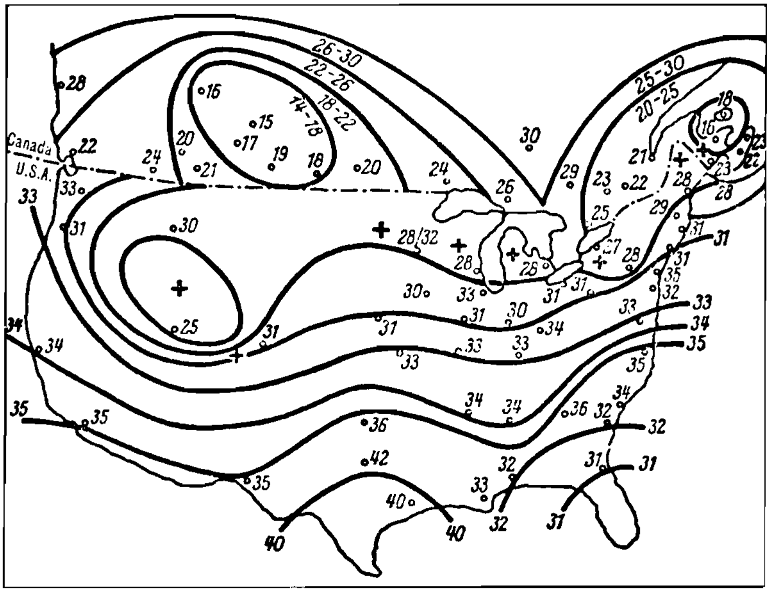

Figure 2.2. Retail prices of potatoes in the United States and Canada, in cents per 10 pounds, 1936. + denotes surplus regions. (Sources: Dominion Bureau of Statistics, Canada; Bureau of Labor Statistics, Canada; Bureau of Labor Statistics, United States.)

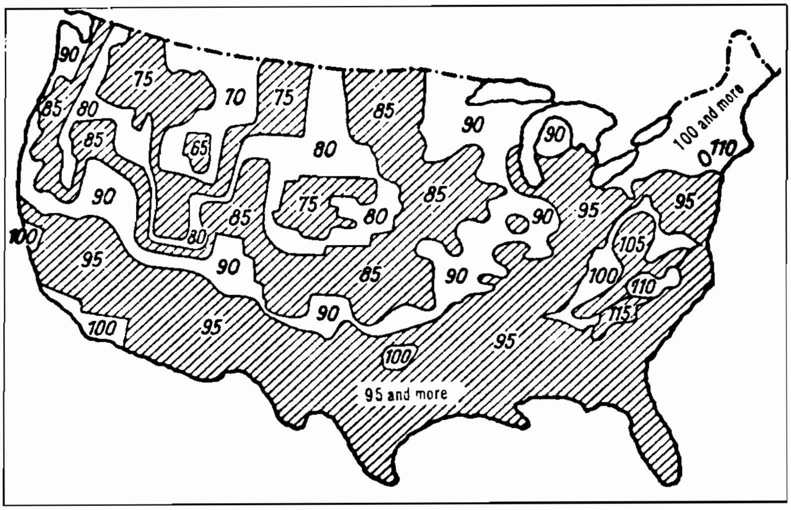

Figure 2.3. Spatial pattern of producer prices for wheat in the United States, in cents per bushel, 1910–24. (From F.A. Fetter, *The Masquerade of Monopoly*, p. 295; after L.B. Zapoleon, *Geography of Wheat Prices*, US Department of Agriculture Bulletin No. 594, Washington, DC, 1918.)

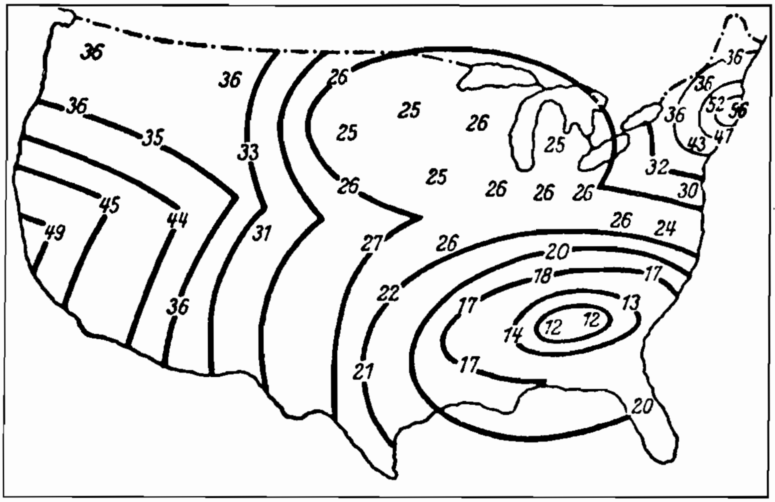

Figure 2.4. Monthly wage, without board, of agricultural workers in the United States, in dollars, 1933. (Source: Ch. Roos, *NRA Economic Planning*, Bloomington, IN, 1937, p. 161.)

other fibrous materials. Another important raw material of this type is wood. Slightly more localized, but still occurring over extensive areas, are various mineral resources, including clay for brickmaking and limestone for the production of cement, mineral raw materials and ores for metal production, and of course the major energy raw materials, coal and oil.

Another resource that is available extensively and used everywhere is human labor.

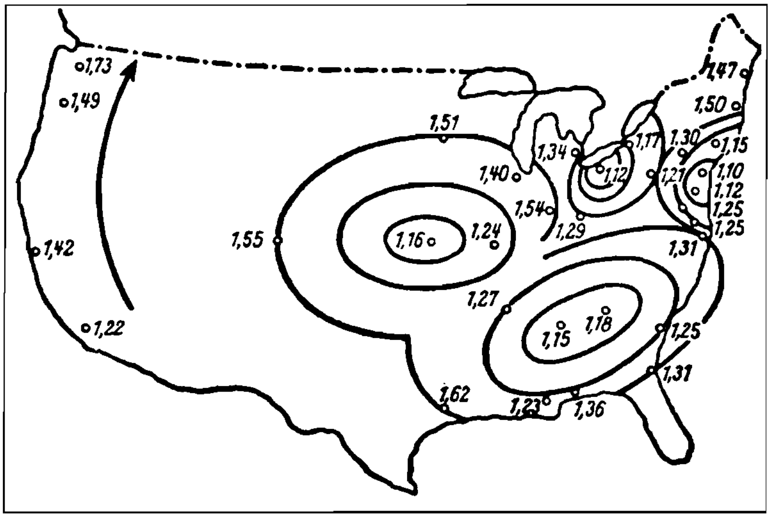

Figure 2.5. Price for resoling and heeling a pair of shoes in the United States, in dollars, 1936.

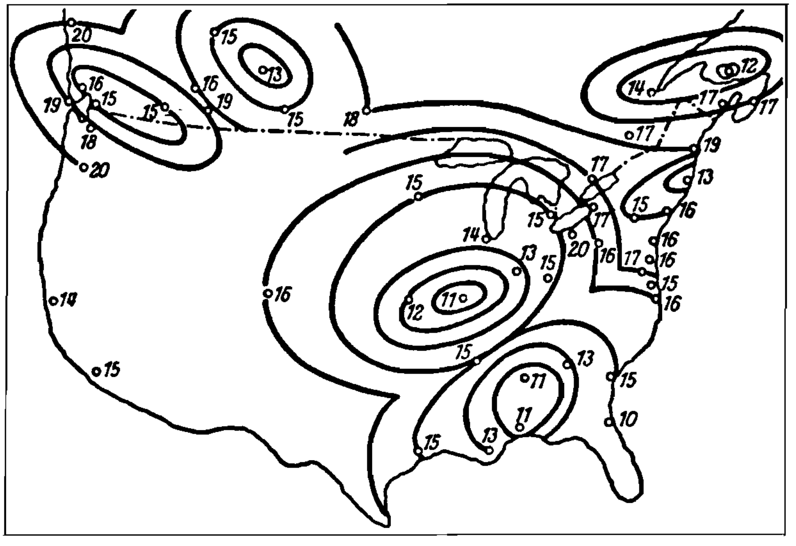

Figure 2.6. Price for laundering a man's shirt in the United States and Canada, in cents, 1936.

In the short term the interlocal flow of labor takes the form of commuting, whereas in the long run it appears as migration.

Finally we must consider money. The excess demand for cash varies between locations. To balance supply and demand requires the transportation of cash, sometimes in significant quantities.

For a spatial market within a closed region to be capable of attaining equilibrium, a necessary and sufficient condition is that aggregate supply and demand must balance. It is sufficient here to consider excess demand q, the difference between demand and supply, rather than dealing with demand and supply separately. Excess demand density is treated here as a given function of location

$$q = q(x_1, x_2)$$

The condition for an equilibrium of the spatial market to exist in a closed region A is that

$$\iint_A q(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 = 0 \tag{1}$$

Except for the trivial case of no interlocal trade, when supply and demand are balanced locally everywhere so that

$$q(x_1, x_2) \equiv 0$$

condition (1) implies that if there are places of surplus with negative excess demand then there must also be places of deficit with positive excess demand. The movement of the commodity in interlocal trade will be, in general, in directions from points of excess supply to points of excess demand. This does not preclude, however, the possibility that the commodity may pass through some points of excess supply or excess demand during trans-shipment.

The commodity movement in interlocal trade will be described by a continuous flow field. At each point (with the exceptions noted below) there is a well defined direction in which the commodity is moved through trade and a volume corresponding to each commodity flow. These together define the local flow vector $\phi(x_1, x_2)$. In areas without trade – where there is neither production nor consumption of the good or where supply and demand are locally balanced and there is no throughflow – the flow field vanishes. At points that are either isolated (centers of supply areas or of market areas) or that extend along lines (e.g. boundaries between supply and market areas) more than one direction of the flow field exists. These singularities occupy an area of measure zero and can be disregarded in all integrals measuring economic costs and/or benefits.

The relationship between the flow fields and the local excess supply corresponds to that between a fluid flow and its sources and sinks and has thus been thoroughly explored in hydrodynamics and thermodynamics. This relationship has the well-known form

$$\begin{aligned} -q(x_1, x_2) &= \operatorname{div} \phi(x_1, x_2) \\ &= \frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2} \end{aligned} \tag{2}$$

where

$$\phi = (\phi_1, \phi_2)$$

is the flow vector. The derivation of (2) is shown in the Appendix.

The assumption that no flow should cross the boundary of an area A has similarly been shown in hydrodynamics to have the form

$$\phi_n = 0 \qquad \text{in } \partial A \tag{3}$$

where n denotes the direction normal to the boundary and pointing in an outward direction; ϕ_n is the vector component in that direction and ∂A denotes the boundary of A. This condition is stronger than the statement that aggregate net exports be zero. It states that there should be no exports or imports across the boundary anywhere.

2.1.3 The Gradient Law

The commodity balance equation (2) is a necessary condition for any spatial system in which the stock of commodity is preserved. It contains no economic meaning but represents a physical constraint. The operation of either economic forces in a competitive market economy or of an efficient planning mechanism must be stated by means of a different principle.

In a spatially extended commodity market under perfect competition resources are efficiently allocated. When supply and demand are given, this means that total transportation cost is minimized. This can be the result of either planning or of competitive market forces, as we shall see in Section 2.1.4.

As in the earlier Section 1.2.3, we let $k(\underline{x})$ denote the cost of transporting a unit of the commodity over unit distance at location x. Let $\lambda(\underline{x})$ denote the price of the commodity as a function of location. Now the gain from trading a unit of the commodity between two "adjacent locations" separated by a distance ds in direction ϕ is

$$D_\phi \, \lambda(\underline{x}) \cdot \mathrm{d}s$$

Here D_ϕ denotes the directional derivative in the direction ϕ. In equilibrium the gain from trade cannot be greater than the required cost $k\,\mathrm{d}s$ of transportation

$$D_\phi \lambda \,\mathrm{d}s \leqslant k \,\mathrm{d}s \tag{4}$$

and this must be true for all possible directions of flow ϕ. Now the direction of steepest increase of λ is the gradient direction, and the value of the directional derivative is then equal to the absolute value of the gradient. The equilibrium condition (4) may therefore be sharpened to

$$|\mathrm{grad}\, \lambda| \leqslant k \tag{5}$$

In order that trade should take place, traders must not suffer losses. This means that the gain from trade exactly equals transportation cost wherever $\phi \neq 0$. Thus

$$|\mathrm{grad}\, \lambda| = k \qquad \text{where } \phi \neq 0 \tag{5a}$$

The direction of trade that achieves a gain equal to transportation cost is then the gradient direction. Thus

$$\phi \parallel \operatorname{grad} \lambda \tag{6}$$

Combining (5a) and (6) one obtains

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \text{wherever } \phi \neq 0 \tag{7}$$

Equations (5) and (7) represent the conditions for price equilibrium in a spatially extended market. Similarly, equation (2) may be said to represent the equilibrium of quantity. Equations (2), (5), and (7) together constitute a complete set of equilibrium conditions in the interior of the region A. On the boundary ∂A only the quantity condition (3) is required.

2.1.4 Perfect Planning

The object of planning is considered here to be the following: a given excess demand $q(\underline{x})$ must be satisfied for the minimum transportation cost. How this transportation cost is actually measured depends on the situation at hand. Sometimes it is measured in commodity units (as in the original von Thünen model) but it is more usually measured in money terms. In a general equilibrium context, transportation cost should be considered an output produced from suitable inputs such as land, labor, and capital (cf. Section 4.1.3 below).

With the previous notation the problem may be stated mathematically as follows

$$\min_{\phi} \iint k|\phi| \, \mathrm{d}x_1 \, \mathrm{d}x_2 = K \tag{8}$$

subject to

$$\operatorname{div} \phi + q = 0 \qquad \text{in } A \tag{2}$$

$$\phi_n = 0 \qquad \text{in } \partial A \tag{3}$$

This is a calculus of variations problem in terms of a vector function ϕ. The objective function of this problem can easily be shown to be convex and the constraints to be linear (see Section 2.2.3). It may also be shown that condition (1) guarantees the existence of a feasible solution (see Section 2.2.1). Moreover the minimand can be shown to be bounded for a bounded region A in two-space. The existence of a piecewise-smooth flow field minimizing (8) follows then from general principles of the calculus of variations (Courant and Hilbert 1953).

The form of the solution may be obtained by using Lagrangean multipliers. The Lagrange function of the problem (considered as a maximum problem) is

$$\iint L \, \mathrm{d}x_1 \mathrm{d}x_2 = \iint -k|\phi| - \lambda[q + \operatorname{div} \phi] \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{9}$$

The Euler–Lagrange condition for a maximum is derived below (in Section 2.2.5) as

$$k \frac{\phi}{|\phi|} = \text{grad}\,\lambda \tag{7}$$

Thus we obtain once more the gradient law.

2.1.5 Discussion of Flow Lines

The fundamental relationship (7), together with constraints (2) and (3), determines uniquely the directions of a flow field ϕ that is a solution to our problem (cf. Section 2.2.6). The principal statement of (7) is that the flow lines of the field ϕ cannot form closed paths. Rather, flow $\phi(x_1, x_2)$ is proportional to a gradient field. The potential function $\lambda = \lambda(x_1, x_2)$ of this gradient field represents, of course, the price of the commodity as a function of location.

Condition (7) needs to be supplemented for the case when ϕ vanishes. By analogy with the situation in convex nonlinear programming, one can show that

$$|\text{grad}\,\lambda| \leqslant k \qquad \text{wherever } \phi = 0 \tag{5}$$

Like any potential function, λ is determined only up to an additive constant. If the flow field is connected, i.e. does not consist of two or more disconnected parts, then it follows that when the price has been fixed at one location, it is uniquely determined at all other locations. This also means that certain types of singularities, such as spirals (Figure 2.7) and centers (Figure 2.8), are ruled out.

The only types of point singularities possible in a gradient field are sources, sinks, i.e. nodes or foci (Figures 2.9, 2.10, 2.13), and saddle points (Figure 2.11). The sources are located at minima and the sinks at maxima of the potential function. At saddles the potential function itself has a minimax or saddle point.

A flow field without singularities is laminar (Figure 2.12).

Suppose that, just inside the boundary ∂A of the region, all flows are directed to the outside (i.e. the boundary). As one traverses a curve parallel to but inside the boundary the direction of the flow field changes by 2π. Now the Poincaré Index theorem (Henle 1979, p. 60) implies that the number of sources and sinks must exceed the number of saddles by one. Therefore there has to be at least one source or sink inside the region.

A similar count applies when all flow vectors near the boundary point toward the inside: there must be at least one source or sink inside.

2.1.6 Shortest Paths and Isovectures

For each location there exists a flow field of special interest. This field is generated by placing a source at this location – that is, a small circle producing a large excess supply immediately around the location – and by letting every other point have an excess demand density of unity i.e. $q = 1$. The flow lines of this field represent the shortest paths to all points from the given location. The potential lines are the curves

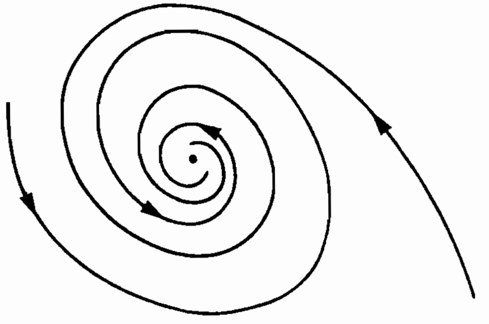

Figure 2.7. Stable spiral.

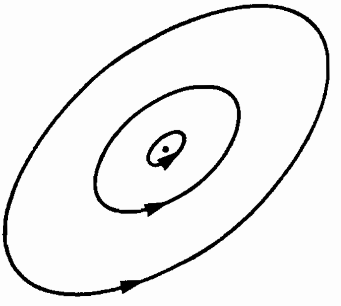

Figure 2.8. A center.

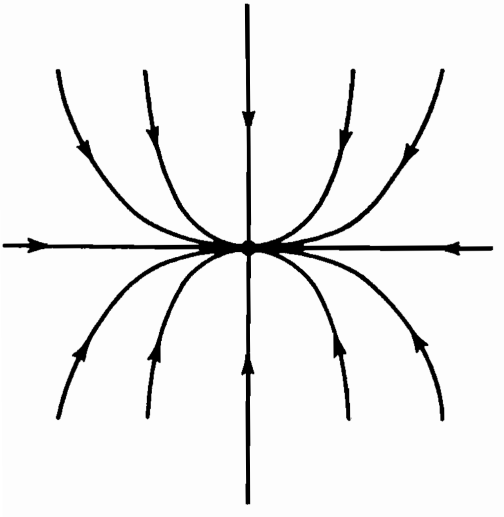

Figure 2.9. A node.

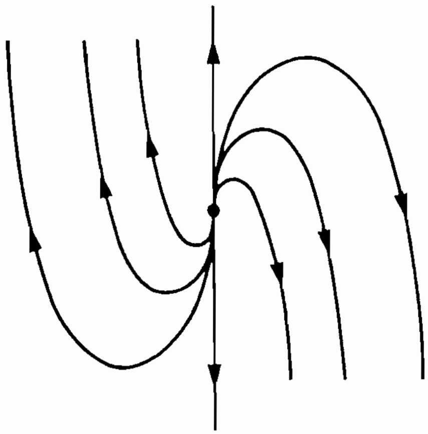

Figure 2.10. An improper node.

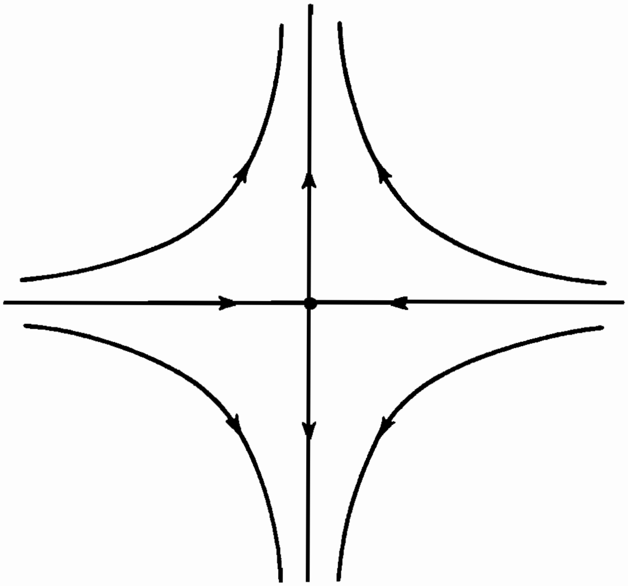

Figure 2.11. A saddle point.

of equal distance or *isovectures*. It is natural to choose $\lambda(\underline{x}_0) = 0$ at the location $\underline{x}_0$ so that $\lambda(\underline{x})$ measures the distance from $\underline{x}_0$ to $\underline{x}$. The isovectures are closed curves unless interrupted by the boundary of the region. The location is called a center of the region if one of its isovectures coincides with the boundary.

For a discussion of the anisotropic case see Section 2.4.2 below.

A special case is that where

$$k(\underline{x}) \equiv k \tag{10}$$

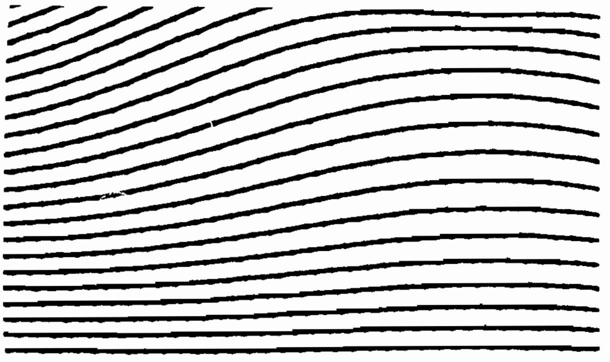

Figure 2.12. Laminar flow: no singularities.

Then all flow lines are straight lines and all isovectures are circles. From the theory of functions it is well known that there always exists a *conformal mapping* preserving angles but changing local scale, which transforms a plane in such a way that $k(\underline{x}) \equiv k = 1$ (say), and that this conformal mapping is achieved by an analytical function. Conversely, the flow fields and potential lines of a region with inhomogeneous $k(\underline{x})$ can always be thought of as being obtained from straight lines and circles through the appropriate conformal mapping (cf. Section 2.5.1).

2.1.7 Uniform Transportation Cost

Suppose

$$k(x_1, x_2) = k \tag{10}$$

is constant, i.e. independent of location. Then the flow lines are straight lines, as mentioned before. This follows from the fact that a flow line connecting a given source with a given sink must follow a shortest path or geodesic (see Section 2.1.6). The integral of transportation cost is then a constant times the ton-mileage, i.e., the integral of total flow

$$\int\int k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 = k\int\int|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

When transportation cost is uniform we can standardize our distance units as transportation cost units so that $k \equiv 1$.

The equilibrium conditions are now simplified as follows. We write

$$\phi = v\psi \tag{10}$$

where ψ is the unit vector giving the direction of ϕ and v is the flow volume. Then (2) becomes

$$\operatorname{div}(v\phi) + q = 0$$

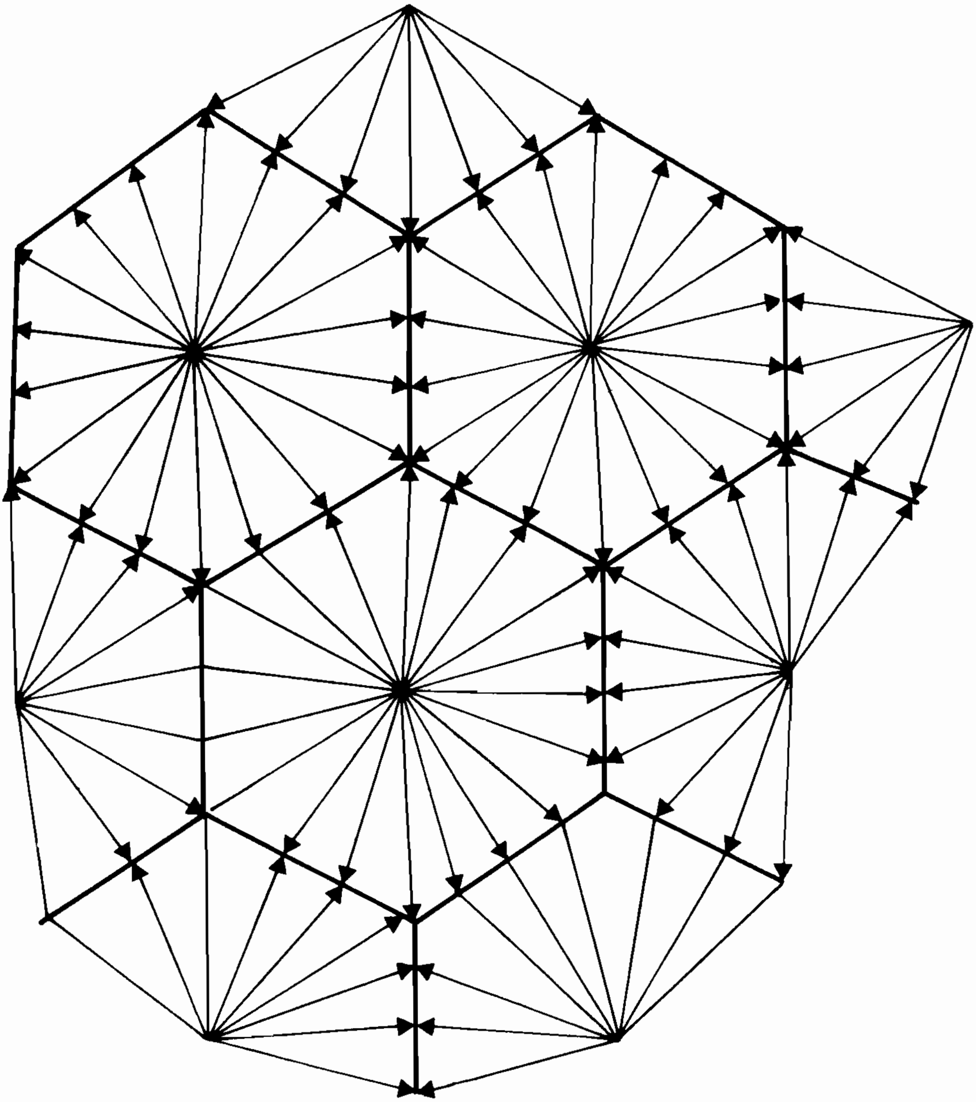

Figure 2.13. A flow field in a Löschian market system. (Each point singularity is a focus.)

Now by a well-known formula of vector analysis

$$\operatorname{div}(v\psi) = v \operatorname{div} \psi + \psi \cdot \operatorname{grad} v$$

Then the equilibrium conditions read

$$\operatorname{grad} \lambda = \psi \qquad |\psi| = 1 \tag{11}$$

and

$$v \operatorname{div} \psi + \psi \cdot \operatorname{grad} v + q = 0 \tag{12}$$

When the price λ is given, ψ is determined by (11) and v is determined by (12). A field of flow directions ψ can thus be associated with different flow volumes v depending on the spatial distribution of excess demand q. Of course along a given flow line of ψ, supply and demand must be balanced (see Section 2.3.5 below).

2.1.8 An Alternative Expression for Transportation Cost

We have seen that a necessary condition for

$$\iint k|\phi| \, dx_1 \, dx_2 \tag{8}$$

to be minimized, subject to

$$\operatorname{div} \phi = -q \tag{2}$$

is the Euler equation

$$k\phi/|\phi| = \operatorname{grad} \lambda \tag{7}$$

Let us now see what we get by taking the line integral of (7) along a flow line. If an arc length element is written ds, then we obtain from equation (7), by multiplication by the unit flow direction vector $\phi/|\phi|$ and by ds

$$\int k \, ds = \int d\lambda \tag{13}$$

Because the right-hand side is an exact differential, we can interpret $\lambda = \int k \, ds$ as equal to transportation cost if we set its value to zero at the point where the trajectory leaves or enters the region studied. If we set λ equal to production cost or to the import price at the boundary, then it represents the local cost obtained by increasing these prices by transportation cost. With this interpretation of λ in mind, let us study the expression

$$\iint \operatorname{div} (\lambda\phi) dx_1 \, dx_2 \tag{14}$$

Owing to Gauss's integral theorem, (14) equals the integral

$$\int \lambda(\phi)_n ds = X - M \tag{15}$$

taken along the boundary of the region. As λ is the "world market price" at the boundary and $(\phi)_n$ is the outward component of the flow normal to the boundary, equation (15) equals exports minus imports in value terms.

On the other hand, $\operatorname{div} (\lambda\phi) = (\operatorname{grad} \lambda)\phi + \lambda \operatorname{div} \phi$, from an elementary identity in vector analysis. Since $\operatorname{div} \phi = -q$ because of (2), we obtain

$$\int\int (\text{grad }\lambda)\phi \, dx_1 \, dx_2 - \int\int \lambda q \, dx_1 \, dx_2 = X - M$$

Finally, we have to interpret (grad λ)ϕ. Because of (7) it obviously equals $k|\phi|$. Therefore

$$\int\int k|\phi| \, dx_1 \, dx_2 - \int\int \lambda q \, dx_1 \, dx_2 = X - M \tag{16}$$

The first term on the left-hand side equals (8) and hence the total transportation cost. To interpret the second term we provisionally suppose that exports and imports balance, i.e. that $X = M$, and interpret λ as a pure transportation cost, so that it has zero value where a trajectory enters the region or issues from a singularity inside it. Then λ is the cost of transportation from this point of origin, and q is the excess demand at the point of destination. We can hence interpret λ as a cost of transportation in the usual sense for a displacement across a finite distance and q as the quantity of goods shipped across the same distance. The integral $\int\int \lambda q \, dx_1 \, dx_2$ hence really equals transportation cost. The equality

$$\int\int k|\phi| \, dx_1 \, dx_2 = \int\int \lambda q \, dx_1 \, dx_2 \tag{16a}$$

thus confirms that (8), obtained by consideration of infinitesimal displacements, is indeed a reasonable expression for transportation costs. If $X \neq M$, then the only difference is that the two cost expressions may deviate by a constant amount, equal to the trade balance. This is reasonable since λ cannot then be taken as zero where flows (not only transverse trajectories) enter or leave the region.

Equations like (16) and (16a) will be utilized in two additional ways in what follows. First, they can be used to establish duality. Second, they can be used to write down aggregate relations between local consumption and production values, transportation costs, exports and imports, etc.

For singularities we need an additional argument. If a singular point is included in the region, we cannot apply Gauss's theorem. What we do then is to surround the singular points by additional boundary curves and add integrals along them to the left-hand side of (15). It is convenient to take $\lambda =$ constant on such curves. By letting λ approach the value it takes at the singularity we can then reduce the areas excluded from the region for which the derivations hold to a set of measure zero.

There is no change in the formulas at all for either of the two cases where the additional boundary integrals vanish in the limiting process, namely when either λ is zero at the singularity or when there is no net outflow from it. Fortunately, these are the cases we are actually dealing with. Moreover, when we consider types of singularities where there is both net outflow and a nonzero λ, we only need to add new constant terms to X and M. Even if the formulas change, nothing in the conclusions is affected.

One further comment should be made here. The reasoning described may appear to apply only to transportation costs computed for optimal flow fields. If this were true, the results would be of rather limited interest. We used the optimality condition (7), but in fact we did not require the condition itself. Instead, we used the condition

$$k = (\text{grad}\ \lambda)\ \phi/|\phi| \tag{7a}$$

obtained from (7) after multiplication by $\phi/|\phi|$. Now, (7a) is a much weaker condition than (7), in that it only states that the scalar product of the vectors grad λ and $\phi/|\phi|$ equals k. In addition, (7) asserts that the two vectors are codirectional.

More important than this is the fact that (7a) makes sense for any flow fields, optimal or not, for which we let $\phi/|\phi|$ represent the direction vector and λ the path integral of k along the lines of the flow field. By the chain rule for derivatives, (grad λ) $\phi/|\phi| = \mathrm{d}\lambda/\mathrm{d}s$. If (13) holds along any line of the flow field, as is supposed by the definition of λ, $k\mathrm{d}s = \mathrm{d}\lambda$ must be true, This, however, is nothing but (7a) multiplied by the arc length element.

Equation (16a) is discussed further in Sections 2.2.8 and 2.2.9.

2.1.9 Relaxation of Constraints

From general economic theory it is well known that a sufficient condition for market equilibrium is that excess supply should be nonnegative (provided there is free disposal). If it is positive, the associated price must be zero. In this section we develop the spatial equivalent of this weaker equilibrium condition.

We begin by reinterpreting q, which we will now use to denote local demand minus local availability. The requirement that aggregate supply and demand should balance is now relaxed in the following way: local demand must still be satisfied everywhere, but instead of supply, local availability is given and represents an upper bound on local supply. In view of the sign of q (positive for excess demand, negative for excess availability), the relaxed constraint has the form

$$\text{div}\ \phi + q \leqslant 0 \tag{2a}$$

The feasibility condition for a closed region must also be relaxed. By the Gauss integral theorem (Courant and John 1974, pp. 59–61)

$$\begin{aligned} 0 &= \int_{\partial A} \phi_n \mathrm{d}s = \iint_A \text{div}\ \phi\ \mathrm{d}x_1 \mathrm{d}x_2 \\ &\leqslant -\iint_A q(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 \qquad \text{using (2a)} \end{aligned} \tag{17}$$

The relaxed feasibility condition is therefore

$$\iint_A q(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 \leqslant 0 \tag{1a}$$

Aggregate excess demand must be nonpositive. Notice that this is in agreement with the well-known equilibrium condition for competitive markets that excess supply of each commodity must be nonnegative (cf., e.g. Arrow and Hahn 1970).

We turn now to the second part of the equilibrium condition.

$$\lambda(\underline{x}) = 0 \text{ wherever } q(\underline{x}) < 0$$

Consider the problem of

$$\min_{\phi} \int\int k|\phi| \, dx_1 \, dx_2 \qquad \text{subject to (2a)}$$

A first consequence of the inequality sign in the constraint is that Lagrangean multipliers, i.e. prices, are now necessarily nonnegative. Moreover, when the < sign applies in (1a) then there must exist points (x_1, x_2) at which the < sign applies in (2a). The constraint (2a) is ineffective there: the minimum would not be changed by dropping the constraint there, and hence the Lagrangean multiplier λ must vanish there

$$\lambda(x_1, x_2) = 0 \qquad \text{where } \operatorname{div} \phi + q(x_1, x_2) < 0$$

Therefore the level of λ is no longer arbitrary. In fact:

Lemma: *Suppose that the flow field ϕ vanishes nowhere in A, and that*

$$\int\int_A q \, dx_1 \, dx_2 < 0$$

Then the function $\lambda = \lambda(x_1, x_2)$ is uniquely determined in A.

Proof: It is shown below (in Section 2.2.6) that for a flow field ϕ that does not vanish anywhere in A, any two potentials λ and μ must differ by a constant c

$$\lambda(x_1, x_2) = \mu(x_1, x_2) + c \tag{18}$$

Moreover the flow directions $\phi/|\phi|$ are uniquely determined. Consider now

$$K = \int\int k|\phi| + \lambda(\operatorname{div} \phi + q) dx_1 \, dx_2$$

$$= \int\int k|\phi| + \mu(\operatorname{div} \phi + q) dx_1 dx_2$$

Subtraction yields

$$\int\int (\lambda - \mu)(\operatorname{div} \phi + q) dx_1 \, dx_2 = 0$$

or, using (18)

$$c\int\int \operatorname{div} \phi \, dx_1 \, dx_2 + c\int\int q \, dx_1 \, dx_2 = 0$$

Then the first term vanishes, because of (17), leaving

$$c \cdot \int\int q dx_1 \, dx_2 = 0$$

and this, together with the < sign in (1a), implies $c = 0$. Therefore the potential function is unique. *Q.E.D.*

In this connection we also note the following:

Lemma: *Let*

$$\operatorname{div}\phi + q \leqslant 0 \tag{2a}$$

$$\iint q \, dx_1 \, dx_2 = 0 \qquad \phi_n = 0 \tag{1}$$

then

$$\operatorname{div}\phi = -q \qquad \textit{everywhere in } A$$

Proof:

$$0 \geqslant \iint (\operatorname{div}\phi + q)\, dx_1\, dx_2 \qquad \text{using (2a)}$$

$$= \int \phi_n \, ds + \iint q \, dx_1 \, dx_2 = 0$$

by (1) and by the Gauss integral theorem. Therefore the nonpositive integrand div $\phi + q$ vanishes everywhere except on a set of measure zero. *Q.E.D.*

The same argument applies to (1) and

$$\operatorname{div}\phi + q \geqslant 0 \tag{2b}$$

Thus if aggregate excess demand is zero, the divergence equation may be replaced everywhere by the same inequality.

2.1.10 Sensitivity

When constraint (2) is not binding but is relaxed as in (2a), the effect of changes in the local excess demand functions q – and of transportation cost k – on the minimum of aggregate transportation cost can be discussed. Denote the value of the minimand (8) by K

$$K = \min_{\phi} \iint k|\phi| \, dx_1 \, dx_2$$

subject to (2a) and (3). In terms of the Lagrange function we have an unrestricted expression for K

$$-K = \iint L \, dx_1 \, dx_2 = \iint -k|\phi| - \lambda[\operatorname{div}\phi + q]\, dx_1 \, dx_2$$

Now, taking partial derivatives

$$\begin{aligned}\frac{\partial K}{\partial k(x_1,x_2)} &= |\phi(x_1,x_2)|\mathrm{d}x_1\,\mathrm{d}x_2 \\ \frac{\partial K}{\partial q(x_1,x_2)} &= \lambda(x_1,x_2)\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned} \tag{19}$$

From (19) we see that a unit increase in local transportation cost at location $\underline{x}$ (and nowhere else) raises minimum total transportation cost by the flow density $|\phi|$ times the (infinitesimal) area $\mathrm{d}x_1\mathrm{d}x_2$. Similarly we conclude that a unit increase of demand in location $\underline{x}$ but nowhere else raises minimum total transportation cost by $\lambda(x_1,x_2)$, while a unit increase of availability lowers total transportation cost by $-\lambda(x_1,x_2)$ times the (infinitesimal) area $\mathrm{d}x_1\,\mathrm{d}x_2$.

A simultaneous increase of demand in one location $(\bar{x}_1,\bar{x}_2)$ and increase of availability in another location (x'_1,x'_2) may increase or decrease aggregate transportation cost, depending on the sign of the difference

$$\lambda(\bar{x}_1,\bar{x}_2)-\lambda(x'_1,x'_2) \gtrless 0$$

The fact that a "program" increase (simultaneous increase of supply and demand, but in different locations) may lower total transportation cost has been considered a paradox by some writers (e.g. Charnes), but it is in fact quite natural in the light of the different values of commodity availability in different locations.

2.1.11 A Minimax Theorem

What follows is a generalization of the Kuhn–Tucker theorem to the special convex minimum problem with linear constraint that is represented by the continuous transportation model.

Consider the function $K(\phi,\lambda)$ defined by

$$K(\phi,\lambda) = \int\!\!\int k(\underline{x})\,|\phi(\underline{x})| + \lambda(\underline{x})[\operatorname{div}\phi(\underline{x}) + q(\underline{x})]\,\mathrm{d}x_1\,\mathrm{d}x_2$$

Let

$$\phi = \hat{\phi} \text{ and } \lambda = \hat{\lambda}$$

represent flows and Lagrangean multipliers associated with an optimal solution of the continuous transportation model where the constraint has the form (2a). Then $K(\phi,\lambda)$ has a saddle point in $(\hat{\phi},\hat{\lambda})$

$$K(\hat{\phi},\lambda) \leqslant K(\hat{\phi},\hat{\lambda}) \leqslant K(\phi,\hat{\lambda}) \tag{20}$$

over the space of nonnegative scalar functions $\lambda(x) \geqslant 0$ and of piecewise-smooth vector fields $\phi(\underline{x})$ satisfying boundary condition (3).

This may be proved by taking a sequence of discrete nonlinear programs to the limit and applying the Kuhn–Tucker theorem to these. The theorem represents the limit of the Kuhn–Tucker statements. Notice that the constraint qualification is not required since the constraints are linear.

2.1.12 Comparative Statics

Consider two problems with different transportation costs but identical source–sink distributions

$$K = \min_{\phi} \int\int k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$M = \min_{\psi} \int\int (k + \Delta k)\,|\psi|\,\mathrm{d}x_1\mathrm{d}x_2$$

subject to

$$\operatorname{div}\phi + q \leqslant 0 \qquad \operatorname{div}\psi + q \leqslant 0 \tag{2a}$$

and denote the minimizing solutions by ϕ and $\phi + \Delta\phi$, respectively. Then minimization implies

$$\int\int (k + \Delta k)\,|\phi + \Delta\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant \int\int (k + \Delta k)\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{21}$$

and

$$\int\int k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant \int\int k\,|\phi + \Delta\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{22}$$

Adding (21) and (22)

$$\int\int \Delta k\,|\phi + \Delta\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant \int\int \Delta k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

or

$$\int\int \Delta k \cdot [|\phi + \Delta\phi| - |\phi|]\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant 0 \tag{23}$$

An isolated local change in transportation cost can only lead to a change with opposite sign in the local volume of flow.

To study the effects of changing q we compare the dual problems (cf. Section 2.2.9)

$$\max_{|\mathbf{grad}\,\lambda| \leqslant k} \int\int \lambda q\;\mathrm{d}x_1\,\mathrm{d}x_2 \quad \text{and} \quad \max_{|\mathbf{grad}\,\mu| \leqslant k} \int\int \mu\;(q + \Delta q)\mathrm{d}x_1\,\mathrm{d}x_2$$

Write the solutions as λ and $\lambda + \Delta\lambda$, respectively, and then by virtue of maximality

$$\int\int \lambda q \; \mathrm{d}x_1 \mathrm{d}x_2 \geqslant \int\int (\lambda + \Delta\lambda) q \; \mathrm{d}x_1 \mathrm{d}x_2 \tag{24}$$

and

$$\int\int (\lambda + \Delta\lambda)(q + \Delta q) \mathrm{d}x_1 \mathrm{d}x_2 \geqslant \int\int \lambda (q + \Delta q) \mathrm{d}x_1 \mathrm{d}x_2 \tag{25}$$

Adding (24) and (25)

$$\int\int (\lambda + \Delta\lambda) \cdot \Delta q \, \mathrm{d}x_1 \mathrm{d}x_2 \geqslant \int\int \lambda \Delta q \; \mathrm{d}x_1 \mathrm{d}x_2$$

or

$$\int\int \Delta\lambda \Delta q \; \mathrm{d}x_1 \mathrm{d}x_2 \geqslant 0 \tag{26}$$

A local change in excess demand can only lead to change of the local efficiency price in the same direction.

When both q and k are allowed to change simultaneously, the application of the same type of argument to the combined problem

$$\max_{\lambda \geqslant 0} \min_{\phi} \int\int k |\phi| + \lambda [\operatorname{div} \phi + q] \, \mathrm{d}x_1 \mathrm{d}x_2$$

shows that

$$\int\int \{-\Delta k \cdot (|\phi + \Delta\phi| - |\phi|) + \Delta q \cdot \Delta\lambda\} \mathrm{d}x_1 \mathrm{d}x_2 \geqslant 0 \tag{27}$$

which of course is a weaker statement than (23) and (26) taken separately.

2.2 MATHEMATICAL ASPECTS

In this section we present various aspects of the continuous transportation model that are of a more mathematical rather than economic nature. They may be skipped at a first reading. The region A to be considered is assumed to be simply connected and bordered by a Jordan curve ∂A.

2.2.1 Feasibility

As with every mathematical programming problem, the question of feasibility arises here. For the continuous transportation model with homogeneous boundary conditions $\phi_n = 0$ in ∂A, feasibility is a simple matter, which can be disposed of as follows.

Lemma: *Let q be continuous and let $\iint q \, dx_1 dx_2 = 0$. Then there exists a flow field ϕ such that*

$$\operatorname{div} \phi + q = 0 \qquad \textit{all } (x_1, x_2) \in A \tag{28}$$

$$\phi_n = 0 \qquad \textit{all } (x_1, x_2) \in \partial A \tag{29}$$

This lemma can actually be refined as follows.

Theorem: *Let q be continuous and let $\iint q \, dx_1 dx_2 = 0$. Then there exists a flow field proportional to a gradient field (i.e. a flow field whose trajectories are gradient lines) such that*

$$\operatorname{div} \phi + q = 0 \qquad \textit{all } (x_1, x_2) \in A \tag{28}$$

$$\phi_n = 0 \qquad \textit{all } (x_1, x_2) \in \partial A \tag{29}$$

Proof: To construct a gradient field ψ consider

$$\min_{\psi : \operatorname{div} \psi = -q} \int\int \tfrac{1}{2} |\psi|^2 \, dx_1 dx_2$$

Its solution is

$$\psi = \operatorname{grad} \mu$$

Substituting in (28)

$$-q = \operatorname{div} \psi = \operatorname{div} (\operatorname{grad} \mu) \qquad \text{in } A$$

$$(\operatorname{grad} \mu)_n = 0 \qquad \text{in } \partial A$$

The resulting equation, the Poisson equation, may be solved. Hence there exist feasible solutions that are gradient fields. *Q.E.D.*

Another (more intuitive) way of constructing a feasible solution is as follows: place a laminar flow field with parallel directions, say, through A, enter the flow for all sources and sinks into this laminar field and circulate any resulting excess flow (positive or negative) around the boundary. This satisfies both the source–sink equation and the boundary condition.

2.2.2 Bounds

One lower bound for total transportation cost is clearly zero

$$K \equiv \int\int k |\phi| \, dx_1 dx_2 \geqslant 0$$

To obtain a better bound consider

$$\begin{aligned}\int\int \lambda q \,\mathrm{d}x_1\,\mathrm{d}x_2 &= -\int\int \lambda \,\mathrm{div}\,\phi \,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= -\int\int \mathrm{div}\,(\lambda\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &\quad+\int\int \phi \,\mathrm{grad}\,\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= -\int \lambda\,\phi_n \mathrm{d}s + \int\int \phi\,\mathrm{grad}\,\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &\leqslant \int\int \phi\cdot k\frac{\phi}{|\phi|}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

for all ϕ satisfying the boundary condition and any λ such that

$$|\mathrm{grad}\lambda| \leqslant k$$

We have therefore the following:

Lemma: *For any ϕ satisfying*

$$0 = \mathrm{div}\,\phi + q \qquad \textit{in } A \tag{28}$$

$$\phi_n = 0 \qquad \textit{on } \partial A \tag{29}$$

and any λ with $|\mathrm{grad}\,\lambda| \leqslant k$ one has

$$\int\int k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant \int\int \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{30}$$

An upper bound on total transportation cost may be obtained as follows. Let

$$k^* = \max_{\underline{x}\in A} k(\underline{x})$$

then

$$\int\int k|\phi|\,\mathrm{d}x \leqslant \int\int k^*|\phi|\,\mathrm{d}x = k^*\int\int|\phi|\,\mathrm{d}x$$

Now for uniform transportation cost k^* the flow lines are straight lines (cf. Section 2.1.4). An upper bound on total flow $\int\int|\phi|\mathrm{d}x$ is obtained by assuming that all flow originates in one point and moves to another point at maximum distance R. Let R be this largest distance, the diameter of region A. Then

$$\int\int|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant k\cdot Q\cdot R$$

where Q is aggregate demand

$$Q = \int\int \max\,[q(\underline{x}), 0]\,\mathrm{d}x$$

$$= \tfrac{1}{2}\int\int |q(x)|\,\mathrm{d}x$$

in view of $\int\int q\ \mathrm{d}x = 0$. Thus, finally

$$\int\int k|\phi|\,\mathrm{d}x \leqslant k^* \cdot R \cdot \tfrac{1}{2}\int\int |q|\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{31}$$

2.2.3 Convexity

The continuous transportation model is a convex programming problem. This may be seen from the following:

Lemma: min $\int\int k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$ *subject to (28) and (29) has a convex objective function and linear constraints. Further*

$$\max_{\lambda} \int\int \lambda q\ \mathrm{d}x_1\,\mathrm{d}x_2$$

subject to $|\mathrm{grad}\,\lambda| \leqslant k$ *has a linear objective function and constraints defining a convex set.*

Proof: We will show that the function $k|\phi|$ is convex in ϕ. Let $0 < a < 1$, and suppose

$$a|\phi| + (1-a)\ |\psi| \leqslant |a\phi + (1-a)\psi|$$

Squaring,

$$(a|\phi| + (1-a)|\psi|)^2 \leqslant |a\phi + (1-a)\psi|^2$$

or,

$$a^2|\phi|^2 + 2a(1-a)|\phi|\,|\psi| + (1-a)^2\ |\psi|^2 \leqslant a^2|\phi|^2 + 2a(1-a)\phi\cdot\psi + (1-a)^2|\psi|^2$$

i.e.,

$$2a(1-a)|\phi|\,|\psi| \leqslant 2a(1-a)\phi\psi$$

By the Cauchy–Schwarz inequality, this is false except when $\phi \| \psi$. Thus $|\phi|$ is a convex function of ϕ and is strictly convex with respect to directions $\phi/|\phi|$. *Q.E.D.*

2.2.4 Euler–Lagrange Equations

Constraint (1) of the earlier maximization problem 2.1.4 may be incorporated in the maximand by the method of Lagrange multipliers

$$-K = \max_{\phi} \int\int -k|\phi| - \lambda[\mathrm{div}\,\phi + q]\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{32}$$

In components

$$-K = \max_{\phi_1, \phi_2} \int\int -k[(\phi_1^2 + \phi_2^2)]^{1/2} - \lambda \left[\frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2} + q \right] \mathrm{d}x_1 \, \mathrm{d}x_2$$

The Euler–Lagrange equations of this two-dimensional calculus of variations problem are

$$-k \frac{\phi_i}{[(\phi_1^2 + \phi_2^2)]^{1/2}} = -\frac{\partial \lambda}{\partial x_i} \qquad i = 1,2$$

or, reverting to vector notation

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \tag{33}$$

The left-hand side of this equation is defined only for nonvanishing ϕ. The efficiency condition for the case of vanishing ϕ must be derived in a different way.

2.2.5 Alternative Derivation of Efficiency Conditions

For every feasible flow field, i.e. every flow field satisfying the source–sink relationship (28) and every differentiable λ, we have

$$\begin{aligned} \int\int k|\phi| \mathrm{d}x_1 \, \mathrm{d}x_2 &= \int\int k|\phi| + \lambda(\operatorname{div} \phi + q) \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int\int [k|\phi| + \operatorname{div}(\lambda\phi) - \phi \cdot \operatorname{grad} \lambda + q\lambda] \, \mathrm{d}x_1 \, \mathrm{d}x_2 \end{aligned} \tag{34}$$

since $\operatorname{div}(\lambda\phi) = \phi \cdot \operatorname{grad} \lambda + \lambda \operatorname{div} \phi$. By the Gauss integral theorem

$$\int\int \operatorname{div}(\lambda\phi) \mathrm{d}x_1 \, \mathrm{d}x_2 = \int \lambda \phi_n \mathrm{d}s = 0$$

in view of the boundary condition (29). Rewriting

$$k|\phi| = k \frac{\phi \cdot \phi}{|\phi|}$$

one has

$$\int\int k|\phi| \mathrm{d}x_1 \, \mathrm{d}x_2 = \int\int \left(k \frac{\phi}{|\phi|} - \operatorname{grad} \lambda \right) \cdot \phi + q\lambda \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

This can be a maximum with respect to nonvanishing ϕ if and only if the parenthetical term vanishes

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda$$

A vanishing ϕ is optimal where and only where the inner product is positive. This is true wherever

$$|\text{grad}\,\lambda| < k$$

since then

$$\begin{aligned} k\frac{\phi\cdot\phi}{|\phi|} - \text{grad}\,\lambda\cdot\phi &= k|\phi| - \text{grad}\,\lambda\cdot\phi \\ &\geqslant k|\phi| - |\text{grad}\,\lambda|\,|\phi| \\ &= (k - |\text{grad}\,\lambda|)\,|\phi| \\ &> 0 \qquad \text{for } |\phi| \neq 0 \end{aligned}$$

Therefore, (28) can be a minimum only if

$$\frac{k\phi}{|\phi|} = \text{grad}\,\lambda \qquad \text{for } \phi \neq 0 \tag{33}$$

$$k \geqslant |\text{grad}\,\lambda| \qquad \text{for } \phi = 0 \tag{34}$$

2.2.6 Uniqueness

Suppose

$$k\frac{\phi}{|\phi|} = \text{grad}\,\lambda \text{ in } A \qquad \phi_n = 0 \text{ on } \partial A$$

$$k\frac{\psi}{|\psi|} = \text{grad}\,\mu \text{ in } A \qquad \psi_n = 0 \text{ on } \partial A$$

are two solutions. Consider

$$\eta = a\phi + (1-a)\psi \qquad 0 < a < 1$$

Then

$$\eta_n = a\phi_n + (1-a)\psi_n = 0 \qquad \text{on } \partial A$$

and

$$\begin{aligned} \text{div}\,\eta &= a\,\text{div}\,\phi + (1-a)\,\text{div}\,\psi \\ &= -aq - (1-a)q = -q \end{aligned}$$

so that η satisfies the constraints. The value of the objective function is then

$$\begin{aligned} \iint k|\eta|\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \iint k|a\phi + (1-a)\psi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &< a\iint k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 + (1-a)\iint k\,|\psi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \end{aligned}$$

by convexity, unless $\phi \| \psi$. This inequality contradicts the assumption that

$$K = \int\int |\phi| \mathrm{d}x_1 \mathrm{d}x_2 = \int |\psi| \mathrm{d}x_1 \mathrm{d}x_2$$

was a minimum. Therefore $\phi \| \psi$.

Assume now that $\phi \neq 0$ everywhere in A (except for a set of measure zero). It follows that

$$\operatorname{grad} \lambda = k \frac{\phi}{|\phi|} = \operatorname{grad} \mu$$

The only solution of

$$\operatorname{grad} (\lambda - \mu) = 0 \qquad \text{everywhere in } A$$

is

$$\lambda = \mu + \text{constant}$$

Summarizing, we have shown the following:

Uniqueness Theorem

The solution $\phi/|\phi|$ is uniquely determined. If ϕ does not vanish anywhere (except at singularities on a set of measure zero), then λ is unique up to an additive constant. On any set of nonzero measure with $\phi \equiv 0$, λ is arbitrary but must satisfy

$$|\operatorname{grad} \lambda| \leqslant k \tag{34}$$

An alternative proof of uniqueness is as follows. Let $k\, \phi/|\phi| = \operatorname{grad} \lambda$ and $k\psi/|\psi| = \operatorname{grad} \mu$ be two solutions. Since both satisfy the boundary condition (2) it follows that

$$0 = \int (\lambda - \mu)(\phi - \psi)_n \mathrm{d}s$$

$$= \int\int \operatorname{div}((\lambda - \mu)(\phi - \psi)) \mathrm{d}x_1 \mathrm{d}x_2 \qquad \text{by the Gauss integral theorem}$$

$$= \int\int (\lambda - \mu) \operatorname{div}(\phi - \psi) + (\phi - \psi)\, [\operatorname{grad}(\lambda - \mu)]\, \mathrm{d}x_1 \mathrm{d}x_2$$

The first term vanishes since

$$\operatorname{div}(\phi - \psi) = \operatorname{div}\phi - \operatorname{div} \psi = -q + q = 0$$

Thus

$$
\begin{aligned}
0 &= \iint (\phi - \psi)\,\mathrm{grad}(\lambda - \mu)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \iint \phi\cdot\mathrm{grad}\,\lambda + \psi\cdot\mathrm{grad}\,\mu - \phi\cdot\mathrm{grad}\,\mu - \psi\cdot\mathrm{grad}\,\lambda\ \mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \iint k|\phi| + k|\psi| - k\phi\cdot\frac{\psi}{|\psi|} - k\psi\cdot\frac{\phi}{|\phi|}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \iint k\phi\cdot\left(\frac{\phi}{|\phi|} - \frac{\psi}{|\psi|}\right) + k\psi\cdot\left(\frac{\psi}{|\psi|} - \frac{\phi}{|\phi|}\right)\,\mathrm{d}x_1\,\mathrm{d}x_2
\end{aligned}
$$

Now the two terms of the integrand are nonnegative since, e.g.

$$\frac{\phi\cdot\phi}{|\phi|} = |\phi| \quad \text{and} \quad \frac{\phi\cdot\psi}{|\psi|} \leqslant |\phi|$$

They are zero if and only if

$$\frac{\phi}{|\phi|} = \frac{\psi}{|\psi|}$$

It follows that

$$\frac{\phi}{|\phi|} \equiv \frac{\psi}{|\psi|}$$

everywhere in A so that the flow lines must coincide. Moreover, the potential functions must be identical except for an additive constant. (The constant may be different for disconnected parts of the flow field.)

That the quantities of flow $|\phi|$ need not be unique follows from the discussion of singular flow in the next section.

2.2.7 Digression on Singular Flows

With suitable precautions, the validity of the divergence and gradient laws can be extended to singular flow fields. The simplest case is that of a single source of finite output. In principle, this could be handled by constructing a circular boundary of small radius around the single source and treating its output as a cross-boundary flow. But formally all equations remain valid. The feasibility condition is once more

$$\iint q\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0$$

or

$$q_0 = -\iint\limits_{\substack{\text{area}\\\text{outside}\\\text{source}}} q\,\mathrm{d}x_1\,\mathrm{d}x_2$$

where q_0 is the output of the single source.

A second case of interest concerns a single finite source and a single sink of equal strength. With linear transportation cost the flow field consists of a single flow line along a geodesic joining source and sink. For uniform transportation cost this flow line is straight. In any case the flow line carries a finite amount of flow through an infinitesimal cross section.

Also of interest is the case of a finite set of sources and sinks, each of finite yield, such that total net output is zero. This is essentially the linear programming transportation model to be discussed later in Section 2.5.5. It is also an example for which the flow field is unique while the flow quantities on particular flow lines need not be unique. This arises when "neutral circuits" occur (see Figure 2.14), i.e. when two distinct flow paths of equal length exist.

Since opposite sides of the parallelogram are of equal length, one has

$$\phi_1 + \phi_2 = 1 \qquad \phi_1 = \phi_3$$

$$\phi_3 + \phi_4 = 1 \qquad \phi_2 = \phi_4$$

which is solved by any $0 \leqslant \phi_1 \leqslant 1$.

Finally, one may consider flows originating and/or terminating on lines with a finite line density (cf. Section 2.5.2). Singular flows may also occur along the boundary ∂A of the admissible region A. The boundary condition restricts the flow entering or leaving, however.

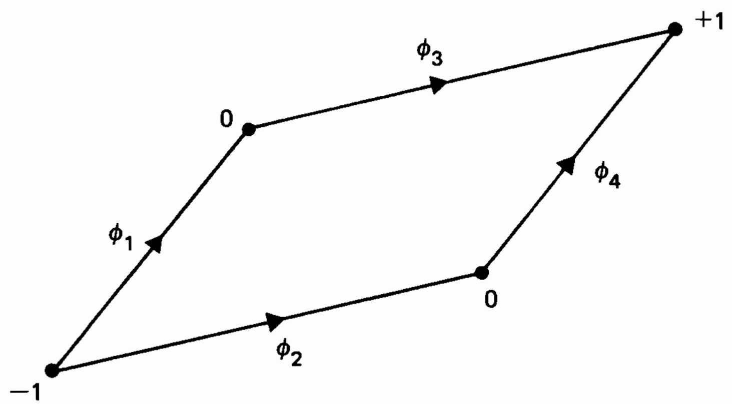

Figure 2.14. Neutral circuits.

2.2.8 Alternative Expression for Minimal Total Transportation Cost

At this point a concise restatement of Section 2.1.8 is in order.

$$K = \min_{\phi} \iint_A k|\phi| \mathrm{d}x_1 \mathrm{d}x_2 = \iint_A k|\phi| + \lambda(\operatorname{div}\phi + q)\mathrm{d}x_1 \mathrm{d}x_2 \qquad \text{for } \operatorname{div}\phi = -q$$

$$= \int\int_A \phi \operatorname{grad} \lambda + \lambda(\operatorname{div}\phi + q)\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\int_A [\operatorname{div}(\lambda\phi) + \lambda q]\,\mathrm{d}x_1\mathrm{d}x_2$$

$$= \int_{\partial A} (\lambda\phi)_n\,\mathrm{d}s + \int\int_A \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$K = \int\int_A \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{35}$$

since the boundary integral vanishes by the boundary condition (29).

2.2.9 Duality

Theorem:

$$\min_{\phi} \int\int_A k|\phi|\mathrm{d}x_1\,\mathrm{d}x_2 = \max_{\lambda} \int\int_A \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{36}$$

$$\phi\text{: } \operatorname{div}\phi + q = 0 \qquad |\operatorname{grad}\lambda| \leqslant k$$

$$\phi_n = 0 \text{ on } \partial A \tag{37}$$

Proof: For all ϕ satisfying (28)

$$\int\int k|\phi|\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int k|\phi| + \lambda(\operatorname{div}\phi + q)\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\int k|\phi| + \lambda q + \operatorname{div}(\lambda\phi) - \phi\cdot\operatorname{grad}\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\int \phi\cdot\left(k\frac{\phi}{|\phi|} - \operatorname{grad}\lambda\right) + \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int (\lambda\phi)_n\mathrm{d}s$$

$$\geqslant \int\int \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad \text{since } \lambda\phi_n = 0 \text{ on } \partial A$$

and since

$$\phi\cdot\left(k\frac{\phi}{|\phi|} - \operatorname{grad}\lambda\right) \geqslant 0$$

for all λ satisfying (37). Thus

$$\min_{\phi:\,(28)} \int\int |\phi|\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant \max_{\lambda:\,(37)} \int\int \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2$$

Now the equals sign is attained as shown in Section 2.2.8. Therefore, on the assumption that the primal problem

$$\min_{\phi:\,(28)} \int\!\!\int_0 k\,|\phi|\,dx_1\,dx_2$$

has a solution, the theorem follows. *Q.E.D.*

We now show:

Lemma: *The efficiency conditions of the dual are the constraints of the primal.*

Proof: Consider

$$\max_{\lambda} \int\!\!\int q\lambda\ dx_1\,dx_2$$

subject to

$$(\text{grad}\ \lambda)^2 \leqslant k^2 \tag{38}$$

This has the concave Lagrangean

$$\max \int\!\!\int \lambda q + \frac{\mu}{2}\ [k^2 - (\text{grad}\ \lambda)^2]\,dx_1\,dx_2 \tag{39}$$

The Euler–Lagrange equation is

$$q + \text{div}(\mu\ \text{grad}\ \lambda) = 0$$

Letting

$$\mu\ \text{grad}\ \lambda = \phi \tag{40}$$

this is the constraint of the primal. The free boundary condition of (39) is

$$-\left(\frac{\mu}{2}\text{grad}\ \phi\right)_n = 0$$

or, in view of (40)

$$\phi_n = 0$$

which is the boundary condition of the primal. *Q.E.D.*

Suppose that A is partitioned and excess demand is such that

$$q(\underline{x}) = \begin{cases} 1 & \underline{x} \in A_0 \\ -1 & \underline{x} \in A_1 \end{cases} \qquad A = A_0 \cup A_1$$

$$|A_0| = |A_1|$$

The duality principle states that

$$\iint_{A_0} \lambda(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 - \iint_{A_1} \lambda(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

is maximized subject to (37).

Subject to the gradient condition (37), the price spread between consumption and production areas is thus maximized. This maximum decreases when the cost k decreases.

2.3 EXTENSIONS: BOUNDARY CONDITIONS

2.3.1 Free Boundary Condition

In the continuous transportation model the boundary condition $\phi_n = 0$ serves to isolate the region A economically from the rest of the world. What happens when no such boundary condition is imposed? Flows across the boundary are even then not entirely unrestricted. For, by the Gauss integral theorem

$$\begin{aligned}\int_{\partial A} \phi_n \,\mathrm{d}x &= \iint \operatorname{div} \phi \,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= -\iint q \,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

by the divergence law. If we impose the previous condition that aggregate excess demand be balanced

$$\iint q \,\mathrm{d}x_1\,\mathrm{d}x_2 = 0$$

then

$$\int_{\partial A} \phi_n \,\mathrm{d}s = 0 \tag{41}$$

This "free boundary condition" states that aggregate flow between the region and the outside world must be balanced. Notice that this balance condition is in physical and not value terms. We shall show that optimization implies that this balance equation must hold in value terms as well.

As before, we write

$$\begin{aligned}-K &= \iint -k|\phi| - \lambda\,[\operatorname{div}\phi + q]\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \iint -k|\phi| - \operatorname{div}(\lambda\phi) + \phi\cdot\operatorname{grad}\lambda - q\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int \lambda\phi_n\,\mathrm{d}s + \iint -q\lambda + \phi\cdot\left[\operatorname{grad}\lambda - k\frac{\phi}{|\phi|}\right]\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

For any flow field $\phi/|\phi|$ satisfying the gradient law

$$k\frac{\phi}{|\phi|} = \text{grad}\,\lambda \tag{42}$$

this becomes

$$-K = \int \lambda\phi_n \mathrm{d}s - \int\int q\,\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2$$

The gradient law (42) prescribes only the direction of flow, not its volume. We may still vary the net outflow ϕ_n on the boundary provided we satisfy the free boundary condition (41). K can be a maximum only if λ is constant throughout ∂A. Therefore

$$\int\lambda\,\phi_n \mathrm{d}s = \lambda\int\phi_n \mathrm{d}s = 0 \tag{43}$$

$$\lambda = \text{constant} \tag{44}$$

so that the flow condition (41) implies the value condition (43). Since the level of λ is arbitrary, we may in particular choose

$$\lambda = 0 \qquad \text{on } \partial A$$

Under a free boundary condition the value of the minimum is, of course, lower than under any boundary condition that imposes an effective constraint.

We have considered the situation where the product may be obtained freely from anywhere outside the region provided we release the same amount of this commodity at other points of the boundary. This is economically meaningful and consistent with an optimum only when the economic value λ of the commodity is the same everywhere along the boundary.

The boundary condition (41) is not implied but does impose an extra constraint when the divergence law is relaxed

$$\text{div}\,\phi + q \leqslant 0 \tag{45}$$

Consider the case where

$$\int\int q\,\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant 0 \tag{46}$$

The combination of (45) and (46) implies that net imports are now nonpositive

$$\begin{aligned} 0 &\leqslant \int\int q\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &\leqslant \int\int -\text{div}\,\phi\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad \text{by (45)} \\ &= -\int \phi_n \mathrm{d}s \end{aligned}$$

The problem considered should then be one of minimizing the sum of import costs and transportation costs (cf. Section 2.3.3).

2.3.2 Inhomogeneous Boundary Conditions

Consider

$$\min \iint k|\phi|\,dx_1\,dx_2$$

subject to

$$\operatorname{div}\phi + q = 0 \qquad \text{in } A \tag{47}$$

$$\phi_n = g(x_1, x_2) \qquad \text{on } \partial A \tag{48}$$

The conditions are consistent if and only if

$$\iint_A q\;dx_1\,dx_2 + \int_{\partial A} g\,ds = 0 \tag{49}$$

For any flow field satisfying (47) and (48) and any piecewise smooth scalar function $\lambda(x_1, x_2)$

$$\begin{aligned}
\iint k|\phi|\,dx_1\,dx_2 &= \iint k|\phi| + \lambda\,[\operatorname{div}\phi + q]\,dx_1\,dx_2 \\
&= \iint k\frac{\phi}{|\phi|}\phi + \lambda\operatorname{div}\phi + \lambda q\;dx_1\,dx_2 \\
&= \iint k\frac{\phi}{|\phi|}\phi + \operatorname{div}(\lambda\phi) - \phi\operatorname{grad}\lambda + q\lambda\;dx_1\,dx_2 \\
&= \int_{\partial A}(\lambda\phi)_n\,ds + \iint_A \phi\left(k\frac{\phi}{|\phi|} - \operatorname{grad}\lambda\right) + q\lambda\;dx_1\,dx_2 \\
&= \int_{\partial A}\lambda g\,dx + \iint_A \phi\left(k\frac{\phi}{|\phi|} - \operatorname{grad}\lambda\right) + q\lambda\;dx_1\,dx_2 \\
&\geqslant \int_{\partial A}\lambda g\,ds + \iint_A q\lambda\;dx_1\,dx_2
\end{aligned}$$

for all λ satisfying $|\operatorname{grad}\lambda| \leqslant k$, and "=" for $\operatorname{grad}\lambda = k\phi/|\phi|$. This proves

$$\min_{\phi:}\iint k|\phi|\,dx_1\;dx_2 = \max_{\lambda}\int_{\partial A}\lambda g\,ds + \iint q\lambda\;dx_1\,dx_2 \tag{50}$$

$$\operatorname{div}\phi = q \text{ in } A \qquad |\operatorname{grad}\lambda| \leqslant k$$

$$\phi_n = g \text{ in } \partial A$$

The duality theorem is thus modified. However, nothing is changed in the uniqueness proof.

2.3.3 Free Trade at Given Boundary Prices

As the next possibility we consider boundary conditions in terms of price rather than quantity. This leads to a change in the objective function, because the gains from commodity trade at the boundary may now offset to some extent the cost of transportation

$$\max_{\phi} \int p\, \phi_n \,\mathrm{d}s - \int\int k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{51}$$

subject to

$$\operatorname{div}\phi + q = 0 \tag{47}$$

No restriction is now required with regard to $\int\int q\ \mathrm{d}x_1\,\mathrm{d}x_2$. On the other hand, some restriction must be imposed on p in order to make (51) bounded. Economically speaking the price gain along any path must never exceed the amount of transportation cost incurred. Otherwise infinite profits would be attainable. We may state this formally as follows.

Let $p(\underline{x})$ be any extension of $p(s)$ from the boundary ∂A to the interior of A. Then, $p(\underline{x})$ is admissible if everywhere

$$|\operatorname{grad} p| \leqslant k(\underline{x}) \tag{52}$$

The Lagrangean of the free trade problem is

$$-K = \int p\,\phi_n\,\mathrm{d}s - \int\int k\,|\phi| + \lambda[\operatorname{div}\phi + q]\,\mathrm{d}x_1\,\mathrm{d}x_2$$

which is transformed in the usual manner to

$$-K = \int (p-\lambda)\phi_n\ \mathrm{d}s + \int\int \left[\left(\operatorname{grad}\lambda - k\,\frac{\phi}{|\phi|}\right)\phi - \lambda q\right]\mathrm{d}x_1\ \mathrm{d}x_2$$

We conclude that the necessary conditions for a maximum are

$$k\,\frac{\phi}{|\phi|} = \operatorname{grad}\lambda \qquad \text{in } A \tag{42}$$

$$\lambda = p \qquad \text{in } \partial A \tag{53}$$

The prices λ must be a continuous extension to the interior of prices p on the boundary.

Boundary conditions have important mathematical implications as well. In Section 4.5.2 it will be shown how Huygens' principle may be applied to solve the differential equation (44). This requires, however, that it is known in which direction λ increases. In fact, this is often known. For instance, when (as frequently happens) production of a commodity is confined to a smaller region than is consumption, it may be inferred that commodity prices fall as one moves in the direction of and towards the interior of the production region.

Suppose, in particular, that

$$q \leqslant 0$$

everywhere in A. This is the case of "production for export." The price must fall as one moves from the boundary to the interior, and Huygens' principle is readily applied. The same is true with the direction of price increase reversed when

$$q \geqslant 0$$

everywhere in A. The region is then one of excess demand everywhere. The problem is then one of optimally supplying imports to meet local excess demands in the region. Finally, let

$$q \equiv 0 \qquad \text{in } A \tag{54}$$

We then have no local excess demand. All trade consists of transactions with the outside, for (54) implies

$$\int \phi_n \, \mathrm{d}s = \int \int \operatorname{div} \phi \, \mathrm{d}x_1 \mathrm{d}x_2 = -\int \int q \, \mathrm{d}x_1 \mathrm{d}x_2 = 0$$

This problem is stated here for a competitive external economy imposing its prices $p(s)$ at the region's boundary. In the next section we consider the case where the region's trade is large enough to affect prices in the outside world.

2.3.4 The Export–Import Problem

In the continuous model of transportation the divergence law represented a non-homogeneous differential equation and the boundary condition was homogeneous. This meant that all trade was confined to the interior while no trade took place across the boundary. The opposite situation occurs when the market is entirely concentrated on the boundary and no transactions occur in the interior.

Suppose that given quantities $g(s)$ have been contracted for net export across the boundary

$$\phi_n(s) = g(s) \qquad \text{on } \partial A \tag{55}$$

For this program to be carried out with zero excess demand in the interior then, in order that no accumulation or rundown of commodity stocks occurs, one must have that

$$\int_{\partial A} g(s) \, \mathrm{d}s = 0 \tag{56}$$

as a consequence of the Gauss integral theorem applied to a vanishing divergence.

We now consider the cheapest way of carrying out this import–export program, which means that the associated transportation cost must be minimized

$$\min_{\phi} \int\int k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{57}$$

subject to (55) and to the flow constraint applicable in the interior

$$\operatorname{div}\phi = 0 \tag{58}$$

The Lagrangean is now

$$K = -\int\int_A [k|\phi| + \lambda \operatorname{div}\phi]\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int\int_{\partial A} \mu[\phi_n - g]\,\mathrm{d}s$$

To maximize this concave Lagrangean observe that

$$-\lambda \operatorname{div}\phi = -\operatorname{div}(\lambda\phi) + \phi\cdot\operatorname{grad}\lambda$$

Substituting and applying the Gauss integral theorem to $\operatorname{div}(\lambda\phi)$

$$-K = -\int \lambda\,\phi_n\,\mathrm{d}s + \int\int \phi\cdot\operatorname{grad}\lambda - k|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int \mu[\phi_n - g]\,\mathrm{d}s$$

Rewriting

$$k|\phi| = \phi\cdot\frac{\phi}{|\phi|}$$

substituting, and rearranging terms

$$-K = -\int \mu g\,\mathrm{d}s + \int (\lambda - \mu)\phi_n\,\mathrm{d}s + \int\int \phi\cdot\left[\operatorname{grad}\lambda - k\frac{\phi}{|\phi|}\right]\mathrm{d}x_1\,\mathrm{d}x_2$$

This represents a maximum in ϕ only if

$$\lambda(s) = \mu(s) \qquad \text{on } \partial A \tag{59}$$

$$k\frac{\phi}{|\phi|} = \operatorname{grad}\lambda \qquad \text{in } A \tag{42}$$

Condition (42) is the familiar gradient law, which must also apply when trade occurs only at the boundary. Condition (59) states that prices μ on the boundary must be a continuous extension of prices λ in the interior. This rules out any windfall profits that could be obtained by transferring commodity from the interior to the boundary.

So far, quantities g have been given at the boundary or prices μ imposed. Consider now the more general situation where excess demand functions are given at the boundary

$$\lambda = \pi(\phi_n) \tag{60}$$

which relates prices λ for cross-boundary trade at location s to net exports ϕ_n at s. The region will now be considered as a monopolist seeking to maximize gains from international trade

$$\max_{\phi} \int \pi(\phi_n)\phi_n \, \mathrm{d}s - \int\int k|\phi| \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

subject to (58). Maximization with respect to ϕ in the interior yields the gradient law (42). Maximization with respect to ϕ_n yields the Amoroso–Robinson formula for monopoly pricing

$$\pi + \phi_n \frac{\mathrm{d}\pi}{\mathrm{d}\phi_n} = 0 \tag{61}$$

This becomes, in effect, a "mixed boundary condition" for the problem (57) subject to (58). Here we prescribe neither price nor flow itself but rather a profit-maximizing combination of both as given by (61). The function π will in general depend on s, i.e. on local conditions. When the profit function $\pi = \pi(\phi)$ is replaced by a utility function, e.g. a consumers' surplus

$$u(\phi_n) = \int_0^{\phi_n} p(g) \, \mathrm{d}g \tag{62}$$

then the monopolist is replaced by perfect competition. The object of the market is not maximizing profits to the region but maximizing welfare on the boundary of the region. This remains meaningful when the commodity is also consumed inside the region and a consumers' surplus function is attached to this.

2.3.5 Macro Relationships for Gradient Flow

In this section we take up the argument of Section 2.1.8 and extend it to the situation described by the broader boundary conditions of Section 2.3.2.

We assume that transportation is described by a gradient flow

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \text{wherever } \phi \neq 0 \text{ in } A \tag{42}$$

Define

$$K = \int\int_A k|\phi| \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

Substitute (42)

$$\begin{aligned} K &= \int\int \phi \cdot \operatorname{grad} \lambda \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int\int [\operatorname{div}(\lambda\phi) - \lambda \operatorname{div} \phi] \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int (\lambda\phi)_n \, \mathrm{d}s - \int\int \lambda \operatorname{div} \phi \, \mathrm{d}x_1 \, \mathrm{d}x_2 \end{aligned}$$

by the Gauss integral theorem

$$K = \int \lambda g \, \mathrm{d}s - \int\int \lambda \operatorname{div} \phi \, \mathrm{d}x_1 \mathrm{d}x_2$$

by the boundary condition (48). Using (47), finally

$$K = \int \lambda g \, \mathrm{d}x + \int\int \lambda q \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{63}$$

Now write

$$g = e - m \qquad e \geqslant 0, m \geqslant 0$$

where e = exports and m = imports. Similarly, let

$$q = d - z \qquad d \geqslant 0, z \geqslant 0$$

where d = demand and z = supply. Then the terms in equation (63) may be rewritten as follows

$$\begin{aligned} K &= \int_{\partial A} e \lambda \, \mathrm{d}s - \int_{\partial A} m \lambda \, \mathrm{d}s + \int\int_{A} d\lambda \, \mathrm{d}x_1 \mathrm{d}x_2 - \int\int_{A} z \lambda \, \mathrm{d}x_1 \mathrm{d}x_2 \\ &= E - M + C - Z \end{aligned}$$

where the aggregates E, M, C, and Z represent the following measures for the commodity in question:

Z = value of aggregate output,
C = value of aggregate consumption,
E = value of aggregate exports,
M = value of aggregate imports,
K = aggregate cost of transporting the commodity.

Thus

$$Z + K = C + E - M \tag{64}$$

Notice that this is a macroeconomic accounting equation for a single commodity.

If demand C were further broken down into private consumption, investment, and government consumption, we would have obtained the usual equation of macroeconomic income accounting, stated however for a single commodity. Notice that the value of output here is broken down into the value due to production Z and the value due to transportation K. This is because transportation constitutes an exogenous activity (cf. Chapter 4 for an endogenous treatment of transportation.)

2.4 EXTENSIONS: METRICS

2.4.1 Manhattan Metric

When transportation is restricted to the North–South and East–West directions, but the network may be considered infinitely dense, then the appropriate measure of transportation cost is given by

$$\int\int k_1 |\phi_1| + k_2 |\phi_2| \, dx_1 \, dx_2 \tag{65}$$

where ϕ_i denotes the vector components of ϕ. The efficiency conditions are now

$$k_i \operatorname{sign} \phi_i = \frac{\partial \lambda}{\partial x_i} \qquad i = 1,2 \tag{66}$$

and may thus be stated separately in the two components.

The possible directions of the gradient are then restricted to

$$\text{arc tan} \left(\frac{\partial \lambda}{\partial x_2} \Big/ \frac{\partial \lambda}{\partial x_1} \right) = \pm \frac{k_2}{k_1}, \quad \pm 0, \quad \pm \infty \tag{67}$$

(where we allow for the possibility that one flow component vanishes). In particular, if transportation costs are constant $k(x) \equiv k$, then

$$\text{arc tan} \left(\frac{\partial \lambda}{\partial x_2} \Big/ \frac{\partial \lambda}{\partial x_1} \right) = \pm 1, \quad \pm 0, \quad \pm \infty$$

so that the possible directions of the gradient are ± 0, $\pm 45°$, and $\pm 90°$ and are described by the star shown in Figure 2.15. It follows that the isopotential lines consist of straight-line segments. While the gradient law still applies, the interpretation in flow terms means that at any point the flow moves in either of the allowed directions, in such a way as to form an angle of, at most, 45° with the gradient. This means the flow lines are obtained by projecting the gradient on the admissible lines of motion.

2.4.2 Anisotropic Linear Homogeneous Metric

Consider now transportation models in which the transportation cost function is linear homogeneous with respect to flow

$$k(t\phi, x_1, x_2) = tk(\phi, x_1, x_2) \tag{68}$$

Of course, k must be a metric, i.e.

$$k \geqslant 0 \text{ and } k = 0 \text{ only for } \phi = 0 \tag{69}$$

$$k(\phi + \psi) \leqslant k(\phi) + k(\psi) \tag{70}$$

The Euler–Lagrange equation of the problem

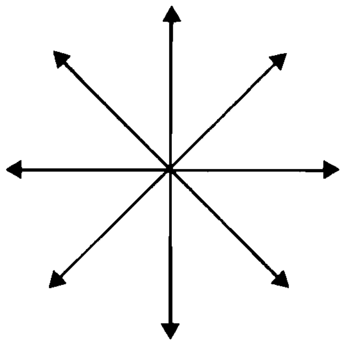

Figure 2.15. Possible gradient directions.

$$\min_{\phi} \int\int k(\phi, x_1, x_2)\,dx_1\,dx_2 \tag{71}$$

subject to

$$\operatorname{div}\phi + q = 0 \qquad \text{in } A \tag{72}$$

$$\phi_n = 0 \qquad \text{on } \partial A \tag{73}$$

is now

$$\frac{\partial k}{\partial \phi} = \operatorname{grad}\lambda \tag{74}$$

The Euler theorem for homogeneous functions states that

$$\phi \cdot \frac{\partial k}{\partial \phi} = k \tag{75}$$

Therefore

$$\begin{aligned} k(\phi) + \lambda[\operatorname{div}\phi + q] &= \phi \cdot \frac{\partial k}{\partial \phi} + \lambda \operatorname{div}\phi + q\lambda \\ &= \phi \cdot \operatorname{grad}\lambda + \lambda \operatorname{div}\phi + q\lambda \qquad \text{using (74)} \\ &= \operatorname{div}(\lambda\phi) + q\lambda \end{aligned}$$

so that

$$\begin{aligned} \int\int k(\phi) + \lambda[\operatorname{div}\phi + q]\,dx_1\,dx_2 &= \int\int \operatorname{div}(\lambda\phi) + q\lambda\,dx_1\,dx_2 \\ &= \int \lambda\phi_n\,ds + \int\int q\lambda\,dx_1\,dx_2 \end{aligned}$$

or

$$\int\int k(\phi)\,dx_1\,dx_2 = \int\int q\lambda\,dx_1\,dx_2 \tag{76}$$

by the Gauss integral theorem with the boundary condition as before. Moreover, for any vector ψ satisfying

$$\operatorname{div}\psi + q = 0 \qquad \text{in } A \tag{72}$$

$$\psi_n = 0 \qquad \text{on } \partial A \tag{73}$$

one has

$$\begin{aligned} k(\psi) &= k(\psi) + \lambda[\operatorname{div}\psi + q] \\ &= \psi\cdot\frac{\partial k}{\partial\psi} + \lambda\operatorname{div}\psi + q\lambda \\ &= \psi\cdot\left(\frac{\partial\psi}{\partial k} - \operatorname{grad}\lambda\right) + \operatorname{div}(\lambda\psi) + q\lambda \end{aligned}$$

Integrating

$$\begin{aligned} \int\!\!\int k(\psi)\mathrm{d}x_1\mathrm{d}x_2 &= \int\!\!\int \psi\cdot\left(\frac{\partial k}{\partial\psi} - \operatorname{grad}\lambda\right) + q\lambda + \operatorname{div}(\lambda\psi)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int\!\!\int \psi\cdot\left(\frac{\partial k}{\partial\psi} - \operatorname{grad}\lambda\right) + q\lambda\,\mathrm{d}x_1\mathrm{d}x_2 + \int(\lambda\psi)_n\,\mathrm{d}s \end{aligned}$$

by the Gauss integral theorem. The boundary integral vanishes by the boundary condition. Thus

$$\int\!\!\int k(\psi)\mathrm{d}x_1\mathrm{d}x_2 = \int\!\!\int k(\psi) - \psi\cdot\operatorname{grad}\lambda + q\lambda\,\mathrm{d}x_1\mathrm{d}x_2 \tag{77}$$

Now define

$$\lambda_\psi = \frac{\psi}{|\psi|}\operatorname{grad}\lambda \tag{78}$$

to be the directional derivative of λ in the direction of ψ.

Lemma: *For all ψ satisfying (72) and (73) and all λ such that*

$$\lambda_\psi \leqslant \frac{k(\psi)}{|\psi|} \tag{79}$$

one has

$$\int\!\!\int k(\psi)\,\mathrm{d}x_1\mathrm{d}x_2 \geqslant \int\!\!\int \lambda q\,\mathrm{d}x_1\mathrm{d}x_2 \tag{80}$$

provided $k(\psi)$ is linear homogeneous.

Proof: From (77)

$$\int\int k(\psi)\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int k(\psi) - \psi\cdot\operatorname{grad}\lambda + q\lambda\ \mathrm{d}x_1\,\mathrm{d}x_2$$

Substitute

$$\psi\cdot\operatorname{grad}\lambda = |\psi|\lambda_\psi$$

from (78) in the integral to obtain

$$\int\int k(\psi)\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int k(\psi) - |\psi|\lambda_\psi + q\lambda\ \mathrm{d}x_1\,\mathrm{d}x_2$$

$$\geqslant \int\int q\lambda\ \mathrm{d}x_1\,\mathrm{d}x_2$$

using (79). *Q.E.D.* Therefore

Theorem:

$$\min_{\phi}\int\int k(\phi)\mathrm{d}x_1\,\mathrm{d}x_2 = \max_{\lambda}\int\int \lambda q\ \mathrm{d}x_1\,\mathrm{d}x_2 \tag{81}$$

$$\operatorname{div}\phi + q = 0 \quad \textit{in } A \qquad\qquad \lambda_\psi \leqslant \frac{k(\psi)}{|\psi|}$$

$$\phi_n = 0 \quad \textit{on } \partial A$$

This is the duality theorem for a general linear homogeneous metric.

In the same way the uniqueness proof may be extended. Thus the mathematics of the simple continuous transportation model extends in a natural way to all continuous transportation models with linear homogeneous metrics. These include

$$k(\psi) = (\phi' K\phi)^{1/2} \qquad K \text{ positive definite: Riemann metric} \tag{82}$$

$$k(\psi) = [k_1|\psi_1|^\omega + k_2|\psi_2|^\omega]^{1/\omega} \quad \text{Minkowski metric } \omega \geqslant 1 \tag{83}$$

$$k(\psi) = k_1|\phi_1| + k_2|\phi| \qquad \text{Manhattan metric} \tag{84}$$

These metrics may be illustrated by isovectures, the curves of equal distance from a given point. By virtue of linear homogeneity all isovectures are enlarged versions of the unit isovecture. Figure 2.16 gives some examples.

In fact, by means of Hölder's inequality it may be seen that the Minkowski isovectures for $\omega_2 < \omega_1$ are located inside the isovecture for ω_1 (except for four points of contact), and that all unit isovectures are convex and fall between the tilted square of diagonal length 2 and the upright square of side length 2 (see Figure 2.17).

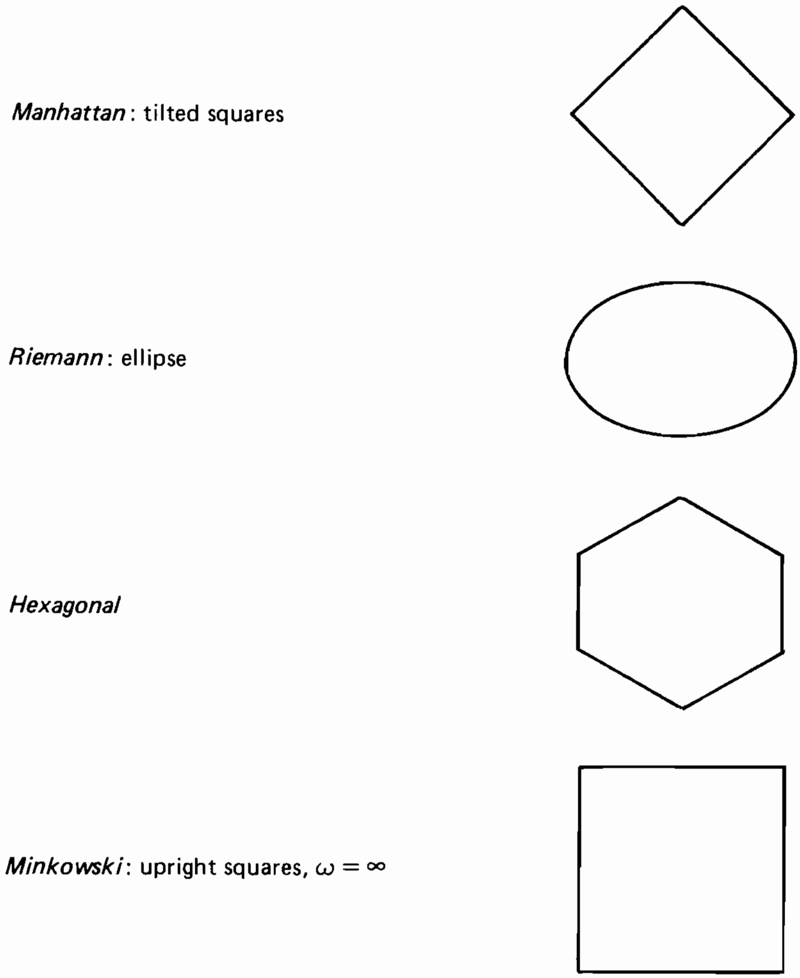

Figure 2.16. Isovectures corresponding to various metrics.

We may also mention here the special metric for water movement

$$k(\phi, x_1, x_2) = \tfrac{1}{2}(\mu \cdot \phi + |\mu|\,|\phi|) \tag{85}$$

$$= \begin{cases} 0 & \text{if } \dfrac{\phi}{|\phi|} = -\dfrac{\mu}{|\mu|} \\ |\mu| \cdot |\phi| & \text{if } \dfrac{\phi}{|\phi|} = \dfrac{\mu}{|\mu|} \end{cases}$$

Thus no cost is incurred in the direction $-\mu$ of natural water flow, while maximum cost is incurred in the opposite direction.

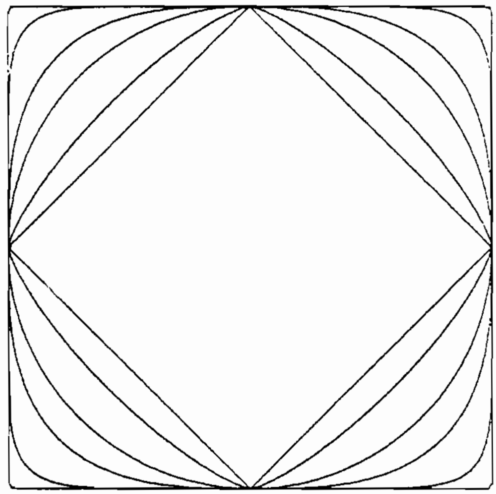

Figure 2.17. Isovectures for Minkowski metrics.

2.4.3 Transportation Cost as a Power Function of Flow

A special feature of the problem solution in the standard case is that the gradient of the flow field bears no relationship to the strength of the flow. This is, however, no longer true when transportation cost is nonlinear with respect to flow density $|\phi|$

$$k = k(x_1, x_2, |\phi|) \tag{86}$$

Consider, in particular, the "Cobb–Douglas" case where k is a power function of $|\phi|$

$$k = k(x_1, x_2) \cdot |\phi|^b \qquad b > 1 \tag{87}$$

Unit cost is an increasing function of flow if and only if $b > 1$. The case of decreasing unit cost is considered in Section 2.4.9. Total transportation cost is a convex function of flow if $b \geqslant 1$. The gradient law assumes the form

$$bk|\phi|^{b-1} \cdot \frac{\phi}{|\phi|} = \text{grad}\,\lambda$$

or

$$bk|\phi|^{b-2}\phi = \text{grad}\,\lambda \tag{88}$$

Taking the norm in (88) one has

$$bk|\phi|^{b-1} = |\text{grad}\,\lambda|$$

It follows from this that the volume of flow increases with the gradient of price provided $b > 1$. This increasing flow entails an increasing marginal cost until the balance between $|\text{grad}\ \lambda|$ and $|\phi|$ is established. It follows furthermore that there is a positive flow wherever local prices differ, for a positive value of $|\text{grad } \lambda|$ implies a nonvanishing flow level $|\phi|$. The flow field vanishes only where prices are uniform.

2.4.4 General Convex Transportation Cost: Congestion

The case where unit transportation cost increases with the volume of flow $|\phi|$ is usually associated with congestion. As the traffic volume $|\phi|$ rises, the unit cost of transportation $k(|\phi|)$ increases and does so at an increasing rate

$$k'(|\phi|) > 0$$

$$k''(|\phi|) > 0$$

Total cost of transportation is then an increasing and convex function of flow

$$\frac{\mathrm{d}}{\mathrm{d}|\phi|}\,[k(|\phi|)\cdot|\phi|] \;=\; k + k'\cdot|\phi| > 0$$

$$\frac{\mathrm{d}}{\mathrm{d}|\phi|^2}\,[k(|\phi|)\cdot|\phi|] \;=\; 2k' + k''\cdot|\phi| > 0$$

Efficient transportation requires that the gradient law be satisfied

$$[k + k'\,|\phi|]\cdot\frac{\phi}{|\phi|} \;=\; \text{grad } \lambda \tag{89}$$

This is the Euler–Lagrange condition of the calculus of variations problem

$$\min_{\phi}\int\!\!\int k(|\phi|)\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

subject to (72). Now the term $k'|\phi|$ is not experienced by travelers unless it is imposed on them as a "congestion toll charge." Otherwise the traveler merely bears the cost $k(|\phi|)$. Here k is average cost and $k + k'|\phi|$ is marginal cost. Notice that, for increasing k, marginal cost is larger than average cost.

A distortion of flow away from the optimum occurs when the difference between marginal cost and average cost is not levied as an extra charge on users of the transportation system. The only exception to this is when marginal cost and average cost are strictly proportional. This is true if and only if transportation cost is a power function of flow

$$k(|\phi|) \;=\; k\cdot|\phi|^b$$

for then

$$\frac{Mc}{Ac} = \frac{bk|\phi|^{b-1}}{k|\phi|^{b-1}} \;=\; b$$

is constant, with Mc denoting marginal cost and Ac average cost. As long as demand is fixed, the optimal routing of traffic is thus not distorted under a power-function relationship between cost and flow. With price (cost) dependent demand, there will, however, be too much traffic generated since $b > 1$.

The most important aspect of increasing unit cost is the following.

Lemma: *Suppose $k(0) = 0$ and $k'(|\phi|) > 0$. Then $\phi = 0$ if and only if* grad $p = 0$.

It follows from this that all singularities are at stationary points of the potential function λ. In particular, a reversal of direction of flow occurs only at points (or along lines) where flow vanishes. Therefore

Corollary: *ϕ is continuous and λ is continuously differentiable in A.*

By contrast, when $k(0) > 0$ any change of direction of ϕ is associated with a discontinuity in ϕ. In particular, the boundaries of market or supply areas constitute line singularities of the flow field. The potential function λ is not differentiable on the singular lines.

2.4.5 Quadratic Transportation Cost: Poisson's Equation

The quadratic case

$$k(\phi, \underline{x}) = k|\phi|^2 \qquad k = \text{constant}$$

plays a special part in the physics of heat transport and diffusion. It can be argued that it is also also relevant to the economics of migration (Hotelling 1929, Beckmann 1957, Tobler 1975); the volume of migration is proportional to the incentive to migrate, i.e. the wage gradient. For these reasons a closer study of this particular specification is in order.

Consider first the case where the cost still depends on location

$$k(\phi, \underline{x}) = k(\underline{x})\,|\phi|^2$$

The efficiency condition (88) assumes the form

$$2k(\underline{x})\phi = \operatorname{grad}\lambda$$

Substituting this into the divergence law, the flow variable ϕ is eliminated and a second-order partial differential equation in λ is obtained

$$\operatorname{div}\left[\frac{1}{k(\underline{x})}\operatorname{grad}\lambda\right] + q = 0 \qquad \text{in } A \tag{90}$$

The boundary condition takes the form

$$(\operatorname{grad}\lambda)_n = 0 \qquad \text{in } \partial A \tag{91}$$

These equations are in terms of gradient λ. It is clear that λ is determined only up to an additive constant.

The most interesting case is that when k = constant = 1/2 (say), i.e. it is independent of location. Equation (90) then becomes Poisson's equation

$$\Delta\lambda + q = 0 \qquad \text{in } A \tag{92}$$

$$(\text{grad }\lambda)_n = 0 \qquad \text{on } \partial A \tag{91}$$

or in component form

$$\frac{\partial^2\lambda}{\partial x_1^2} + \frac{\partial^2\lambda}{\partial x_2^2} + q(x_1, x_2) = 0 \qquad \text{in } A \tag{92a}$$

$$\frac{\partial\lambda}{\partial n} = 0 \qquad \text{on } \partial A \tag{91a}$$

Since $k|\phi|^b$, with $b > 1$, is a strictly convex function of ϕ, the solution of (91), (92) is always unique.

In the theory of electrostatics the term is stated as

$$q = 2\pi e$$

where e represents the density of electric charge and λ is the electric potential of the field. The boundary condition

$$(\text{grad }\lambda)_n = \frac{\partial\lambda}{\partial n} = 0 \tag{91}$$

specifies that no electric flux (or, by analogy, commodity flow) crosses the boundary ∂A.

2.4.6 Vanishing Divergence: Potential Theory

This case is introduced here for its mathematic rather than its economic interest. Assume, as in the export–import problem of Section 2.3.4, that

$$\text{div }\phi \equiv 0 \qquad \text{in } A \tag{93}$$

$$\phi_n = g \qquad \text{on } \partial A \tag{94}$$

This means that all demand/supply is external to the region. The boundary conditions are feasible if and only if

$$\int g \, ds = 0$$

in view of the Gauss integral theorem applied to (93), (94).

Assume once more $k \equiv 1/2$. The equilibrium conditions then assume the form

$$\Delta\lambda = 0 \qquad \text{in } A \tag{92}$$

$$(\text{grad }\lambda)_n = g$$

or

$$\frac{\partial \lambda}{\partial n} = g \qquad \text{on } \partial A \tag{91a}$$

This is the so-called Neumann problem (Duff and Naylor 1966, p. 139). Alternatively we may specify prices on the boundary

$$\lambda(s) = p(s) \qquad \text{on } \partial A \tag{91b}$$

With the boundary conditions (91b), equation (92) represents the standard problem of potential theory.

Equation (92) is known as Laplace's equation. As Duff and Naylor (1966, pp. 133, 134) put it:

> "Laplace's equation is the most famous and most universal of all partial differential equations. No other single equation has so many deep and diverse mathematical relationships and physical applications. A few leading cases are:
> 1 Theory of functions $f(z)$ of a complex variable $z = x + iy$, and the associated conformal mapping theory.
> 2 Theory of gravitational or Newtonian potentials.
> 3 Electrostatic potentials.
> 4 Potentials of steady current flows (magneto-statics).
> 5 Stationary heat flow problems.
> 6 Potentials of incompressible inviscid fluid flow.
> 7 Probability density in random-walk problems.
> 8 Harmonic and biharmonic potentials in two-dimensional elasticity.
> 9 Water-wave potentials for unsteady motion."

The most important property for economic purposes is that equation (92) implies:

Theorem: $\lambda(\underset{\sim}{x})$ *assumes its maximum value and its minimum value on the boundary.*

This means the absence of supply or market areas with interior centers: these centers must lie on the boundary.

2.4.7 Boundary Points: Single Source and Sink

A problem of the export–import type also arises when a finite number of point singularities is introduced. The classic case is that of a single source and a single sink at given locations, each representing a boundary point. Specifically, let a single source of strength $-q$ be placed at

$$(x_1, x_2) = (-1, 0)$$

and a single sink of strength q at

$$(x_1, x_2) = (1, 0)$$

The solution of the Neumann problem is then

$$\ln \frac{1}{((x_1+1)^2+x_2^2)^{1/2}} - \ln \frac{1}{((x_1-1)^2+x_2^2)^{1/2}} = \ln \lambda$$

or

$$\ln \frac{((x_1+1)^2+x_2^2)^{1/2}}{((x_1-1)^2+x_2^2)^{1/2}} = \ln \lambda$$

$$(x_1+1)^2+x_2^2 = \lambda^2(x_1-1)^2+\lambda^2 x_2^2$$

$$\left[x_1+\frac{1+\lambda^2}{1-\lambda^2}\right]^2+x_2^2 = \left(\frac{1+\lambda^2}{1-\lambda^2}\right)^2-1$$

This represents a family of circles (see Figure 2.18).

2.4.8 Duality for Quadratic Transportation Cost

Consider

$$\min_{\phi} \int\int \tfrac{1}{2}\, k|\phi|^2 + \lambda(\operatorname{div}\phi + q)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

which is solved by

$$\phi = \frac{1}{k}\operatorname{grad}\lambda \tag{95}$$

Substituting the solution in the integral

$$\begin{aligned} V &= \int\int \tfrac{1}{2}\phi\cdot\operatorname{grad}\lambda + \lambda\operatorname{div}\phi + q\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int\int -\tfrac{1}{2}\,\phi\cdot\operatorname{grad}\lambda + \operatorname{div}(\lambda\phi) + q\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int (\lambda\phi)_n\,\mathrm{d}s + \int\int q\lambda - \tfrac{1}{2}\phi\cdot\operatorname{grad}\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2 \end{aligned} \tag{96}$$

where the line integral vanishes by the boundary condition. Finally

$$V = \int\int q\lambda - \frac{1}{2k}|\operatorname{grad}\lambda|^2\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{97}$$

Consider now the problem of maximizing the concave function (97) with respect to λ. This (unconstrained) calculus of variations problem is solved by

$$q = -\operatorname{div}\left(\frac{1}{k}\operatorname{grad}\lambda\right) \tag{98}$$

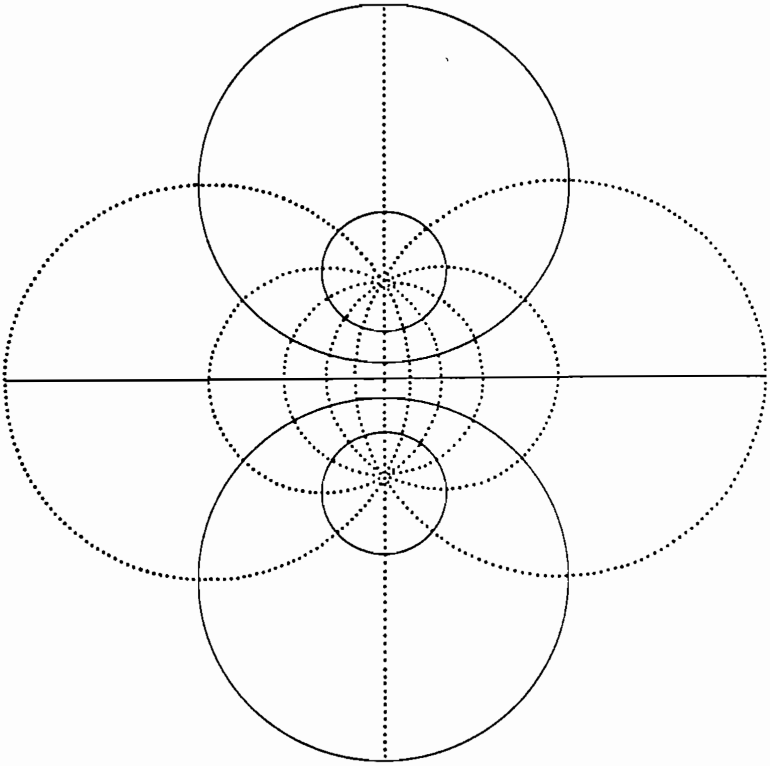

Figure 2.18. Flow lines and potential lines in the dipole: one source and one sink.

Utilizing (95), one sees that the constraint

$$\operatorname{div} \phi + q = 0 \tag{72}$$

is satisfied. The free boundary condition is

$$-\frac{1}{k}(\operatorname{grad} \lambda)_n = 0 \quad \text{or} \quad \phi_n = 0 \qquad \text{on } \partial A \tag{91}$$

We have thus shown:

Lemma: *The dual of the transportation cost minimization problem with quadratic transportation cost function is the unconstrained problem*

$$\max_{\lambda} \int\int q\lambda - \frac{1}{2k}(\operatorname{grad} \lambda)^2 \;\mathrm{d}x_1\,\mathrm{d}x_2 \tag{99}$$

Consider the special case of an export–import problem or a transportation problem with discrete sources and sinks. Then (99) takes the form

$$\min \iint (\text{grad } \lambda)^2 \, dx_1 dx_2 \tag{100}$$

subject to the boundary conditions

$$\text{grad } \lambda_n = g \qquad \text{on } \partial A \tag{101}$$

Problem (100), (101) is known as Dirichlet's principle and plays a major role in potential theory (Courant and Hilbert 1953).

2.4.9 Remarks on Concave Transportation Costs

We note in passing that continuous flow is not consistent with strict concavity of transportation costs with respect to volume of flow. In that case a channeling of flow would take place so that separate flow paths of high density would emerge. This is in fact what happens to a water flow in nature when it is channeled into a discrete network of watercourses. Notice, however, that a fine structure of water channels may still be approximated by a macroscopic description in terms of a flow field whose local direction is that of the water channels and whose density is that of the water run-through per unit area.

2.5 FURTHER EXTENSIONS

2.5.1 Mappings of the Solution Space

In some cases it may make it easier to visualize the flow lines that represent integral curves to a Euler equation, or even simplify the solution process, if we map the actual (flat) region studied homeomorphically onto another (flat or curved) surface. On this latter surface, every point of the actual region has a unique image point, and every point on the surface represents a unique point of the region. In the same manner, flow lines in the original region are mapped one-to-one onto curves on the new surface.

We now require from this surface that, for any image of a trajectory, the arc length on the surface must equal the transportation cost along the original trajectory in the region. As optimal routes reduce to a minimum the transportation cost for any pair of endpoints, we conclude that their images on the surface are geodesics, i.e. curves of minimum length or the "straightest" curves that exist on a (possibly) curved surface.

Wardrop (1969) has explored the possibilities of using complex analytic functions for flat representation so that the optimal routes are mapped onto straight-line segments, whereas Angel and Hyman (1976), following a conjecture by Warntz (1967), have investigated the possibility of mappings onto curved surfaces of revolution, cones, cylinders, and spheres. We will treat this problem in terms of Gaussian differential geometry as this simplifies the general discussion.

As before, we denote the actual region by R and let S denote the (possibly) curved image. It is, of course, two-dimensional and can be embedded in ordinary three-dimensional Euclidean space. Denote a point in this space $\xi = (\xi^1, \xi^2, \xi^3)$. We obtain the most natural parameterization of S by using three coordinate functions $\xi^1(x_1, x_2)$, $\xi^2(x_1, x_2)$, and $\xi^3(x_1, x_2)$, using the original Cartesian coordinates x_1 and x_2 for points in R as parameters. These functions immediately allow us to compute the image of any point x_1, x_2 and of any curve $x_1(s), x_2(s)$ in R.

Let us now denote by ξ_1 and ξ_2 the vectors $(\partial\xi^1/\partial x_1, \partial\xi^2/\partial x_1, \partial\xi^3/\partial x_1)$ and $(\partial\xi^1/\partial x_2, \partial\xi^2/\partial x_2, \partial\xi^3/\partial x_2)$. By keeping one of the parameters x_1, x_2 constant at various values and varying the other, we get a net of coordinate curves covering the whole surface S. Then ξ_1 and ξ_2 are tangent vectors along these coordinate curves. By taking products of these tangent vectors we obtain a set of interesting expressions. Starting with the "dot" or scalar products we get the Gaussian first fundamental coefficients, traditionally denoted

$$E = \xi_1 \cdot \xi_1, F = \xi_1 \cdot \xi_2, G = \xi_2 \cdot \xi_2 \tag{102}$$

These completely define the metric structure of the surface S in terms of distances, areas, and angles. An arc length element along the image of a parameterized curve $x_1(s), x_2(s)$ is

$$[E(\mathrm{d}x_1/\mathrm{d}s)^2 + 2F(\mathrm{d}x_1/\mathrm{d}s)(\mathrm{d}x_2/\mathrm{d}s) + G(\mathrm{d}x_2/\mathrm{d}s)^2]^{1/2}\,\mathrm{d}s \tag{103}$$

The formula is easily verified by using the chain rule. As transportation cost for an infinitesimal displacement is $k\mathrm{d}s$, we require the square root in (103) to equal k in order that distance on the surface be equal to transportation cost. Assuming s to be an arc length parameter, by which we lose no generality, we have $(\mathrm{d}x_1/\mathrm{d}s)^2 + (\mathrm{d}x_2/\mathrm{d}s)^2 \equiv 1$ and can put $\mathrm{d}x_1/\mathrm{d}s = \cos\theta$, $\mathrm{d}x_2/\mathrm{d}s = \sin\theta$. Hence

$$E\cos^2\theta + 2F\cos\theta\sin\theta + G\sin^2\theta = k^2 \tag{104}$$

If local transportation cost k depends only on the location coordinates x_1, x_2 but not on the direction of passage θ (so that we are dealing with an isotropic problem), then (104) must hold for all θ with a constant right-hand side. Putting θ equal to 0 and $\pi/2$, in turn, yields

$$E = G = k^2 \tag{105}$$

Substitution from (105) into (104) yields

$$2F\cos\theta\sin\theta = 0$$

As this must be true for all θ, we conclude that

$$F = 0 \tag{106}$$

Equation (106) tells us that the coordinate lines on S intersect orthogonally, like the coordinate lines for x_1, x_2 in R, so that angles of intersecting curves are not altered. Equations (105) tell us that the linear magnification factors are equal in the two coordinate directions, and in fact in all other directions as well. This means that the mapping is conformal and hence that the surface S must be conformal to the plane.

We now turn to the "cross" or vector product $\xi_1 \times \xi_2$. Unlike the scalar product this is a vector, orthogonal to both ξ_1 and ξ_2. As $\xi_1 \times \xi_2$ is thus orthogonal to both coordinate lines it is orthogonal to the tangent plane and hence has the interpretation of a normal vector to S. By normalizing to unit length we get the unit normal vector

$$\xi_1 \times \xi_2 / |\xi_1 \times \xi_2| \tag{107}$$

This will be needed in defining curvature, but we first note that $|\xi_1 \times \xi_2| dx_1 dx_2$ is the area of the image on S of the rectangle $dx_1 dx_2$ in R. Accordingly, $|\xi_1 \times \xi_2|$ is the areal magnification factor. By an identity in vector algebra $|\xi_1 \times \xi_2|^2 \equiv (\xi_1 \cdot \xi_1)(\xi_2 \cdot \xi_2) - (\xi_1 \cdot \xi_2)^2$, which by (102) equals $EG - F^2$. So

$$(EG - F^2)^{1/2} \tag{108}$$

is the areal magnification factor of the mapping.

Owing to (105), (106) we get $(EG - F^2)^{1/2} = k^2$. Thus areal magnification is completely determined by the transformation needed to make the images of optimal routes geodesics. It is not possible to employ yet another transformation to make some areal densities constant, as was conjectured by Bunge (1962). This was noted by Angel and Hyman (1976).

Besides the metric structure of a surface its curvature structure is also important. If we denote by ξ_{11}, ξ_{12}, and ξ_{22} the three-component vectors of second partial derivatives taken with respect to the parameters indicated by the subscripts, then the scalar products of these with the unit normal vector

$$L = \xi_{11} \cdot \xi_1 \times \xi_2 / |\xi_1 \times \xi_2| \tag{109}$$

$$M = \xi_{12} \cdot \xi_1 \times \xi_2 / |\xi_1 \times \xi_2| \tag{110}$$

$$N = \xi_{22} \cdot \xi_1 \times \xi_2 / |\xi_1 \times \xi_2| \tag{111}$$

using the traditional notation, are the second fundamental coefficients.

Together with the first fundamental coefficients, these second fundamental coefficients determine the curvature structure, and in fact, if E, F, G, L, M, and N are known functions of x_1, x_2, they supply sufficient information to determine the surface completely, except for rigid translations and rotations in the surrounding space. This is true provided only that such a surface exists, and for this a couple of integrability conditions are required. These conditions impose constraints on the fundamental coefficients. As we have seen, E, F, and G are determined by the k function. On the other hand we can choose L, M, and N as we wish. The question is whether it is possible to choose them so that the compatibility conditions are fulfilled. If not, the idea of mapping the solution space lacks meaning, except in a few special cases.

The curvature of the surface in a normal section with direction dx_2/dx_1 (as referred to the parameter plane) is

$$\frac{L(dx_1)^2 + 2M(dx_1)(dx_2) + N(dx_2)^2}{E(dx_1)^2 + 2F(dx_1)(dx_2) + G(dx_2)^2} \tag{112}$$

and this expression takes, at any point, a maximum and a minimum value in two specific directions. These curvatures are called the principal curvatures. If they are of equal sign then the surface lies on one side of its tangent plane in the neighborhood of the point of tangency. If they have opposite signs the surface crosses the tangent plane at the point of tangency. Umbilical points are points where the curvature is equal in all directions. In the conformal case, we conclude that the numerator of (112) must then be constant, as the denominator is, and hence $L = N$ and $M = 0$. A sphere is a surface with exclusively umbilical points. On the other hand, all normal curvatures must be zero for a plane surface and hence $L = M = N = 0$.

The product of the principal curvatures is called the Gaussian curvature and has the simple formula

$$\kappa = (LN - M^2)/(EG - F)^2 \tag{113}$$

There is a remarkable theorem by Gauss, which he himself called the "*teorema egregium*," which states that, even though L, M, and N can to some extent be chosen freely, the expression $LN - M^2$ is determined by E, F, G, and their derivatives. Since we have seen that the latter are determined by k in the conformal case, we get the simple expression

$$LN - M^2 = -k^2 \text{ div grad } (\ln k) \tag{114}$$

by the *teorema egregium*. This is one of the integrability equations. It must be noted that the general form is much more complicated than that for the conformal case stated here.

The other two conditions, the Mainardi–Codazzi equations, are obtained by considering that various mixed third-order derivatives are invariant with respect to the order of derivation. They too are particularly simple for the conformal case and read

$$N_1 - M_2 = (L + N)k_1/k \tag{115}$$

and

$$L_2 - M_1 = (L + N)k_2/k \tag{116}$$

where the subscripts denote partial derivatives with respect to x_1 and x_2.

We can now see how these formulas can be used. Let us first discuss the question: in which cases can the surface S be a plane or a sphere? For the plane, $L = M = N = 0$, as we have seen, and hence (115) and (116) are trivially fulfilled. Then (114) reads div grad $(\ln k) = 0$. Assuming that k is a function of $r = (x_1^2 + x_2^2)^{1/2}$ only, the solution is $k = \alpha r^\beta$. This agrees with Wardrop's (1969) conclusion that the only cost functions with circular symmetry that admit plane transformations by analytic complex functions are of the power variety. It should be observed that the conclusion applies not only to genuine planes but to surfaces like cones and cylinders as well, where Gaussian curvature is zero.

For the spherical surface we have noted that all points are umbilical, i.e. $L = N$ and $M = 0$. Applied to equations (115) and (116), this information gives $N_1/N = 2k_1/k$ and $L_2/L = 2k_2/k$. These partial differential equations have a possible solution $L = N = \kappa^{1/2} k$. Substituting into (114) we get div grad $(\ln k) = -\kappa\, k^2$. Assuming again that k is a function of r only, we get the solution $k = 1/(\alpha + \beta r^2)$. This again agrees with a result of Angel and Hyman (1976). In this case the mapping could help us to find the solution. The geodesics on a sphere are great circles and their stereographic projections on a plane (which is a conformal mapping) are circles or straight lines. Finding these routes directly from the complicated k function could be quite difficult.

Let us close this section by saying something about the existence of a suitable surface S. Bonnard's fundamental theorem says that if E, F, G are any C^2 functions of x_1, x_2 and L, M, N any C^1 functions of x_1, x_2 satisfying the Gauss and Mainardi–Codazzi equations, then there exists a unique corresponding surface. In our case, we have a free choice of L, M, N as long as the equations are not violated. As $EG - F^2 = k^2 > 0$, which is required for the general theorem as well, it has been demonstrated that the second coefficients can always be so chosen that the equations (114)–(116) are fulfilled. Hence the surface always exists (Guggenheimer 1963).

2.5.2 Bounded Flow

Suppose that there is a capacity limit

$$c = c(x_1, x_2)$$

on flow through point (x_1, x_2)

$$|\phi| \leqslant c(x_1, x_2) \tag{117}$$

Then the cost-minimization problem is further restricted by (117). The Lagrangean assumes the form

$$\iint L \, \mathrm{d}x_1 \, \mathrm{d}x_2 = \iint - k|\phi| - \lambda[q + \operatorname{div} \phi] + \mu[c - |\phi|] \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

Maximizing this concave Lagrangean yields

$$(k + \mu) \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \phi \neq 0 \tag{118}$$

$$k + \mu \geqslant |\operatorname{grad} \lambda| \qquad \phi = 0 \tag{119}$$

$$\left.\begin{array}{l} \mu = 0 \longleftrightarrow |\phi| < c \\ \mu \geqslant 0 \longleftrightarrow |\phi| = c \end{array}\right\} \tag{120}$$

When no flow is permitted through a certain point, μ can be positive even where flow vanishes.

Example: Consider a single source and a single sink. The line connecting the points passes through a rectangular area where capacity is restricted. Then, in order to induce flow to diversify and even use a northern bypass route, tolls μ have to be charged for any flow passing through the area of restricted capacity (Figure 2.19). Another example is that shown in Figure 2.20.

2.5.3 Max Flow–Min Cut

The two-dimensional version of a famous network capacity problem (Ford and Fulkerson 1962) is best illustrated by water flow through a river bed (Figure 2.20). The river is bounded by two edges that are considered piecewise smooth. At each point

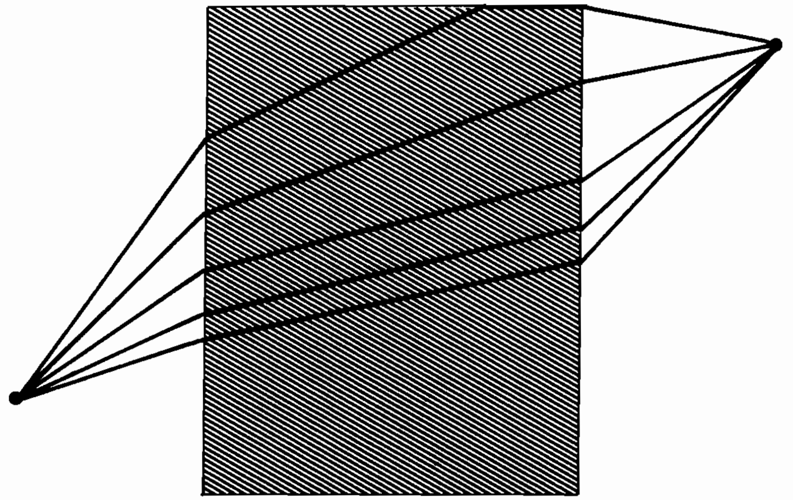

Figure 2.19. A region of bounded flow.

in the river there is a capacity $c(x_1, x_2)$ limiting the flow. Suppose that the capacity limits are applicable over a finite length of the river bed between curves C and D. The object is to maximize total flow through curves C or D.

The Lagrangean of this problem is

$$\int_D \phi_n \, ds + \int\int -\lambda [0 + \operatorname{div} \phi] + \mu(c - |\phi|) dx_1 \, dx_2 \tag{121}$$

giving rise to the efficiency condition

$$\operatorname{grad} \lambda = \mu \frac{\phi}{|\phi|} \tag{122}$$

The flow is a gradient flow. The potential function λ has constant values on C and D. These we may standardize

$$\lambda = 0 \text{ on } C \qquad \lambda = 1 \text{ on } D \tag{123}$$

and all isopotential curves in the bottleneck section are "min cut" lines. This means that the capacity limits on through flow are effective on each potential line considered as a cross section. As a degenerate case we have the possibility of a single potential curve, representing a singular bottleneck.

2.5.4 Finite Approximation

For purposes of calculation, a continuous region must be subdivided into a set of cells, and the continuous flow field must be approximated by a discrete set of vectors, one associated with each cell. For convenience let

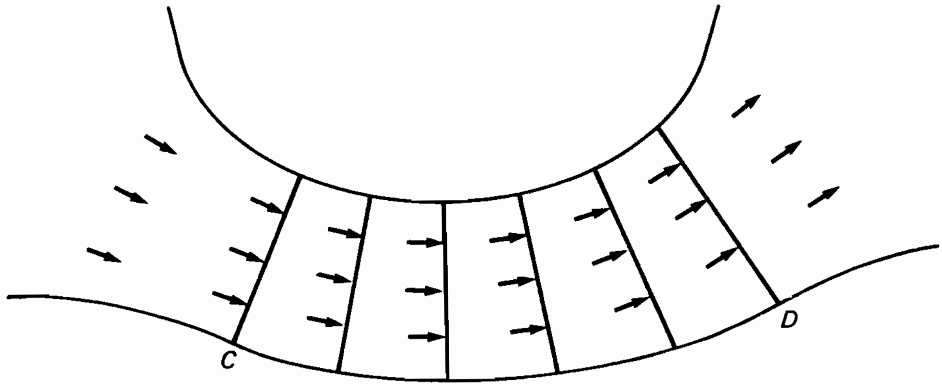

Figure 2.20. Flow through a bottleneck.

$$x_1 = ia \qquad i = 0,1,\ldots,m \tag{124}$$

$$x_2 = jb \qquad j = 0,1,\ldots,n \tag{125}$$

be such a discretization of the continuous variables x_i. Then rewrite the flow vector

$$\phi(x_1,x_2) = \phi(i,j) = (\phi_1(i,j), \phi_2(i,j)) \tag{126}$$

To obtain the divergence law, consider the outflow from and inflow into cell i, k across its four boundary lines (see Figure 2.21)

$$-q(i,j) = \text{net outflow from } (i,j)$$

$$-q(i,j)ab = [\phi_1(i+1,j) - \phi_1(i,j)]b + [\phi_2(i,j+1) - \phi_2(i,j)]a \tag{127}$$

The cost-minimization problem becomes

$$\min \sum_{i=0}^{m} \sum_{j=1}^{n} k(\phi_1^2(i,j) + \phi_2^2(i,j))^{1/2} \tag{128}$$

subject to (127). The efficiency conditions of this convex nonlinear program with linear constraints are then

$$\begin{aligned} \frac{k}{b}\,\frac{\phi_1(i,j)}{(\phi_1^2(i,j) + \phi_2^2(i,j))^{1/2}} &= \lambda(i,j) - \lambda(i-1,j) \\ \frac{k}{a}\,\frac{\phi_2(i,j)}{(\phi_1^2(i,j) + \phi_2^2(i,j))^{1/2}} &= \lambda(i,j) - \lambda(i,j-1) \end{aligned} \tag{129}$$

When $a = b$ then (129) simplifies to

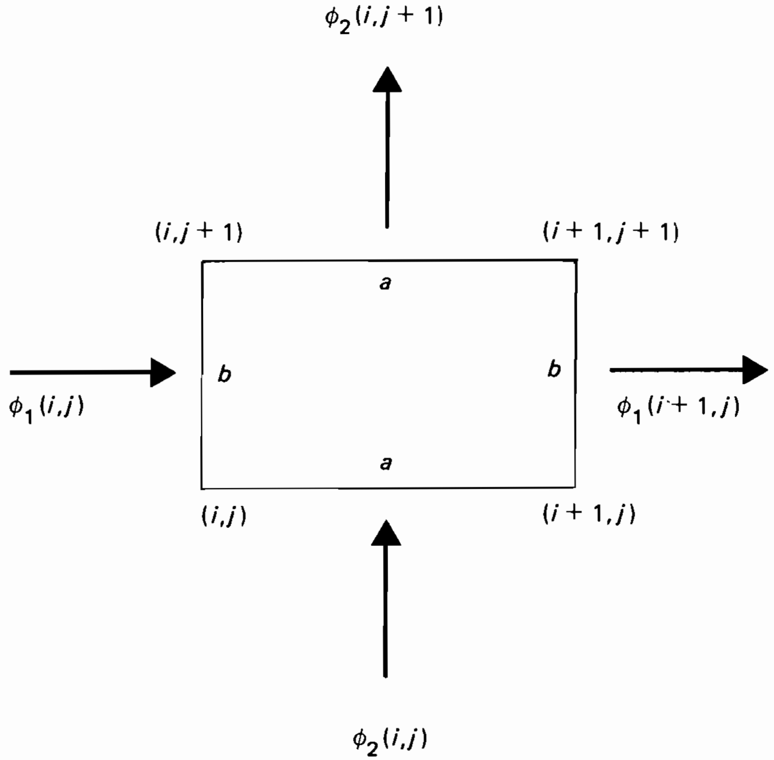

Figure 2.21. Flow through a rectangular cell.

$$
\begin{aligned}
\frac{k\,\phi_1(i,j)}{(\phi_1^2(i,j)+\phi_2^2(i,j))^{1/2}} &= \lambda(i,j)-\lambda(i-1,j) \\
\frac{k\phi_2(i,j)}{(\phi_1^2(i,j)+\phi_2^2(i,j))^{1/2}} &= \lambda(i,j)-\lambda(i,j-1)
\end{aligned}
\tag{130}
$$

where λ is now a times the previous λ. When both components ϕ_1, ϕ_2 vanish at some i, j then the left-hand side is undefined.

Equations (127), (129), and (130) are the discrete analogs of the two flow field equations in the continuous case. Observe that, when the ϕ components do not both vanish

$$\phi_1^2 + \phi_2^2 \neq 0$$

the squares of the left-hand side of (130) add up to k^2

$$[\lambda(i,j)-\lambda(i-1,j)]^2 + [\lambda(i,j)-\lambda(i,j-1)]^2 = k^2$$

The finite approximation of the cost function for Manhattan metric is

$$\sum_i \sum_j k_1 |\phi_1(i,j)| + k_2 |\phi_2(i,j)| \tag{131}$$

with the source–sink equation unchanged. The efficiency conditions are

$$\begin{aligned} k_1 \operatorname{sign} \phi_1 &= \lambda(i,j) - \lambda(i-1,j) \\ k_2 \operatorname{sign} \phi_2 &= \lambda(i,j) - \lambda(i,j-1) \end{aligned} \tag{132}$$

The conditions are not defined for $\phi_m = 0$, $m = 1$ and 2. Writing $\phi_m = z_m - y_m$, with $z_m \geqslant 0$ and $y_m \geqslant 0$, and the objective function as

$$\sum_{m=1}^{2} \sum_{i=1}^{m} \sum_{j=1}^{n} k_m \cdot (z_m + y_m)$$

one can show that

$$\left.\begin{aligned} \phi_1 &= 0 \quad \text{whenever } |\lambda(i,j) - \lambda(i-1,j)| < k_1 \\ \phi_2 &= 0 \quad \text{whenever } |\lambda(i,j) - \lambda(i,j-1)| < k_2 \end{aligned}\right\} \tag{133}$$

In a similar way one can show that, for Euclidean metric

$$\phi_1 = \phi_2 = 0$$

whenever

$$[\lambda(i,j) - \lambda(i-1,j)]^2 + [\lambda(i,j) - \lambda(i,j-1)]^2 < k^2 \tag{134}$$

The discretization of the continuous transportation model with Euclidean metric thus leads to a nonlinear program. It should be compared with the linear programming version of the minimum cost flow problem when sources and sinks occur only in isolated locations, viz. the transportation problem (Section 2.5.5). In the transportation problem the underlying metric of transportation costs may be Euclidean or not. The flow variables are no longer interpretable as vector components. We turn now to a brief description of this situation.

2.5.5 Linear Programming Formulation

The linear programming formulation of the spatial market equilibrium problem with given demand and supply is the so-called trans-shipment problem. Denote locations by i, excess demand by q_i, and flows from i to j by x_{ij}. The divergence law takes the form

local excess demand = inflow − outflow

or

$$q_i = \sum_{j \in R_i} x_{ji} - \sum_{j \in S_i} x_{ij} \tag{135}$$

The set R_i is the set of origin locations j from which shipments can reach market location i, while the set S_i is the set of destination locations j to which shipments can go from market location i.

Efficiency of resource allocation in a competitive market results in minimum transportation cost. Let k_{ij} be the cost of shipping one unit of the commodity from i to j. Total transportation cost to be minimized is then

$$\sum_{ij \in T} k_{ij} x_{ij}$$

where T is the set of all admissible connections, i.e. of all links in the transportation network joining locations i and j. T may be constructed from R_i or S_i: $T = \cup R_i = \cup S_i$. The equation–inequality corresponding to the gradient law is the Koopmans price theorem

$$\begin{aligned} \lambda_j - \lambda_i &= k_{ij} \longleftrightarrow x_{ij} > 0 \\ \lambda_j - \lambda_i &\leqslant k_{ij} \longleftrightarrow x_{ij} = 0 \end{aligned} \tag{136}$$

As in the continuous case, the level of the λ is undetermined. The formal similarity is obvious. The flow pattern is now one-dimensional: it is a tree, or can always be made into a tree through the elimination of "neutral circuits" (Koopmans 1949, Koopmans and Reiter 1951). Isopotential lines reduce to sets of isolated points on the flow tree (see Figure 2.22).

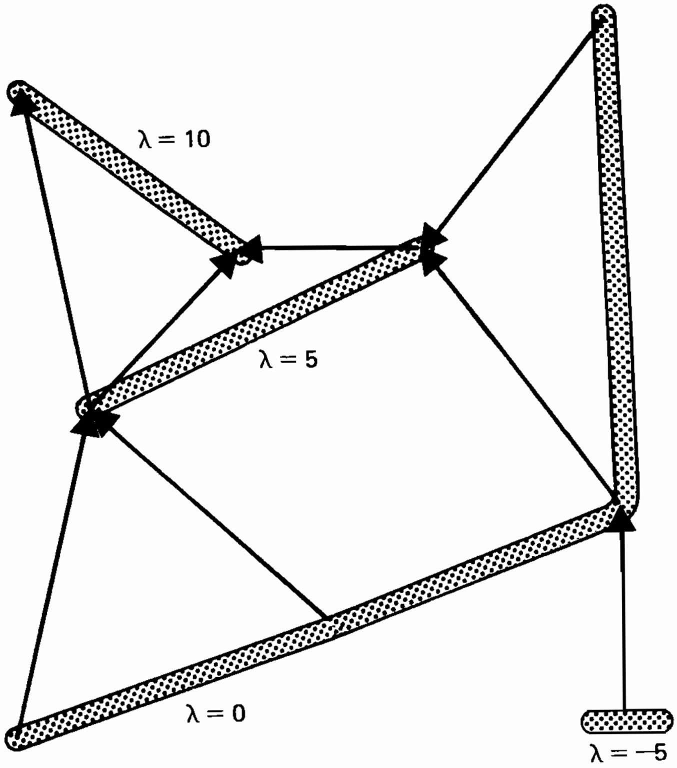

Figure 2.22. Isopotential lines reduce to sets of isolated points on a flow tree.

3 *Short-Run Equilibrium and Stability*

3.1 PARTIAL EQUILIBRIUM OF SPATIAL MARKETS

When studying spatial markets we can distinguish between three broad time periods, in the same way as in general economic theory. In the very short run, a given fixed supply is rationed among competing demanders (see Section 3.1.1). In the intermediate run, the supply of given facilities may be expanded or contracted (see Section 3.2). In the long run, facilities may be introduced at new locations or may depart from old locations (see Chapter 6).

On a formal basis two types of supply function are of particular interest: the vertical supply curve of Section 3.2 resulting from capacity limitations of facilities in place (including land) and the horizontal supply curve of the procurement problem, treated in Section 6.1.1.

3.1.1 Excess Demand Dependent on Price

Recall, so far, that the local price λ has been considered arbitrary up to an additive constant. This was quite natural in the context of supply and demand balanced at an aggregate level. We can immediately add realism to the market model by allowing the excess demand function q to depend on local price

$$q = q(\lambda, x_1, x_2)$$

$$\operatorname{div} \phi(x_1, x_2) + q(\lambda, x_1, x_2) = 0 \tag{1}$$

Assume, in fact, that excess demand is a strictly decreasing function of λ

$$\frac{\partial q}{\partial \lambda} < 0 \qquad \text{all} \quad \lambda, x_1, x_2 \tag{2}$$

We now show that λ is then uniquely determined everywhere. Assume two solutions

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \tag{3}$$

$$k \frac{\psi}{|\psi|} = \operatorname{grad} \mu \tag{4}$$

and consider the boundary integral

$$\int (\lambda - \mu)(\phi_n - \psi_n)\mathrm{d}s = 0$$

since both ϕ and ψ must satisfy the boundary condition. By the Gauss integral theorem

$$0 = \int_{\partial A} (\lambda - \mu)(\phi - \psi)_n \cdot \mathrm{d}s = \iint_A \operatorname{div}[(\lambda - \mu)(\phi - \psi)]\, \mathrm{d}x_1\, \mathrm{d}x_2$$

$$= \iint_A (\phi - \psi)\cdot \operatorname{grad}(\lambda - \mu) + (\lambda - \mu)\operatorname{div}(\phi - \psi)\, \mathrm{d}x_1\, \mathrm{d}x_2$$

Now ϕ vanishes wherever

$$k\frac{\phi}{|\phi|} \neq \operatorname{grad}\lambda \tag{3a}$$

and the same applies to ψ. Since only the nonvanishing ϕ and ψ contribute to the integral, we may replace grad λ and grad μ throughout by the expressions (3) and (4). Thus

$$0 = \iint k(\phi - \psi)\cdot\left(\frac{\phi}{|\phi|} - \frac{\psi}{|\psi|}\right) + (\lambda - \mu)\operatorname{div}(\phi - \psi)\, \mathrm{d}x_1\, \mathrm{d}x_2$$

Substituting (1) in the second term

$$0 = \iint k(\phi - \psi)\cdot\left(\frac{\phi}{|\phi|} - \frac{\psi}{|\psi|}\right) + (\lambda - \mu)\,[-q(\lambda) + q(\mu)]\, \mathrm{d}x_1\, \mathrm{d}x_2$$

Now the first term was shown in Section 2.3.6 (uniqueness) to be nonnegative, and zero only if

$$\frac{\phi}{|\phi|} = \frac{\psi}{|\psi|} \tag{5}$$

Therefore, omitting this term we have

$$0 \geqslant -\iint (\lambda - \mu)\,[q(\lambda) - q(\mu)]\, \mathrm{d}x_1\, \mathrm{d}x_2 \tag{5a}$$

But for any strictly decreasing q

$$(\lambda - \mu)[q(\lambda) - q(\mu)] \leqslant 0$$
$$\text{and } = 0 \qquad \text{only if} \quad \lambda = \mu \tag{6}$$

This may be shown as follows: if $\lambda > \mu$ then

$$q(\lambda) < q(\mu)$$

and if

$$\lambda < \mu$$

then

$$q(\lambda) > q(\mu)$$

In either case the two terms are nonvanishing and have opposite signs. Combining (5a) and (6) one obtains

$$0 \geqslant \int\int (\mu - \lambda)\,[q(\lambda) - q(\mu)]\,\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant 0$$

Except on a set of measure zero it must therefore be true that

$$\lambda(x_1, x_2) \equiv \mu(x_1, x_2)$$

On sets of measure zero the conclusion follows by continuity (for λ is piecewise differentiable and therefore continuous). We have thus just proved:

Theorem: *If excess demand is a strictly decreasing function of price, then any equilibrium price distribution $\lambda(x_1, x_2)$ is unique.*

In spatial market equilibrium, one possible solution is that where markets are balanced locally at prices whose gradients do not permit profitable arbitrage. Then there is no integrated spatial market but merely a system of locally autarchic markets. As transportation cost declines through technical progress such a system may become an integrated spatial market. As a result of this integration the price spread must decrease. In fact, importing locations must now have lower prices and exporting locations higher prices.

In general terms this conclusion arises from the dual transportation problem, where spread is maximized subject to the constraint that gradients may not exceed transportation cost; hence reducing transportation cost necessarily reduces gradients (cf. Section 2.2.9). Note also that the efficiency conditions are linear homogeneous in k. Hence a proportional reduction of all k reduces the λ in the same proportion and hence all differences of λ, in other words the price spread.

3.1.2 Welfare Maximization

As a special case, let the excess demand function

$$q(p, \underset{\sim}{x}) = a(\underset{\sim}{x}) - bp$$

be linear with constant slope $-b$ and an intercept $a(\underset{\sim}{x})$ that depends on location. The consistency condition for a closed region is

$$0 = \int\int q(p, x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int a(x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2 - b\int\int p(x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2$$

from which

$$\bar{p} = \frac{\bar{a}}{b} \tag{7}$$

where

$$\bar{p} = \frac{\iint p(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2}{\iint \mathrm{d}x_1\,\mathrm{d}x_2} = \text{average price}$$

$$\bar{a} = \frac{\iint a(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2}{\iint \mathrm{d}x_1\,\mathrm{d}x_2} = \text{average intercept}$$

Average price is thus proportional to the average intercept. The utility function u (consumers' surplus) is obtained by integrating the demand function

$$p = \frac{a - q}{b}$$

$$u = \int_0^q \frac{a - t}{b}\,\mathrm{d}t = \frac{a}{b}q - \frac{1}{2}\frac{q^2}{b}$$

It is well known that competitive market equilibrium can be described as the result of maximizing the sum of consumers' surplus and profits, sometimes called social surplus. When transfers are cancelled, this is equivalent to maximizing the integral of the demand function minus costs

$$\max_{q,\phi} \iint \left[\frac{a(\underset{\sim}{x})}{b}\,q - \frac{1}{2b}\,q^2 - k\,|\phi| \right] \mathrm{d}x_1\,\mathrm{d}x_2$$

subject to (1). Or, eliminating q by means of (1)

$$\max_{\phi} \iint \left[-\frac{a}{b}\,\mathrm{div}\,\phi - \frac{1}{2b}\,\mathrm{div}^2\phi - k\,|\phi| \right] \mathrm{d}x_1\,\mathrm{d}x_2$$

The Euler equation of this calculus of variations problem is

$$\mathrm{grad}\left[\frac{a}{b} + \frac{\mathrm{div}\,\phi}{b} \right] = k\,\frac{\phi}{|\phi|}$$

or, in view of

$$\mathrm{div}\,\phi = -q = -a + b\lambda$$

$$\lambda = \frac{a}{b} + \frac{\mathrm{div}\,\phi}{b}$$

$$\mathrm{grad}\,\lambda = k\,\frac{\phi}{|\phi|}$$

which is the equilibrium condition (3).

The generalization of this to an arbitrary demand function $q(\lambda, \underset{\sim}{x})$ is as follows. Let

$$\lambda = p(q, \underset{\sim}{x})$$

be the inverse function of q. The consumers' surplus is defined as

$$u(q,\underline{x}) = \int_0^q p(t,\underline{x})\mathrm{d}t \tag{8}$$

consistent with (7). Competitive spatial market equilibrium can then be characterized as the solution of the maximization problem

$$\max_{\phi} \int_A\int \left[\int_0^{-\mathrm{div}\,\phi} p(t,\underline{x})\mathrm{d}t - k\,|\phi|\right]\mathrm{d}x_1\,\mathrm{d}x_2$$

yielding the Euler equation

$$\mathrm{grad}\, p(q,\underline{x}) = k\frac{\phi}{|\phi|}$$

where

$$q = -\,\mathrm{div}\,\phi$$

p now represents the local price, hitherto denoted by λ.

An analogous construction for a discrete spatial market may be found in Samuelson (1952).

3.1.3 Monopolistic Price Policies

So far, supply and demand have been considered to be perfectly competitive. In this section we assume, in contrast, that there is a single supplier. Initially, the supplier operates a single plant, which we assume to be located at the origin $\underline{x} = (0,0)$. Three specific spatial price policies are generally distinguished for special attention out of the set of all possible policies:

- Mill pricing or f.o.b. (free on board): the supplier charges a price p at the mill and the demander bears the transportation cost.
- Uniform pricing or c.i.f. (cost, insurance, and freight): the supplier charges the same price p and absorbs the transportation cost. However, the supplier may refuse service to certain geographical points.
- Perfect discrimination: the supplier charges profit maximizing prices for each location but bears the transportation cost.

These cases have been thoroughly studied in the literature, mainly on the assumptions of a uniform plane and of linear demand curves. The purpose of this section is to formulate these policies for arbitrary transportation cost functions $k(\underline{x})$ and the demand densities $q(\underline{x})$, while retaining for the most part the assumption of linear demand curves. Production cost is assumed proportional to total output. In the case of linear demand it can in fact be ignored (Beckmann 1976b).

It is easiest to begin with perfect discrimination. When charging prices $p(\underline{x})$, local revenue is $p(\underline{x})\cdot q(\underline{x})$ and aggregate revenue is the integral of this expression. Total cost consists of transportation costs and production costs

$$\int\int k|\phi| + c\cdot q \, \mathrm{d}x_1 \mathrm{d}x_2$$

Profits

$$G = \int\int (p-c)q(p) - k|\phi| \mathrm{d}x_1 \mathrm{d}x_2 \tag{9}$$

are to be maximized subject to the constraint

$$\operatorname{div}\phi + q(p) = 0 \tag{10}$$

Observe that the region is "punctured" at $\underline{x} = \underline{0}$ and that a free boundary condition at $\underline{x} = \underline{0}$ permits the necessary flow to enter the region.

Consider now the Lagrangean

$$\int\int L \, \mathrm{d}x_1 \mathrm{d}x_2 = \int\int (p-c)\, q(p) - k|\phi| - \lambda\cdot(\operatorname{div}\phi + q)\mathrm{d}x_1 \mathrm{d}x_2 \tag{11}$$

The efficiency conditions are

$$\operatorname{grad}\lambda = k\frac{\phi}{|\phi|} \qquad \phi \neq 0 \tag{12}$$

and

$$q + (p - c - \lambda)q' = 0 \tag{13}$$

The first of these conditions is familiar. The second may be rewritten as the Amoroso–Robinson formula in the following way

$$\frac{p-c-\lambda}{p} = -\frac{q}{pq'} = \frac{1}{\epsilon}$$

where ϵ is the elasticity of demand. Solving

$$p = \frac{c+\lambda}{1-(1/\epsilon)} \tag{14}$$

If $\lambda(\underline{x})$ is tentatively identified as transportation cost to location $\underline{x}$ then $c + \lambda$ is marginal cost and (14) states that prices should equal marginal cost times the Lerner factor of the degree of monopoly $1/(1-(1/\epsilon))$. In particular, let q be a linear function times population density $\rho(\underline{x})$. After standardization, i.e. a proper choice of units (Beckmann 1976b)

$$q = \rho(\underline{x})(1-p) \tag{15}$$

Equation (13) then becomes

$$\rho(1-p)-\rho\cdot[p-c-\lambda] = 0$$

or

$$p = \frac{1+c+\lambda}{2} \tag{16}$$

Here $(1+c)/2$ is the optimal price at the mill. To this is added half of the freight cost λ, a result first reported by Singer (1937).

To show that $\lambda(\underline{x})$ is indeed transportation cost from $\underline{0}$ to $\underline{x}$, observe that λ in (12) is arbitrary and one may set $\lambda(\underline{0})=0$ without restriction. (Alternatively $\lambda(\underline{0})=0$ may be derived from the free boundary condition at $\underline{0}$.) Equation (12) shows that λ increases by $\underline{k}$ in the direction of actual shipment ϕ.

The dual problem may be obtained as before by rewriting the Lagrangean function (11)

$$\int\int L\,\mathrm{d}x_1\mathrm{d}x_2 = \int\int (p-c-\lambda)q - k|\phi| - \operatorname{div}(\lambda\phi) + \phi\cdot\operatorname{grad}\lambda\,\mathrm{d}x_1\mathrm{d}x_2 \tag{17}$$

using the identity

$$\operatorname{div}\lambda\phi = \lambda\operatorname{div}\phi + \phi\operatorname{grad}\lambda \tag{18}$$

By the Gauss integral theorem

$$\int\int \operatorname{div}(\lambda\phi)\mathrm{d}x_1\mathrm{d}x_2 = \int \lambda\,\phi_n\mathrm{d}x$$

On the outer boundary ϕ_n vanishes by assumption; on the boundary at $\underline{x} = \underline{0}$, λ vanishes. Therefore

$$\int\int \operatorname{div}(\lambda\phi)\mathrm{d}x_1\mathrm{d}x_2 = 0$$

Thus

$$\int\int L\,\mathrm{d}x_1\mathrm{d}x_2 = \int\int (p-c-\lambda)q + \phi\cdot\left(\operatorname{grad}\lambda - k\frac{\phi}{|\phi|}\right)\mathrm{d}x_1\mathrm{d}x_2$$

$$\leqslant \int\int (p-c-\lambda)q\,\mathrm{d}x_1\mathrm{d}x_2$$

for all λ such that

$$|\operatorname{grad}\lambda| \leqslant k$$

This proves the following duality principle

$$\max_{\phi,p}\int\int (p-c)q - k|\phi|\mathrm{d}x_1\mathrm{d}x_2 \quad = \quad \max_{\lambda,p}\int\int (p-c-\lambda)q\,\mathrm{d}x_1\mathrm{d}x_2 \tag{19}$$

$$\text{subject to } \operatorname{div}\phi + q = 0 \qquad\qquad \text{subject to } |\operatorname{grad}\lambda| \leqslant k$$

The case of uniform pricing is now obtained by making p constant rather than an arbitrary function of $\underline{x}$. It may therefore be considered as a restricted case of discriminatory pricing.

Now suppose that q is linear. Then the maximand in the dual has the form

$$\rho(p-c-\lambda)(1-p) \tag{20}$$

Now define

$$m = 1+c-p \tag{21}$$

and substitute for p in (20)

$$\rho(1-m-\lambda)(m-c) \tag{22}$$

Now m can be considered a mill price, and the first term in parentheses is quantity demanded, if customers pay $m+\lambda$; the second term in parentheses is then unit profit. Comparison of (20) and (22) shows that finding the optimal mill price is mathematically equivalent to finding the optimal uniform price, and that the optimal mill and uniform prices are related by (21). This result is well known for the case of monopoly in a uniform transport cost case (Beckmann and Ingene 1976).

For general demand functions the problem of determining the optimal mill price may be formulated in a similar way using the dual (no formulation is known in terms of the primal)

$$\max_{p,\lambda}\int\int(p-c)q(p+\lambda)\,dx_1\,dx_2 \tag{12}$$

subject to

$$|\text{grad}\,\lambda| = k$$

$$\lambda(0) = 0$$

The Lagrangean is

$$L = (p-c)q(p+\lambda)+\psi\cdot(\text{grad}\,\lambda-k\epsilon)$$

where ϵ is a suitable unit vector and the vector ψ a Lagrangean multiplier. We now show that ψ may be interpreted as the flow vector. First we write

$$\int\int L\,dx_1\,dx_2 = \int\int (p-c)q(p+\lambda)-k\psi\cdot\epsilon+\text{div}(\lambda\psi)-\lambda\,\text{div}\,\psi\,dx_1\,dx_2$$

Now

$$\int\int \text{div}\,\lambda\psi\,dx_1\,dx_2 = \int\lambda\psi_n\,ds$$

vanishes on the outer boundary by the boundary condition and at $\underline{0}$ since $\lambda(\underline{0})=0$. Consider the maximum with respect to λ of

$$\int\int (p-c)\,q(p+\lambda) - \lambda \operatorname{div} \psi \, dx_1 \, dx_2$$

Differentiating the integrand with respect to λ and setting this equal to zero

$$(p-c)q' - \operatorname{div} \psi = 0 \tag{23}$$

If the maximand (22) is maximized with respect to p one has

$$\bar{q} + (p-c)q' = 0 \tag{24}$$

and comparing (23) and (24) yields

$$\operatorname{div} \psi + q = 0$$

confirming the interpretation of ψ as the flow vector.

In the case of linear demand function the optimal uniform and mill prices may be given explicitly. Differentiating the Lagrangean integral with respect to the uniform price p and setting it equal to zero, we have

$$0 = \int\int \frac{d}{dp}((p-c-\lambda)\rho(1-p))\, dx_1 \, dx_2$$

or

$$0 = \int\int \rho(1-2p+c+\lambda)\, dx_1 \, dx_2$$

from which

$$p = \frac{1}{2} \; \frac{\int\int (1+c+\lambda)\rho \, dx_1 \, dx_1}{\int\int \rho \, dx_1 \, dx_2} = \frac{1}{2} \left(1 + c + \frac{\int\int \lambda \, \rho \, dx_1 \, dx_2}{\int\int \rho \, dx_1 \, dx_2}\right)$$

or

$$p = \frac{1+c+\bar{\lambda}}{2} \tag{25}$$

Here $\bar{\lambda}$ represents the average transportation cost from the mill to the customers. Using the identity (21) we immediately obtain the optimal mill price

$$m = 1 + c - p = \frac{1+c-\bar{\lambda}}{2} \tag{26}$$

These results, too, are well known for the special case of a uniform transportation plane. Here they have been extended to general transportation cost metrics.

3.1.4 Monopoly with Multiple Facilities

A monopolist may control multiple outlets operating (in general) with different marginal costs. In such a situation uniform pricing and mill pricing are much less natural;

in fact, each plant should have its own mill price. In contrast, under "basing point pricing" a mill price is defined for one outlet only, and prices are charged as if all supplies came from that one point. Uniform price too could depend on the supply point but this would invite retrading across boundaries where prices change discontinuously.

Rather than discuss these complex and often illegal possibilities, we will consider here only a perfectly discriminating monopolist, particularly since this sets a standard that other monopolistic practices can only approximate.

We begin by considering a monopolistic trader who has available quantities $z(\underline{x})$ at locations $\underline{x}$, and these quantities are assumed to have zero opportunity cost. The monopolist faces demand $q(p, \underline{x})$ at location $\underline{x}$ and seeks prices $p(\underline{x})$ that yield maximum profits. In view of the commodity shipments ϕ required for these sales, his profits are

$$G = \int\int pq(p) - k|\phi| \mathrm{d}x_1 \mathrm{d}x_2$$

The constraint is now

$$\operatorname{div} \phi \leqslant z - q \tag{27}$$

with the left-hand side representing net exports and the right-hand side the quantities available for net export from a given location. In order to maximize G subject to constraint (27) we introduce the Lagrangean

$$\int\int L \,\mathrm{d}x_1 \mathrm{d}x_2 = \int\int pq(p) - k|\phi| - \lambda(\operatorname{div} \phi + q - z)\mathrm{d}x_1 \mathrm{d}x_2 \tag{28}$$

Maximization with respect to ϕ and p yields

$$k\frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \phi \neq 0 \tag{29}$$

$$(p - \lambda)q' + q = 0 \tag{30}$$

Equation (30) is the well-known condition for a profit maximum if λ is interpreted as the opportunity cost of supplying the commodity at location $\underline{x}$.

It is possible that the monopolist will choose not to sell the entire supply

$$\int\int z \,\mathrm{d}x_1 \mathrm{d}x_2$$

In that case the $<$ sign in (27) will apply somewhere so that λ will be zero at locations of excess availability.

3.1.5 A Case of Duopoly Pricing

The theory of pricing under spatial oligopoly rightly occupies an important place in spatial economics (Hotelling 1929). In this section we treat one illustrative example in order to demonstrate how a continuous treatment of space may be incorporated into this theory.

Suppose that a duopolist offers a product slightly different from ours so that markets will overlap. Assume an exponential distance effect, attractions a_1 and a_2, product prices p_1 and p_2, and distances (in transportation cost units) λ and r, respectively. Our power of market penetration at any point $\underline{x}$ is then given by

$$a_1 \, e^{-b(p+\lambda)} \tag{31}$$

and the resulting market share is

$$\frac{a_1 \, e^{-b(p_1+\lambda)}}{a_1 \, e^{-b(p_1+\lambda)} + a_2 \, e^{-b(p_2+r)}} = \frac{1}{1 + \dfrac{a_2}{a_1} \, e^{b(p_1-p_2+\lambda-r)}} \tag{32}$$

Writing

$$\frac{a_2}{a_1} \, e^{-bp_2} = a$$

and

$$p_1 = p$$

this simplifies to

$$\frac{1}{1 + a e^{b(p+\lambda-r)}} \tag{33}$$

If the density of demand is ρ, our profit G is given by

$$G = \int\int \frac{(p-c-\lambda)\rho}{1 + a e^{b(p+\lambda-r)}} \, dx_1 \, dx_2 \tag{34}$$

If the duopolist behaves adaptively he chooses a profit-maximizing p regardless of the other duopolist's price p_2, which is implicit in a. Essentially, the problem has now been reduced to the form of a monopoly problem, with a given, nonlinear demand function $q(p+\lambda)$, as studied in Section 3.1.4

$$q(p+\lambda, \underline{x}) = \frac{1}{1 + e^{b(p+\lambda-r(\underline{x}))}}$$

When choosing a mill price, p must be treated as a constant; when selecting a discriminatory price policy one has instead $p = p(\underline{x})$. We illustrate this by considering a discriminating price policy for distant customers. For customers distant from us but close to the other firm, $p + \lambda - r$ is large, so that the integrand is approximately

$$(p-c-\lambda)\frac{\rho}{a} \cdot e^{-b(p+\lambda-r)}$$

Maximizing with respect to p yields

$$0 = e^{-b(p+\lambda-r)} - b(p-c-\lambda)e^{-b(p+\lambda-r)}$$

from which

$$p = \frac{1}{b} + c + \lambda \tag{35}$$

This is a policy of mark-up pricing in which a fixed profit margin $1/b$ is added to the cost $c + \lambda$ of supplying a customer. Here discriminatory pricing is identical with mill pricing. Notice that the terms p_2 and r referring to the effect of competition no longer appear in the price-determination equation, but that $1/b$, which reflects the exponential distance effect, does. In principle this price policy is still valid when more than two firms are present, provided the market is always divided with one (for example, the nearest) of the several competitors.

3.2 LAND USE: PARTIAL EQUILIBRIUM

Having derived demand from utility functions (cf. Section 3.1.2), the next step is to obtain supply from production decisions. We start with the simplest type of production function. Throughout it is assumed that producers maximize profit.

3.2.1 An Introductory Model

It is a fact of economic theory that supply is accessible to much deeper analysis than demand. So far we have considered the short run, where supply is given. Having discussed demand as the result of utility maximization (in a one-commodity model requiring no budget constraint), we now turn to the supply side in the intermediate run: all facilities are in place but they are subject to capacity constraints. Up to the limit of capacity, production is possible at constant unit cost. This is consistent with (but does not require) constant input coefficients. For linguistic simplicity, the production facilities will be identified with the land they occupy. Therefore we consider in this section that land at location $\underline{x}$ can produce the product in question up to an amount $a(\underline{x})$ at constant unit costs $c(\underline{x})$. The supply curve is thus of the type shown in Figure 3.1.

Initially we assume demand to be given independent of price $q(\underline{x})$. The competitive market will allocate production so as to

$$\min_{\phi, z} \int\int k\,|\phi| + cz \;\mathrm{d}x_1\,\mathrm{d}x_2 \tag{36}$$

subject to

$$\operatorname{div} \phi + q - z = 0 \tag{37}$$

$$0 \leqslant z \leqslant a \tag{38}$$

where all variables depend on location $\underline{x}$.

Consider now the Lagrangean

$$\int\int L \;\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int (-k\,|\phi| - cz - \lambda(\operatorname{div} \phi + q - z) + \mu(a - z))\;\mathrm{d}x_1\,\mathrm{d}x_2$$

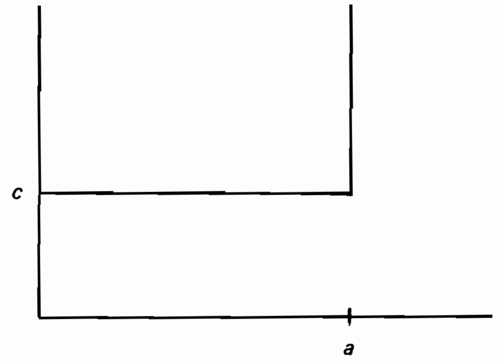

Figure 3.1. A simple supply function.

Optimization yields the efficiency conditions

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \phi \neq 0 \tag{39}$$

$$\left.\begin{array}{lll} z = a & \text{for} & \lambda > c \\ 0 \leqslant z \leqslant a & \text{for} & \lambda = c \\ z = 0 & \text{for} & \lambda < c \end{array}\right\} \tag{40}$$

Production takes place only where the value for the product λ, its competitive market price, is at least equal to its marginal cost c. Production occurs at the maximum rate a when the product's value exceeds its marginal cost.

Now, at locations where the commodity is consumed but not produced, its value λ must be larger than at some location whence it is obtained. It follows that, at all locations

$$\lambda \geqslant \min_{\underline{x}} c(\underline{x})$$

The minimum level of marginal cost is a lower bound for the commodity's market price. This bound is particularly interesting when $c(\underline{x}) \equiv c$ is constant throughout. Then c is in fact a lower bound on commodity prices everywhere. For

$$q \leqslant a,\ \lambda \geqslant c, \text{ and } \lambda = c \text{ only where } q \leqslant a$$

When

$$\int\!\!\int q \,\mathrm{d}x_1 \mathrm{d}x_2 = \int\!\!\int a \,\mathrm{d}x_1 \mathrm{d}x_2$$

the aggregate production capacity is just sufficient to satisfy aggregate demand. We have then, in fact, once again the continuous model of transportation.

When demand is price-dependent, the market will maximize a welfare function, in which $u(q)$ is the consumers' surplus

$$\max_{\phi, z, q} \int\!\!\int u(q) - cz - k|\phi| \,\mathrm{d}x_1 \mathrm{d}x_2$$

subject to (37), (38). An additional efficiency condition determines consumption q

$$\left.\begin{array}{lll} q = 0 & \text{for} & u'(q) < \lambda \\ q \geqslant 0 & \text{for} & u'(q) = \lambda \end{array}\right\} \tag{41}$$

Thus no consumption takes place where marginal utility, even at the zero level of consumption, falls short of the product's market price λ; otherwise marginal utility equals λ.

The dual of the problem with fixed demand is obtained as follows

$$\int\int -k|\phi| - cz + \lambda[z - q - \operatorname{div}\phi] + \mu[a - z]\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\int -\phi\cdot\operatorname{grad}\lambda + [\lambda - \mu - c]z - \lambda\operatorname{div}\phi + \mu a - \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$\leqslant -\int \lambda\phi_n\,\mathrm{d}s + \int\int \mu a - \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad \text{for all} \quad \lambda - \mu \leqslant c$$

$$= \int\int [\mu a - \lambda q]\,\mathrm{d}x_1\,\mathrm{d}x_2$$

since ϕ_n vanishes, by the boundary condition. Thus

$$\begin{array}{ccc} \min\limits_{\phi} \displaystyle\int\int k|\phi| - cz\,\mathrm{d}x_1\,\mathrm{d}x_2 & = & \max\limits_{\mu,\lambda\geqslant 0} \displaystyle\int\int [\mu a - \lambda q]\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ \text{subject to (37) and (38)} & & \text{subject to } \lambda - \mu \leqslant c \\ & & |\operatorname{grad}\lambda| \leqslant k \end{array} \tag{42}$$

The dual is in terms of efficiency prices λ and μ, while the primal is in terms of a vector field ϕ. For price-dependent demand, the dual is instead

$$\max_{\lambda,\mu\geqslant 0} \int\int \int_{\lambda}^{\infty} q(p)\,\mathrm{d}p + \mu a\,\mathrm{d}x_1\,\mathrm{d}x_2$$

subject to

$$\lambda - \mu \leqslant c$$

$$|\operatorname{grad}\lambda| \leqslant k$$

This may be shown along the same lines as those of Section 3.1.2.

3.2.2 Variable Proportions

In this section we consider production of a single commodity with mobile labor and fixed land according to a constant-returns production function

$$z = f(l, m) \tag{43}$$

where l is labor and m land. The cost of production is

$$wl + gm$$

where w is the wage rate and g the land rent, and land input is subject to the constraint

$$m \leqslant a \tag{44}$$

Once again we consider demand to be fixed. A competitive market now allocates resources so as to

$$\min_{l, m, \phi} \int\int wl + gm + k|\phi| \mathrm{d}x_1 \mathrm{d}x_2$$

subject to (44) and

$$\operatorname{div} \phi = f(l, m) - q \tag{37a}$$

Consider the Lagrangean

$$L = -wl - gm - k|\phi| - \lambda(\operatorname{div} \phi + q - f(l, m)) + \mu(a - m)$$

Optimization with respect to ϕ, l, and m yields the efficiency conditions

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \phi \neq 0 \tag{45}$$

$$\left.\begin{aligned} l &= 0 \quad \text{for} \quad \lambda \frac{\partial f}{\partial l} < w \\ l &\geqslant 0 \quad \text{for} \quad \lambda \frac{\partial f}{\partial l} = w \end{aligned}\right\} \tag{46}$$

$$\left.\begin{aligned} m &= 0 \quad \text{for} \quad \lambda \frac{\partial f}{\partial m} < g + \mu \\ m &\geqslant 0 \quad \text{for} \quad \lambda \frac{\partial f}{\partial m} = g + \mu \end{aligned}\right\} \tag{47}$$

Now, if in conditions (46), (47) the lower alternatives apply for some l, $m > 0$, then they apply for all positive multiples of l and m so that production can be expanded to the limit $m = a$. Conditions (46), (47) state the familiar marginal productivity principle: each factor is employed to the point where the value of its marginal product equals its wage.

For illustrative purposes consider a Cobb–Douglas production function

$$z = b(\underset{\sim}{x}) l^{\alpha} m^{\beta} \qquad \alpha + \beta = 1 \tag{48}$$

Notice that productivity b depends on location, but the output elasticities α and β are technologically determined and independent of location. Conditions (46), (47) now take the form

$$\left.\begin{array}{lll} l = 0 & \text{for} & \alpha\lambda\dfrac{z}{l} < w \\ l \geqslant 0 & \text{for} & \alpha\lambda\dfrac{z}{l} = w \end{array}\right\} \tag{46a}$$

$$\left.\begin{array}{lll} m = 0 & \text{for} & \beta\lambda\dfrac{z}{m} < g + \mu \\ m \geqslant 0 & \text{for} & \beta\lambda\dfrac{z}{m} = g + \mu \end{array}\right\} \tag{47a}$$

Wherever production takes place the factor proportions are determined by

$$\frac{m}{l} = \frac{w}{g+\mu}\,\frac{\beta}{\alpha} \qquad g + \mu \neq 0 \tag{49}$$

Substituting l from (49) into the production function (48) yields

$$z = b\left(\frac{g}{w}\,\frac{\alpha}{\beta}\right)^{\alpha} m$$

The production cost associated with a unit input of land is

$$g + \mu + w\cdot\left(\frac{g+\mu}{w}\,\frac{\alpha}{\beta}\right) = (g+\mu)\left(1 + \frac{\alpha}{\beta}\right) = \frac{g+\mu}{\beta}$$

Thus

$$\left.\begin{array}{lll} m = a & \text{for} & \dfrac{g+\mu}{\beta} \leqslant \lambda \\ m = 0 & \text{for} & \dfrac{g+\mu}{\beta} > \lambda \end{array}\right\} \tag{50}$$

In particular, when the factors have no alternative use

$$g = 0$$

and

$$\mu = 0 \qquad \text{when } m < a$$

It follows from (50) that

$$m = a \qquad \text{whenever } \lambda > 0 \tag{51}$$

Thus production takes place at full capacity wherever the product price is positive. In that case labor input is determined by

$$w = \lambda \frac{\partial f}{\partial l}(l, a)$$

$$= \frac{\lambda \alpha z}{l}$$

$$= \lambda b \alpha a^{\beta} l^{\alpha - 1}$$

$$= \lambda b \alpha \left(\frac{a}{l}\right)^{\beta}$$

from which

$$l = a \cdot \left(\frac{\lambda \alpha b}{w}\right)^{1/\beta} \tag{52}$$

The case of flexible demand is similar, with the addition of the consumption condition (41) from Section 3.2.1.

It is always possible, even with price-dependent demand, that the value of the commodity falls to zero in remote locations, because it cannot cover the cost of transportation to the consumption locations. In such circumstances even a zero land rent cannot induce its utilization.

Substituting (51) and (52) into the production function yields output

$$z = a\, b^{1/\beta} \left(\frac{\alpha \lambda}{w}\right)^{\alpha/\beta} \tag{53}$$

Finally we obtain profits or rent $g + \mu$ (setting $\mu = 0$)

$$g = \lambda z - wl$$

$$g = \beta a b^{1/\beta} \left(\frac{\alpha \lambda}{w}\right)^{\alpha/\beta} \tag{54}$$

Note that all the variables are proportional to land availability a.

3.2.3 An Illustration: Supply Areas

Let consumption of a particular commodity be concentrated in one small area, assumed to be a circle of negligible radius (or even a point). In the surrounding area, land has no alternative use other than to produce the commodity in question. Labor is available wherever needed at a constant wage rate w and transportation cost k is constant. The production function is the Cobb–Douglas function of the previous section. Under competitive conditions all variables are functions of distance r from the consumption center. In particular

$$\lambda = \lambda(r) = \lambda(0) - kr \tag{55}$$

Land is used productively wherever λ is positive

$$m = a \qquad \text{for} \quad \lambda(r) - kr > 0$$

$$m = 0 \qquad \text{for} \quad \lambda(r) - kr \leqslant 0$$

The limit of production R is determined by

$$\lambda(0) - kR = 0$$

or

$$R = \frac{\lambda(0)}{k} \tag{56}$$

Labor employed per unit of land is then, using (52)

$$l = a \cdot \left(\frac{\alpha b}{w}\right)^{1/\beta} \cdot [\lambda(0) - kr]^{1/\beta} \qquad r \leqslant R \tag{52a}$$

Labor intensity is thus a convex function of r, decreasing with distance from the consumption center to zero for $r \geqslant R$. Output is, using (53)

$$z = ab^{1/\beta} \left(\frac{\alpha}{w}\right)^{\alpha/\beta} [\lambda(0) - kr]^{\alpha/\beta} \tag{53a}$$

Output is proportional to land capacity a, and it also decreases with distance to zero at $r = R$. Output is a convex, linear, or concave function of r depending on whether α is greater than, equal to, or less than β. The standard case is $\alpha > \beta$.

Finally consider profits or rent per unit area

$$g = \beta\, ab^{1/\beta} \left(\frac{\alpha}{w}\right)^{\alpha/\beta} [\lambda(0) - kr]^{\alpha/\beta} \tag{54a}$$

Rent also decreases with distance and reaches zero at $r = R$. It is convex in the standard case when $\alpha > \beta$.

Qualitatively, these results remain true when some of the output is consumed locally and some of the land is used for housing labor locally. In the simplest version only the productivity coefficient b and the land availability a are then reduced, but the analysis remains the same.

3.2.4 Land and Labor Immobile

Let both labor and land be available in fixed quantities

$$m \leqslant a \tag{44}$$

$$l \leqslant d \tag{44a}$$

and have opportunity costs g and w, respectively. Assume once more a linear homogeneous production function

$$z = f(l, m) \tag{38}$$

The competitive market solves an allocation problem

$$\min \int\int (wl + gm + k|\phi|) \mathrm{d}x_1 \mathrm{d}x_2$$

subject to constraints (44), (44a), and

$$\operatorname{div} \phi = f(l, m) - q \tag{37a}$$

Consider the Lagrangean

$$L = -wl - gm - k|\phi| - \lambda(\operatorname{div} \phi - f(l, m) + q) + \mu(a - m) + \nu(d - l)$$

Optimization with respect to l, m, and ϕ yields

$$k \frac{\phi}{|\phi|} = \operatorname{grad} \lambda \qquad \phi \neq 0 \tag{45}$$

$$\left.\begin{aligned} l &= 0 \qquad \text{for} \quad \lambda \frac{\partial f}{\partial l} < w \\ l &\geqslant 0 \qquad \text{for} \quad \lambda \frac{\partial f}{\partial l} = w + \nu \end{aligned}\right\} \tag{46a}$$

$$\left.\begin{aligned} m &= 0 \qquad \text{for} \quad \lambda \frac{\partial f}{\partial m} < g \\ m &\geqslant 0 \qquad \text{for} \quad \lambda \frac{\partial f}{\partial m} = g + \mu \end{aligned}\right\} \tag{47a}$$

We now show that $\mu > 0$, $\nu > 0$ cannot both be true but that, at most, one factor can earn a scarcity rent μ or ν, respectively. In fact, performing the calculations of the previous section (3.2.2) with $w + \nu$ in lieu of w shows that, as before, either $z = 0$ or $m = a$. Moreover, factor proportions depend on $(w + \nu)/(g + \mu)$. But any positive value of that ratio can be achieved by choosing either ν positive or μ positive or both zero. Thus, even when both factors are fully utilized, the scarcity rent need go to only one factor.

It can be shown that factor prices $w + \nu$ and $g + \mu$ are continuous functions of location $\underline{x}$. In particular, if wages w and rents g are continuous, so are the scarcity rents, ν and μ, respectively.

3.2.5 Remarks on Diminishing Returns

Examination of the case of a production function with diminishing returns offers nothing new, except that neither land nor labor need now be fully utilized in locations where production of the commodity takes place. Diminishing returns are due presumably to the presence of other factors such as capital. Production with land, labor, and capital is studied in detail in Chapters 4 and 5.

In the following section we turn from single-commodity models to those dealing with multiple products.

3.2.6 Alternative Land Uses: The Specialization Theorem

In this section we discuss briefly a topic to be more fully developed in Chapter 4, namely, how land should be allocated to competing uses. Our discussion is in terms of a fixed-coefficient technology, but the result applies to any set of linear homogeneous production functions.

Suppose a unit of product i can be produced at location $\underline{x}$ with unit cost c_i of mobile inputs (including labor) and a fixed input of land, m_i. Total land availability is again $a = a(\underline{x})$. Let the quantities desired for consumption be $q_i(\underline{x})$. The competitive market will then minimize the cost of achieving this consumption program. Let z_i denote quantities produced and ϕ_i quantities shipped

$$\min_{z_i, \phi_i} \; -\iint \sum_i c_i z_i + \sum_i k_i |\phi_i|$$

subject to

$$\operatorname{div} \phi_i + q_i - z_i = 0 \tag{37b}$$

$$\sum_i m_i z_i \leqslant a \tag{44b}$$

The first constraint states the commodity balance equation for each commodity and the second the limit on land availability.

We introduce the Lagrangean

$$L = -\sum_i c_i z_i - \sum_i k_i |\phi_i| - \sum_i \lambda_i (\operatorname{div} \phi_i + q_i - z_i) + \mu \left(a - \sum_i m_i z_i \right)$$

The efficiency conditions are

$$k_i \frac{\phi_i}{|\phi_i|} = \operatorname{grad} \lambda_i \tag{45b}$$

$$\left.\begin{aligned} z_i &= 0 \qquad \text{for} \quad \lambda_i < c_i + \mu m_i \\ z_i &\geqslant 0 \qquad \text{for} \quad \lambda_i = c_i + \mu m_i \end{aligned}\right\} \tag{46b}$$

Here μ obviously represents the rent of land that serves to ration its use among the competing products.

Note first that not all land need be utilized, since production cost c_i for every product could exceed product value λ_i in some (remote) location. If a positive program of consumption q is to be carried out, however, land in some locations will have to be used productively. Wherever this is true, a rewriting of (46b) shows that

$$\mu \geqslant \frac{\lambda_i - c_i}{m_i}$$

and

$$\mu = \frac{\lambda_i - c_i}{m_i} \qquad \text{for some } i$$

It appears that μ represents the largest value added per unit of land, a statement expressed more succinctly by

$$\mu = \max_i \left(\frac{\lambda_i - c_i}{m_i} \right) \tag{57}$$

Suppose that two products can be effectively produced in some location $\underline{x}$ because the lower alternative holds in (46b) for more than one i. In order for this to be true in an entire neighborhood, the two products must satisfy

$$\frac{\lambda_i - c_i}{m_i} \equiv \frac{\lambda_j - c_j}{m_j}$$

Suppose that costs c and land inputs m remain in the same proportion. Then the two product prices must also stay in the same proportion. The two products must move in the same direction. Products whose prices and movements are similar to this extent should be considered identical.

We state this result as a

Land Use Specialization Theorem: *Except on sets of measure zero land is used for, at most, one product.*

This theorem states, under more general market conditions, what is already known in the von Thünen case: that zones of specialized land use emerge under competitive conditions.

3.2.7 A Short-Run Model of Urban Structure

As a preliminary to a more comprehensive model of the location of production and residential activities in an urban area, consider the following simple model of competition for space between two types of industry and their labor forces. This is a von Thünen model based on a fixed-coefficient technology and it addresses the same problem of urban structure as the Mills model (Mills 1972). However, the space demand of production activities is expressed not in terms of a certain height of buildings but in terms of floor-space and labor requirements per unit of output and in terms of weight (or transportation cost) per unit of output. Moreover, in the short run the amount of available floor space $\underline{a}$ is assumed given. The total amounts to be produced of each product Z_1, Z_2

are also assumed given; so are the total labor requirements L_1, L_2. Production costs are independent of location. The objective is to locate these activities so as to minimize total transportation cost. This classic principle of location theory has recently acquired a new significance through our heightened consciousness of the need for energy saving in transportation.

Let ϕ_i denote the flow of product $i = 1,2$ and ϕ_0 denote the flow of labor, all measured in the direction of the center. The objective function is then

$$\int\int \sum_{i=0}^{2} k_i |\phi_i| \mathrm{d}x_1 \mathrm{d}x_2$$

where the domain of integration is a circle of radius R.

Let z_i be the density of activity i per unit of floor space. The following constraints apply to the interaction of activities and flows. For production:

$$\operatorname{div} \phi_i = z_i \qquad i = 1,2 \tag{37c}$$

The divergence of the flow equals production. Let b_0 units of labor require one unit of housing, and let production of i require b_i units of labor per unit of floor space. z_0 is the density of housing. Then the commuter flow ϕ_0 is governed by the divergence law

$$\operatorname{div} \phi_0 = b_0 z_0 - b_1 z_1 - b_2 z_2 \tag{58}$$

The total output requirements are

$$Z_i = \int\int z_i \, \mathrm{d}x_1 \mathrm{d}x_2 \qquad i = 1,2 \tag{59}$$

and the total amount of labor needed is

$$L = \int\int \sum_{i=1}^{2} b_i z_i = \int\int z_0 \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{60}$$

Conditions (59) and (60) will be restated as boundary conditions.

The amount of floor space available at $\underset{\sim}{x}$ may be allocated to either housing or production

$$z_0 + z_1 + z_2 \leqslant a \tag{43c}$$

Notice that we assume the space requirements of production activities to be unity. In other words, we measure production levels in terms of their space requirements. The problem is to allocate space in such a way that total transportation cost is minimized subject to constraints (37c), (43c), (58)–(60).

The boundary conditions are as follows: no flow of labor or product through the outer boundary, no flow of labor through the center, but total flows Z_i of product i through the center. We construct a Lagrangean

$$\int\int \left[\sum_{i=0}^{2} k_i |\phi_i| + \sum_{i=1}^{2} \lambda_i (z_i - \operatorname{div} \phi_i) + \lambda_0 (b_0 z_0 - \sum_{i=1}^{2} b_i z_i - \operatorname{div} \phi_0) \right.$$
$$\left. + \mu (a - z_0 - z_1 - z_2) \right] \mathrm{d}x_1 \mathrm{d}x_2 + \sum_{i=1}^{2} \beta_i \left(\int\int z_i \mathrm{d}x_2 \mathrm{d}x_2 - Z_i \right)$$

The efficiency conditions are as follows

$$k_i \frac{\phi_i}{|\phi_i|} = \text{grad}\,\lambda_i \qquad i = 0,1,2 \qquad \phi_i \neq 0 \tag{45b}$$

$$\left.\begin{array}{lll} z_i = 0 & \text{wherever} & \lambda_i < b_i\lambda_0 + \mu \\ z_i \geqslant 0 & \text{wherever} & \lambda_i = b_i\lambda_0 + \mu \end{array}\right\} \quad i = 1,2 \tag{46b}$$

$$\left.\begin{array}{lll} z_0 = 0 & \text{wherever} & b_0\lambda_0 < \mu \\ z_0 \geqslant 0 & \text{wherever} & b_0\lambda_0 = \mu \end{array}\right\} \tag{46c}$$

Here λ_i represents the values of labor and products. Labor commutes in the direction of increasing value (equation (45b)). It is housed where its value equals that of the floor space occupied (equation (46c)). Production takes place where the product value equals the value of land and labor inputs (equation (46b)).

Floor space always goes to the highest bidder, for (46b), (46c) can be rewritten

$$\mu = \max(\lambda_1 - b_1\lambda_0, \lambda_2 - b_2\lambda_0, b_0\lambda_0) \tag{57b}$$

$$= \max(\mu_1, \mu_2, \mu_0) \qquad \text{(for example)} \tag{57c}$$

Since the product necessarily moves to the center, the product values λ_i at distance r from the center are

$$\lambda_i(r) = \lambda_i(0) - k_i r \qquad i = 1,2 \tag{55c}$$

The locations of the two production activities and the housing activity depends on their rent bid functions for floor space.

Consider first the condition under which one industry, say $i = 1$, locates right next to the center. Its rent bid function μ_1 must then fall off more rapidly than the rent bid function μ_0 of labor for housing. Now

$$\text{grad}\,\mu_0 = -k_0 b_0$$

$$\text{grad}\,\mu_1 = \text{grad}(\lambda_1 - b_1\lambda_0)$$

$$= -k_1 + b_1 k_0$$

The condition for "heavy industry" to outbid labor for central locations is

$$\text{grad}\,\mu_1 < \text{grad}\,\mu_0$$

or

$$-k_1 + b_1 k_0 < -b_0 k_0$$

or

$$k_1/k_0 > b_1 + b_0 \tag{61}$$

The transportation cost of product 1 must be sufficiently large relative to the commuting cost of labor. When (61) is satisfied, the inner zone immediately adjacent to the central business district (represented here by a point) is occupied by heavy industry and the next zone by labor commuting to work in the heavy industry zone.

Next we consider the possibility of mixed land use involving housing and industry. An industry 2 using resident labor has a rent bid function

$$\mu_2 = \lambda_2 - b_2\lambda_0 \tag{61a}$$

in view of (57b), (57c). Its bid must be the same as that for housing, μ_0

$$\mu_0 = b_0\lambda_0$$

from which

$$\lambda_0 = \frac{\lambda_2}{b_0 + b_2}$$

and using (61a)

$$\mu_2 = \frac{b_0\lambda_2}{b_0 + b_2} \tag{62}$$

If this zone is to follow that of commuting labor one must have

$$\operatorname{grad}\mu_0 < \operatorname{grad}\mu_2$$

$$-k_0 b_0 < -\frac{b_0}{b_0 + b_2}\cdot k_2$$

or

$$k_2/k_0 < b_0 + b_2 \tag{63}$$

Thus, zones for "heavy industry" using commuting labor, and "light industry" using resident labor and located beyond the housing zone for commuters are distinguished by the criteria (61a) and (63). This model can be augmented by introducing industries whose product is not marketed at the center but is exported through the outer boundary. Once again, three zones are possible. Together with the CBD, here represented by a point, they constitute the seven zones of a city. Incidentally, one of the "heavy industries" may be the housing of rentiers and elderly persons who spend their time in the CBD but experience heavy transportation costs.

The mixed use of land for housing and production might seem to violate the specialization theorem of Section 3.2.6. However, labor is operating here as a mobile intermediate good – and the specialization theorem did not allow for such intermediate commodities. Housing and light manufacturing may, in fact, be considered as a joint activity, characterized by its own constant coefficient of production. By contrast, if two activities could take place in arbitrary mixtures at the same location, such fixed joint coefficients would not exist. In passing, it is interesting to note that exactly this type of joint land use for housing and manufacturing was observed in Chicago after the fire (Fales and Moses 1972).

3.3 A GENERAL EQUILIBRIUM MODEL

3.3.1 One Mobile Product

An economy in which there is only one transportable good cannot be organized as a market economy. For how can the exporters of the commodity be compensated and how can the importers pay for their demand? This is possible, however, when the importers are absentee landlords and the exporters are serfs who pay their rent (personal or land rent) by handing over specified quantities of the product to their landlords. Transportation is then arranged by other (possibly foreign) agents, who are paid in terms of the good by either serfs or landlords. Thus the surplus produced by serfs takes care of both the consumption of landlords and the transportation cost.

Another possibility is the following. Over a person's lifetime, the single product may be net produced by the person while in one location and net consumed after retirement to another location. In this case imports have been prepaid so that a combined intertemporal interregional equilibrium is possible.

Let us return to the general model. Let transportation cost be in units of the single product and equal to h. The equation for local balance of the commodity is then

$$\operatorname{div}\phi = z - q - h|\phi| \tag{64}$$

Production of one unit of the commodity requires unit labor input so that no distinction need be made between labor and product in production. Let labor availability be limited

$$0 \leqslant z \leqslant a(\underline{x}) \tag{65}$$

The object is to achieve the consumption program q with minimum labor input

$$\min \int\int z \,\mathrm{d}x_1 \mathrm{d}x_2$$

subject to (64) and (65). Introduce the Lagrangean

$$L = -z - \lambda(\operatorname{div}\phi + q - z - h|\phi|) + \mu(a - z)$$

The competitive market achieves an efficient allocation of resources characterized by

$$\operatorname{grad}\lambda = h\lambda \frac{\phi}{|\phi|} \tag{66}$$

$$\left.\begin{array}{lll} z = a & \text{for} & \lambda > 1 \\ 0 \leqslant z \leqslant a & \text{for} & \lambda = 1 \\ z = 0 & \text{for} & \lambda < 1 \end{array}\right\} \tag{67}$$

While everything thus operates on a labor standard of value, the product value may lie above or below the unit value of labor. However, in locations where production takes place the commodity value is at least equal to one. The same applies *a fortiori* to consumption locations. In fact, where the commodity must be imported its value must be greater than one. Note that the transportation cost is itself proportional to the commodity value. Thus, equation (66) could be rewritten

$$\text{grad} \log \lambda = h \frac{\phi}{|\phi|} \tag{66a}$$

To see the implications of this consider a von Thünen case where production is dispersed and consumption is concentrated in a single point (or small circle). In terms of distance r from the center, (66a) states that

$$\frac{\mathrm{d} \log \lambda}{\mathrm{d}r} = -h$$

$$\lambda(r) = \lambda(0)e^{-hr} \tag{68}$$

The commodity thus has a positive value at every distance. The present model minimizing total labor input imposes, however, a cut-off distance $r = R$ where $\lambda(R) = 1$ as stated in (67). Suppose, however, that labor is available everywhere at wages fixed in commodity units $w\lambda$, and that land input is fixed at unity. Let the production function be Cobb–Douglas

$$z = bl^{\alpha} \tag{69}$$

Profit per unit of land using labor l on one unit of land equals

$$\lambda bl^{\alpha} - w\lambda l$$

and its maximum with respect to l is the land rent

$$g = \max_{l} \lambda bl^{\alpha} - w\lambda l$$

$$g = \lambda \cdot (1 - \alpha) \left(\frac{\alpha}{w}\right)^{\alpha/(1-\alpha)} b^{1/(1-\alpha)} \tag{70}$$

and this is achieved by a constant labor intensity

$$l = \left(\frac{\alpha b}{w}\right)^{1/(1-\alpha)} \tag{71}$$

resulting in a constant output per unit of land

$$z = b^{1/(1-\alpha)} \left(\frac{\alpha}{w}\right)^{\alpha/(1-\alpha)} \tag{72}$$

The commodity price still falls exponentially with distance according to (68). But since wages and transportation cost are in commodity units, production takes place at constant intensity up to a distance R determined by the demand at the center, $q(0)$.

3.3.2 Exchange of Two Commodities: Rotational Flow of Value

Consider first the budget constraint that applies to local excess demand when economic activities are restricted to trading in these two commodities and their transportation. Specifically, let local income be earned from local transportation plus trade. In locations that sell commodity 1 (and thus buy commodity 2), income is

$$y(\underline{x}) = -q_1 p^1 + k^1 |\phi^1| + k^2 |\phi^2| \tag{73}$$

and this income is spent on commodity 2

$$y(\underline{x}) = q_2 p^2$$

The budget constraint is therefore

$$q_1 p^1 + q_2 p^2 - k^1 |\phi^1| - k^2 |\phi^2| = 0 \tag{74}$$

and it has the same form in locations that sell commodity 2.

Under these conditions, an equilibrium exists in which the two commodity flows are opposite and equal in value terms. We first prove:

Theorem: *The flow of value $p^1 \phi^1 + p^2 \phi^2$ is purely rotational.*

Proof: Consider

$$\begin{aligned}
\operatorname{div}(p^1 \phi^1 + p^2 \phi^2) &= p^1 \operatorname{div} \phi^1 + p^2 \operatorname{div} \phi^2 + \operatorname{grad} p^1 \cdot \phi^1 + \operatorname{grad} p^2 \cdot \phi^2 \\
&= -p^1 q^1 - p^2 q^2 + k_1 \frac{\phi^1}{|\phi^1|} \cdot \phi^1 + k_2 \frac{\phi^2}{|\phi^2|} \cdot \phi^2 \\
&= 0 \qquad \text{by (74)}
\end{aligned}$$

On the boundary one has

$$p^1 \phi_n^1 + p^2 \phi_n^2 = 0 \tag{75}$$

by the boundary condition for a closed region. A flow field whose divergence is identically zero and does not transcend a boundary is purely rotational, for div rot = 0. *Q.E.D.*

The example of Figure 3.2 shows that, in a square region with production of each good localized on one of the two diagonals, a nonvanishing rotating value flow can occur as a solution. When k is isotropic, there is always another simpler solution: the two commodity flows are opposed. Given the first flow the strength of the second flow is chosen so as to make it equal in economic value at each point to that of the first flow. The price of the given commodity increases, while that of the second decreases along

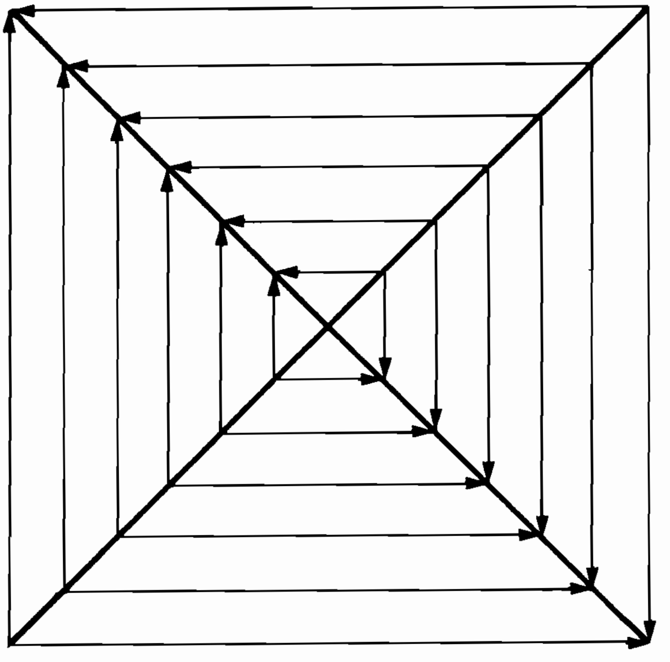

Figure 3.2. Flows of two commodities in a square region, where production of each good is localized on one half of the Northeast–Southwest diagonal.

the original flow lines. Then both gradient and divergence laws are satisfied for the second commodity so that this flow is an optimal solution. (Recall that the efficiency conditions are both necessary and sufficient for equilibrium.)

3.3.3 Extension to Multi-Commodity Trade

Suppose that transportation cost is distributed as income earned locally. Then the net flow of value is purely rotational

$$\operatorname{div} \sum_i p^i \phi^i = 0 \tag{75}$$

Note: In the case of more than two commodities one can no longer conclude that one flow is the opposite of any other, but rather that any flow is now the opposite of the combined value flow of the other commodities.

On the other hand, when incomes from transportation are earned by persons residing outside a subregion, there must be an outflow in value terms just sufficient to purchase the transportation services. Consider a subregion $A_0 \subset A$, which imports its transportation services

$$\int_{\partial A_0} \sum_i p^i \phi_n^i \, ds = \int\int_{A_0} \operatorname{div} \sum_i p^i \phi^i \, dx_1 \, dx_2$$

$$= \int\int_{A_0} \sum_i p^i \operatorname{div} \phi^i + \sum_i \phi^i \cdot \operatorname{grad} p^i \, dx_1 \, dx_2 \tag{76}$$

The budget constraint is now

$$q_1 p_1 + q_2 p_2 = 0 \tag{74a}$$

implying

$$(76) = \int\int_{A_0} \sum_i k_i |\phi^i| \, dx_1 \, dx_2 > 0$$

Therefore net exports are now positive.

3.3.4 Production of Two Commodities

Consider an economy producing two mobile products. Suppose that each location produces exactly one of these $i = 1,2$. (Compare the specialization theorem of Section 3.2.6.) Assume the local outputs to be given independent of price. Let the cost of transportation k be a given multiple of the price of one of the two commodities at a particular given location. Assume demand to be generated by a logarithmic utility function

$$u = \alpha_1 \log q_1 + \alpha_2 \log q_2 \tag{77}$$

The problem is to determine the equilibrium price distributions

$$p_i(x_1, x_2) \qquad i = 1,2$$

Let the income of location (x_1, x_2) be obtained through production only

$$y = \sum_i z_i p_i \tag{78}$$

(where only one of the two summands is positive). The demand functions for the two products are

$$q_i = \frac{\alpha_i y}{(\alpha_1 + \alpha_2) p_i} \qquad i = 1,2 \tag{79}$$

Without restriction assume that

$$\alpha_1 + \alpha_2 = 1$$

Now

$$q_i = \frac{\alpha_i}{p_i} \sum_j p_j z_j \qquad i = 1,2$$

If the chosen location exports product 1 then

$$q_1 = \frac{\alpha_1}{p_1} \cdot p_1 z_1 = \alpha_1 z_1 \tag{80}$$

$$q_2 = \frac{\alpha_2}{p_2} \cdot p_1 z_1 \tag{81}$$

The excess demand functions are therefore

$$q_1 - z_1 = (\alpha_1 - 1) z_1 = -\alpha_2 z_1 \tag{82}$$

$$q_2 - z_2 = \alpha_2 z_1 \cdot \frac{p_1}{p_2} \tag{83}$$

and the equilibrium conditions are

$$\operatorname{div} \phi_1 - \alpha_2 z_1 = 0 \tag{84}$$

independent of price for commodity 1, and

$$\operatorname{div} \phi_2 + \alpha_2 z_1 \frac{p_1}{p_2} = 0 \tag{85}$$

involving the price ratio p_1/p_2, for commodity 2.

$$k \frac{\phi_i}{|\phi_i|} = \operatorname{grad} p_i \qquad i = 1,2 \tag{86}$$

A consequence of the first two equilibrium conditions is that

$$p_1 \operatorname{div} \phi_1 + p_2 \operatorname{div} \phi_2 = 0 \tag{74a}$$

This is the local trade balance equation, stating that payments for imports must equal receipts from exports. Since transportation is supplied by outside agents, boundary condition (76) applies.

At points that export commodity 2, the indices must be interchanged in (84) and (85).

3.3.5 A von Thünen Equilibrium Model

Let the metropolis, considered as a point located at the center, be the producer and exporter of an industrial good and let the surrounding area of radius R produce an agricultural good. The outputs are Z_1 of the industrial good and z_2 per unit area of the agricultural good. The total quantities are then Z_1 and

$$Z_2 = \pi R^2 z_2 \tag{87}$$

Let transportation require only good 1 as an input, so that transportation cost is

$$k_i = h_i \cdot p_1(0) \tag{88}$$

where

$$p_i = p_i(r)$$

is the price of good i at distance r from the center. The income of the metropolis equals

$$Y(0) = p_i(0)Z_1 + p_1(0)\,T$$

where T is the total amount of transportation. We assume that the suppliers of transportation live in the metropolis. Absorption of good 1 in the metropolis equals

$$\begin{aligned} Q_1(0) &= \frac{\alpha_1}{p_1(0)}\,Y(0) \\ &= \alpha_1\,[Z_1 + T] \end{aligned} \tag{89}$$

Consumption of good 2 in the metropolis is, using (79)

$$\begin{aligned} Q_2(0) &= \frac{\alpha_2}{p_2(0)}\,Y(0) \\ &= \alpha_2\,\frac{p_1(0)}{p_2(0)}\,[Z_1 + T] \end{aligned} \tag{90}$$

The prices of the commodities at distance r from the center are then

$$p_1(r) = p_1(0) + h_1 p_1(0) \cdot r \tag{91}$$

for the industrial good, which flows outward, and

$$p_2(r) = p_2(0) - h_2 p_1(0) \cdot r \tag{92}$$

for the agricultural good, which flows to the center. Excess supply or export of the agricultural good per unit area at distance r from the metropolis is given by

$$z_2 - q_2 = \alpha_1 z_2 = \alpha_1 \tag{93}$$

If we standardize the agricultural output of one unit of land as unity, demand for the industrial good or imports per unit area at distance r, using (83), is given by

$$\begin{aligned} q_1 &= \alpha_1\,\frac{p_2}{p_1} \\ &= \alpha_1\,\frac{p_2(0) - h_2 p_1(0)\,r}{p_1(0) + h_1 p_1(0)\,r} \end{aligned} \tag{94}$$

The total flow of agricultural products from exporting areas beyond distance r through the cross section $2\pi r$ at distance r equals

$$F(r) = \int_r^R (z_2 - q_2)\, 2\pi r \,\mathrm{d}r$$
$$= \alpha_1 \pi (R^2 - r^2)$$

The flow through a unit cross section is then

$$|\phi(r)| = \frac{F(r)}{2\pi r} = \frac{\alpha_1}{2}\left(\frac{R^2}{r} - r\right) \tag{95}$$

and the transportation effort for the agricultural product is

$$T_2 = h_2 \int_0^R \int_r^R \alpha_1\, 2\pi r \,\mathrm{d}r$$
$$= h_2 \int_0^R r \cdot 2\pi r \cdot \alpha_1 \,\mathrm{d}r$$
$$= \tfrac{2}{3}\pi\, h_2\, \alpha_1 R^3 \tag{96}$$

The transportation effort for the industrial good is more complicated to calculate since the density of consumption falls with distance. In fact, using (94)

$$T_1 = \int_0^R h_1 r \alpha_1\, 2\pi r \cdot \frac{p_2(0) - h_2 p_1(0) r}{p_1(0) + h_1 p_1(0) r} \,\mathrm{d}r$$
$$= 2\pi\alpha_1 h_1 \left[p_2(0) \int_0^R \frac{r^2}{p_1(0) + h_1 p_1(0)\, r} \,\mathrm{d}r - h_2 p_1(0) \int_0^R \frac{r^3}{p_1(0) + h_1 p_1(0)\, r} \,\mathrm{d}r \right] \tag{97}$$

Substituting

$$T_1 + T_2 = T$$

from (96) and (97) in (89) and (90), one obtains the demand for both goods in the metropolis.

Equilibrium is now established by equating the demands for each good to the supplies Z_1 and $z_2 \pi R^2$. The supply of the industrial good will be assumed given (as a measure of city size), while the radius R of the agricultural supply area remains to be determined

$$Z_1 = \alpha_1 (Z_1 + T) + \alpha_1 \int_0^R \frac{p_2(0) - h_2 p_1(0) r}{p_1(0) + h_1 p_1(0) r}\, 2\pi r \,\mathrm{d}r \tag{98}$$

$$\pi R^2 \alpha_1 z_2 = \alpha_2 \frac{p_1(0)}{p_2(0)} (Z_1 + T) \tag{99}$$

The two equations (98) and (99), in conjunction with (96) and (97) for T_1 and T_2, determine the two unknowns R and $p_1(0)/p_2(0)$. von Thünen was perhaps wise to steer clear of an explicit analysis of this spatial equilibrium problem, the details of which are rather messy! We merely state here without proof that R is an increasing and $p_1(0)/p_2(0)$ a decreasing function of Z_1. The radius of the supply area and the relative price of the agricultural good increase with the size of the metropolis.

3.4 DYNAMICS

3.4.1 Dynamic Adjustment by Gradient Methods

Let λ, ϕ be a trial solution satisfying the boundary condition $\phi_n = 0$ but not necessarily the equilibrium conditions – the divergence law and the gradient law. Choose a suitable time unit and introduce the following rules of adjustment

$$\dot{\phi} = \operatorname{grad} \lambda - k \frac{\phi}{|\phi|} \tag{100}$$

$$\dot{\lambda} = -(q + \operatorname{div} \phi) \tag{101}$$

In this first rule, the vector

$$\operatorname{grad} \lambda - k \frac{\phi}{|\phi|}$$

indicates the direction of shipment that yields maximum profit or minimum loss. In the second rule

$$q + \operatorname{div} \phi$$

is the sum of local net demand and net exports. If this is positive, there is excess supply from this location and therefore prices should be lowered. Consider now the change of the Lagrangean cost integral $K = K(\phi, \lambda)$

$$\frac{dK}{dt} = \frac{d}{dt} \int\int k|\phi| + \lambda(\operatorname{div} \phi + q) dx_1 dx_2 \tag{102}$$

$$= \int\int k \frac{d}{dt} |\phi| + \dot{\lambda} \cdot (\operatorname{div} \phi + q) + \lambda \operatorname{div} \dot{\phi} \, dx_1 dx_2$$

Now

$$\frac{d}{dt} |\phi| = \frac{\phi \cdot \dot{\phi}}{|\phi|} \tag{103}$$

and

$$\int\int \lambda \operatorname{div} \dot{\phi} + \dot{\phi} \operatorname{grad} \lambda \, dx_1 dx_2 = \int\int \operatorname{div} \lambda\dot{\phi} \, dx_1 dx_2 = \int \lambda \dot{\phi}_n ds = 0$$

using the Gauss integral theorem and $\dot{\phi}_n = 0$ on ∂A.

Therefore

$$\int\int \lambda \operatorname{div} \dot{\phi} \, \mathrm{d}x_1 \mathrm{d}x_2 \;=\; -\int\int \dot{\phi} \cdot \operatorname{grad} \lambda \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{104}$$

Substituting (103) and (104) into (102)

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \int\int \dot{\phi}\left(\frac{k\phi}{|\phi|} - \operatorname{grad} \lambda\right) + \dot{\lambda}(\operatorname{div} \phi + q) \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{105}$$

Now the adjustment rules (100), (101) may be applied in (105) to yield

$$\frac{\mathrm{d}K}{\mathrm{d}t} \;=\; -\int\int \left|\operatorname{grad} \lambda - k\frac{\phi}{|\phi|}\right|^2 + [\operatorname{div} \phi + q]^2 \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{106}$$

This is strictly negative as long as any one of the equilibrium conditions is violated. It follows that the adjustment process converges to the equilibrium solution. Here

$$K[\phi, \lambda] - K_{\min} \tag{107}$$

where $K_{\min}$, the minimum value of the transportation cost integral, plays the part of a Liapunow function. It is nonnegative and vanishes only at the equilibrium, and it decreases monotonically with time.

3.4.2 Extension to Price-Dependent Demand

When excess demand is dependent on price, we have a maximand

$$M = \int\int (u(-\operatorname{div} \phi) - k|\phi|) \mathrm{d}x_1 \mathrm{d}x_2 \tag{108}$$

where the utility function is a consumers' surplus

$$u(q) = \int_0^q p(x_0, x_1, x_2) \mathrm{d}x_0 \;=\; u(q, x_1, x_2) \tag{109}$$

Let

$$p \;=\; p\,(-\operatorname{div} \phi, x_1, x_2)$$

throughout, so that local markets are always in equilibrium: excess supplies $-q(x_1, x_2)$ are equal to net exports $\operatorname{div} \phi$. The flow of product is adjusted according to the same rule as before

$$\dot{\phi} \;=\; \operatorname{grad} p - k\frac{\phi}{|\phi|} \tag{100}$$

while keeping regional outflows zero

$$\phi_n = 0 \qquad \text{on } \partial A$$

Now

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \int\int \frac{\partial u}{\partial q}\,\dot{q} - k\frac{\phi\,\dot{\phi}}{|\phi|}\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{110}$$

moreover

$$\frac{\partial u}{\partial q} = p \qquad q = -\operatorname{div}\phi$$

and

$$\int\int -p\operatorname{div}\phi\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int -\operatorname{div}(p\dot{\phi}) + \dot{\phi}\operatorname{grad}p\,\mathrm{d}x_1\,\mathrm{d}x_2$$

By the Gauss integral theorem

$$0 = \int p\dot{\phi}_n\,\mathrm{d}x = \int\int \operatorname{div}(p\dot{\phi})\mathrm{d}x_1\,\mathrm{d}x_2$$

so that

$$\int\int -p\operatorname{div}\dot{\phi}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int \dot{\phi}\operatorname{grad}p\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{111}$$

Substituting (100) into (110) and using (111)

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \int\int \phi\cdot\left(\operatorname{grad}p - k\frac{\phi}{|\phi|}\right)\mathrm{d}x_1\,\mathrm{d}x_2$$

and using (100)

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \int\int\left|\operatorname{grad}p - k\frac{\phi}{|\phi|}\right|^2\mathrm{d}x_1\,\mathrm{d}x_2 > 0 \tag{112}$$

This shows that flow adjustment alone is sufficient for convergence when prices are set to equilibrate local excess supply and net exports. The Liapunow function is here

$$M_{\max} - M[\phi]$$

3.4.3 Price-Flow Dynamics in a One-Commodity Model: Price Waves

Let us formulate another simple dynamic adjustment model. Suppose the local cost of transportation to be a linear function of flow intensity so that we have a slight congestion effect. We put, for convenience

$$k = |\phi|/2 \tag{113}$$

Accordingly, we get total transportation cost as

$$\int\int (|\phi|^2/2)\mathrm{d}x_1\,\mathrm{d}x_2 \tag{114}$$

This expression is to be minimized subject to a constraint on the divergence of the flow and one boundary constraint. Suppose that local excess demand is price dependent, so that we write $q(\lambda)$ with $q' < 0$. Accordingly

$$\operatorname{div}\phi = -q(\lambda) \tag{115}$$

As shown above, the condition for (114) to be a minimum subject to (115) is

$$\phi = \operatorname{grad}\lambda \tag{116}$$

A combination λ^*, ϕ^* that fulfills (115), (116) and the boundary constraint is a solution to the equilibrium problem.

The equations (115) and (116) can also be combined into the single equation

$$\operatorname{div}\operatorname{grad}\lambda = -q(\lambda) \tag{117}$$

which has some similarity to the Laplace and Poisson equations that appear in potential theory.

We examine two simple examples. First, let $q \equiv -4$ and $\lambda = r^2$, where $r = (x_1^2 + x_2^2)^{1/2}$. As $\operatorname{div}\operatorname{grad} r^2 = 4$, equation (117) is fulfilled. We have a case with a constant excess demand everywhere and a price that increases with the square of the distance from the origin. As $\phi = \operatorname{grad} r^2 = (2x_1, 2x_2)$, the flow is radially outward and the intensity $|\phi| = 2r$ increases linearly with the distance from the origin.

For the second example, set $q = -6r$ and $\lambda = r^3$. Then, once again, (117) is fulfilled. The price rise as we move away from the origin is somewhat steeper and so is the increase of intensity, $|\phi| = 3r^2$. Flows are still radial and outward.

We are now ready to introduce the dynamics. Put

$$\dot{\lambda} = a(\operatorname{div}\phi + q(\lambda)) \tag{118}$$

and

$$\dot{\phi} = b(\operatorname{grad}\lambda - \phi) \tag{119}$$

Equation (118) tells us that the rate of change of price is proportional to the extent to which the withdrawal of goods by the flow exceeds the desired excess supply at the local price, or the extent to which the delivery from the flow falls short of the excess demand. To make the process feasible, we have, as always with such adaptive processes, to introduce local stocks of goods. Equation (119) states that, if grad λ and ϕ have the same magnitude, then the direction of the flow is adjusted in proportion to the difference between their directions. If, on the other hand, grad λ and ϕ have the same direction, then (119) tells us that the intensity of the flow is adjusted to the extent to which it deviates from the rate of spatial price increase.

We now define the deviations from the equilibrium price and flow patterns

$$\mu = \lambda - \lambda^* \tag{120}$$

and

$$\psi = \phi - \phi^* \tag{121}$$

Recalling that λ^* and ϕ^* satisfy (115)–(116) and that they are time-invariant, we obtain by substitution into (118), (119)

$$\dot{\mu} = a(\text{div}\ \psi + q(\lambda^* + \mu) - q(\lambda^*)) \tag{122}$$

and

$$\dot{\psi} = b(\text{grad}\ \mu - \psi) \tag{123}$$

Making a Taylor expansion of $q(\lambda^* + \mu)$ around λ^* and truncating after the linear term we obtain the approximation

$$q(\lambda^* + \mu) - q(\lambda^*) = q'\mu \tag{124}$$

This expression is then substituted into (122).

Next, we differentiate (122) with respect to time, to get $\ddot{\mu} = a(\text{div}\ \dot{\psi} + q'\dot{\mu})$, and substitute from (123) to get $\ddot{\mu} = a(b\ \text{div grad}\ \mu - b\ \text{div}\ \psi + q'\dot{\mu})$. Finally, we substitute from (122) for div ψ and get

$$\ddot{\mu} + (b - aq')\dot{\mu} - abq'\mu = ab\ \text{div grad}\ \mu \tag{125}$$

This is a wave equation that includes a linear "friction" term, which may be expected to give stability to the model. To solve the equation we use the method of separating variables, i.e. we put $\mu = T(t)\,X(x_1, x_2)$ and obtain

$$T''/T + (b - aq')T'/T - abq' = ab(\text{div grad}\ X)/X \tag{126}$$

This equation can hold as an identity in space and time only if both its sides equal one and the same constant. Denoting this constant by κ, we get:

$$T'' + (b - aq')T' + (\kappa - abq')T = 0 \tag{127}$$

and

$$ab\ \text{div grad}\ X + \kappa X = 0 \tag{128}$$

Equation (127) is an ordinary differential equation with constant coefficients. Its roots are

$$-(b - aq')/2 \pm ((b - aq')^2 - 4(\kappa - abq'))^{1/2}/2 \tag{129}$$

As the real parts of the roots are negative, we see that any motion in time is damped, always provided that the roots are complex so that there is some oscillatory motion at all.

Equation (128) can again be treated by the separation of variables method, if we assume that the region has a simple form, such as a rectangle or circle. In the circular case, already used for illustration, (128) should be changed into polar coordinates: $r = (x_1^2 + x_2^2)^{1/2}$ and $\omega = \arctan\,(x_2/x_1)$. Assuming $X(x_1, x_2) = R(r)\ \Omega(\omega)$ we arrive at the two equations

$$R'' + R'/r + ((\kappa/ab) - n^2/r^2)R = 0 \tag{130}$$

and

$$\Omega'' + n^2\Omega = 0 \tag{131}$$

The procedure leading to these equations again follows a recognition of the fact that the separated solution $R\Omega$ only fits when the expressions involving only distance or only angle must equal an identical constant. This constant is n^2, where n is an integer. The reason for this is that (131) only makes sense when n^2 is a square of an integer, since only then will the angular wave in space end up at its initial value after a full round. Solutions to (131) are pure sine and cosine functions, representing waves of various lengths travelling circularly around the region.

Equation (130) is Bessel's differential equation, whose solutions are Bessel functions. These represent waves that travel radially between center and boundary. The principle of superposition allows us to use any linear combination of solutions to (130), (131) for various n. Therefore, we get a set of radial and angular waves of different wavelengths, travelling at various speeds. As κ links the waves in space and time together, we can conclude that the cycles travel faster the shorter their wavelengths. The exact mixture is determined by the boundary conditions that determine, for example, whether waves can travel undisturbed across the boundary or whether they are completely damped there.

As the price waves in μ, which we have been studying, die away, so do the changes in the gradient flows $\psi = \text{grad}\ \mu$. Ultimately the oscillating system, disturbed by some initial displacement from equilibrium, returns to the equilibrium state described by $\lambda^*, \phi^* = \text{grad}\ \lambda^*$.

3.4.4 Alternative Derivation of Price Waves

Consider now the case where transportation cost is negligible, but assume the spatial market to be in disequilibrium. The adjustment of flow is proportional to its profitability

$$\frac{\partial \Phi}{\partial t} = a\ \text{grad}\ p \tag{132}$$

The adjustment of price is proportional to excess demand

$$\frac{\partial p}{\partial t} = b\,[q + \text{div}\ \Phi] \tag{133}$$

Differentiating (133) with respect to time

$$\frac{\partial^2 p}{\partial t^2} = b\frac{\partial q}{\partial p}\cdot\frac{\partial p}{\partial t} + b\ \text{div}\frac{\partial \Phi}{\partial t}$$

Now substitute for $\partial\Phi/\partial t$ from (132), use the dot notation for derivatives with respect to time, and

$$q' = \frac{\partial q}{\partial p}$$

to obtain

$$\ddot{p} = bq'\dot{p} + ab\,\Delta p \tag{134}$$

where

$$\Delta = \operatorname{div}\operatorname{grad} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$$

In particular, if excess demand is independent of price (as in the simple continuous transportation model)

$$\ddot{p} = ab\,\Delta p \qquad ab > 0 \tag{135}$$

and this is the wave equation in its simplest form.

We will not discuss the resulting price waves, since their space and time profile depends crucially on the shape of the region. We will merely demonstrate the persistence of fluctuations. Multiplying (135) by $2\dot{p}$ one has

$$\frac{\mathrm{d}}{\mathrm{d}t}(\dot{p})^2 = 2\dot{p}\ddot{p} = 2ab\,\dot{p}\Delta p \tag{136}$$

Now

$$\dot{p}\operatorname{div}\operatorname{grad} p = \operatorname{div}(\dot{p}\operatorname{grad} p) - \operatorname{grad} p \cdot \operatorname{grad}\dot{p} \tag{137}$$

Integrating (137) over the region A and applying the Gauss integral theorem using the boundary condition $\operatorname{grad} p_n = 0$ on ∂A

$$\begin{aligned}\int\!\!\int \dot{p}\operatorname{div}\operatorname{grad} p\;\mathrm{d}x_1\,\mathrm{d}x_2 &= \int\!\!\int \operatorname{div}(\dot{p}\operatorname{grad} p) - \operatorname{grad} p \cdot \operatorname{grad}\dot{p}\;\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int \dot{p}\,(\operatorname{grad} p)_n\;\mathrm{d}s - \int\!\!\int \operatorname{grad} p \cdot \operatorname{grad}\dot{p}\;\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

Assume that the boundary condition remains satisfied, so that cross flows do not change

$$0 = \frac{\partial}{\partial t}(\Phi_n) = \left(\frac{\partial \Phi}{\partial t}\right)_n = a\,(\operatorname{grad}\dot{p})_n$$

Upon integration by parts, the line integral vanishes and we have

$$\begin{aligned}\int\!\!\int \dot{p}\operatorname{div}\operatorname{grad} p\;\mathrm{d}x_1\,\mathrm{d}x_2 &= -\int\!\!\int \operatorname{grad} p \cdot \operatorname{grad}\dot{p}\;\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int\!\!\int (\operatorname{grad} p)^2\;\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned} \tag{138}$$

Integrating (136) over the region A and substituting (138)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\int \dot{p}^2 + ab\,|\mathrm{grad}\, p|^2\, \mathrm{d}x_1\,\mathrm{d}x_2 = 0$$

or

$$\int\int \dot{p}^2 + ab\,|\mathrm{grad}\, p|^2\, \mathrm{d}x_1\,\mathrm{d}x_2 = \text{constant} \tag{139}$$

If the system was initially not in equilibrium, then the constant is positive. If the system ever passes through equilibrium, so that $|\mathrm{grad}\, p|^2 \equiv 0$, then

$$\int\int \dot{p}^2\, \mathrm{d}x_1\,\mathrm{d}x_2 = \text{constant} > 0$$

so that the motion continues. The system can never settle down to equilibrium and is unstable. In the case where excess demand

$$q = q(p)$$

is a strictly decreasing function of price, the same argument applied to (134) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\int \dot{p}^2 + ab\,|\mathrm{grad}\, p|^2\, \mathrm{d}x_1\,\mathrm{d}x_2 = \int\int bq'\,\dot{p}^2\, \mathrm{d}x_1\,\mathrm{d}x_2 < 0 \tag{140}$$

since $q' < 0$ everywhere. Thus price-dependent demand acts as a damping factor with regard to the fluctuations of p in time and space.

3.4.5 Diffusion of Expenditure

Let $y(x_1, x_2, t)$ represent wealth in terms of liquid assets in locations x_1, x_2 at time t. Assume that spending units tend to spend this money at their own and in adjacent locations $x_i \pm \Delta x$ at time $t + \Delta t$. In discrete terms

$$\begin{aligned} y(x_1, x_2, t + \Delta t) = {} & c_0 y(x_1, x_2, t) \\ & + c_1 [y(x_1 - \Delta x, x_2, t) + y(x_1 + \Delta x, x_2, t) \\ & + y(x_1, x_2 - \Delta x, t) + y(x_1, x_2 + \Delta x, t)] \end{aligned} \tag{141}$$

where c_0 is the propensity to spend at home and c_1 is the propensity to spend in any of the adjacent locations. Rewriting the bracketed term

$$\begin{aligned} & [y(x_1 + \Delta x, x_2) - y(x_1, x_2)] - [y(x_1, x_2) - y(x_1 - \Delta x, x_2)] \\ & + [y(x_1, x_2 + \Delta x) - y(x_1, x_2)] - [y(x_1, x_2) - y(x_1, x_2 - \Delta x)] \\ & + 4y(x_1, x_2) \end{aligned}$$

one observes that

$$\frac{y(x_1,x_2,t+\Delta t)-y(x_1,x_2,t)}{\Delta t} = (c_0+4c_1-1)y(x_1,x_2,t)\frac{1}{\Delta t}$$
$$+\frac{\Delta x^2}{\Delta t}\frac{c_1}{\Delta x}\left[\frac{\Delta_1 y(x_1+\Delta x,x_2)}{\Delta x}-\frac{\Delta_1 y(x_1,x_2)}{\Delta x}\right]$$
$$+\frac{\Delta x^2}{\Delta t}\cdot\frac{c_1}{\Delta x}\left[\frac{\Delta_2 y(x_1,x_2+\Delta x)}{\Delta x}-\frac{\Delta_2 y(x_1,x_2)}{\Delta x}\right]$$

Now assume that the total propensity to spend equals unity

$$c_0+4c_1 = 1 \tag{142}$$

Then the first term on the right-hand side disappears. Let the ratio between the time change and the space change be fixed at

$$\frac{\Delta x^2}{\Delta t} = \frac{m}{c_1} = \text{constant} \tag{143}$$

Going to the limit we obtain

$$\frac{\partial y}{\partial t} = m\Delta y$$

or

$$\frac{\partial y}{\partial t} = m\left(\frac{\partial^2 y}{\partial x_1^2}+\frac{\partial^2 y}{\partial x_2^2}\right) \tag{144}$$

the well-known diffusion equation. This equation may be given the following interpretation. Introduce the gradient field grad y. We observe that money flow in the region is then described by

$$\phi = -\operatorname{grad} y \tag{145}$$

and the change in liquid wealth is the divergence of this flow field. Thus liquid wealth can change only through the net yield of money flows

$$\dot{y} = -\operatorname{div}\phi = -m\operatorname{div}(-\operatorname{grad} y)$$
$$= m\Delta y$$

Add now the boundary condition $\phi_n = 0$ on ∂A. Then it follows at once from the Gauss integral theorem that

$$0 = \int \phi_n\,\mathrm{d}s = \int\int \operatorname{div}\phi\,\mathrm{d}x_1\mathrm{d}x_2 = \frac{1}{m}\frac{\mathrm{d}}{\mathrm{d}t}\int\int y\,\mathrm{d}x_1\mathrm{d}x_2$$

Total liquid wealth or money stock must remain unchanged.

Consider now an equal distribution of liquid wealth throughout the region

$$y(x_1, x_2, t) = \bar{y} \tag{146}$$

We now show that this is the only stable equilibrium. Consider

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int\int (y-\bar{y})^2 \,\mathrm{d}x_1\,\mathrm{d}x_2 &= \int\int (y-\bar{y})\cdot\dot{y}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= m\int\int (y-\bar{y})\Delta y\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad \text{using (144)} \\ &= m\int\int (y-\bar{y})\,\mathrm{div\,grad}\,y\,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

Now

$$(y-\bar{y})\,\mathrm{div\,grad}\,y = \mathrm{div}\,[(y-\bar{y})\,\mathrm{grad}\,y] - \mathrm{grad}\,y\cdot\mathrm{grad}\,y$$

Also, by the Gauss integral theorem

$$\int\int \mathrm{div}\,[(y-\bar{y})\,\mathrm{grad}\,y]\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int (y-\bar{y})\phi_n\,\mathrm{d}s = 0$$

by the boundary condition.

Substituting

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int\int (y-\bar{y})^2\,\mathrm{d}x_1\,\mathrm{d}x_2 = -m\int\int (\mathrm{grad}\,y)^2\,\mathrm{d}x_1\,\mathrm{d}x_2 < 0 \tag{147}$$

proving the assertion.

Suppose now that spending by residents of a particular location is not financed entirely by net receipts from trade but by outside earnings as well. Let $z(x_1, x_2)$ be the density of these net earnings, which may be negative. In that case residents of the location pay wages to outsiders, presumably out of earnings from trade. The expenditure or money flow equation is then modified as follows

$$\frac{\partial y}{\partial t} = z - m\,\mathrm{div}(-\,\mathrm{grad}\,y) \tag{148}$$

Constancy of aggregate disposable wealth or aggregate liquid assets implies that

$$\begin{aligned}0 &= \int\int \frac{\partial y}{\partial t}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int\int z\,\mathrm{d}x_1\,\mathrm{d}x_2 + m\int (\mathrm{grad}\,y)_n\,\mathrm{d}s \qquad (149) \\ &= \int\int z\,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

Thus, in equilibrium, aggregate net earnings z from outside locations must add up to zero. Now let y^* be the stationary solution of this system

$$z + m\,\mathrm{div\,grad}\,y^* = 0 \tag{150}$$

Consider

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int\int(y-y^*)^2\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \int\int(y-y^*)\frac{\mathrm{d}y}{\mathrm{d}t}\,\mathrm{d}x_1\,\mathrm{d}x_2\\ &= \int\int(y-y^*)(z+m\,\mathrm{div\,grad}\,y)\,\mathrm{d}x_1\,\mathrm{d}x_2 \quad \text{using (148)}\\ &= m\int\int(y-y^*)\mathrm{div\,grad}(y-y^*)\mathrm{d}x_1\,\mathrm{d}x_2 \quad \text{using (150)}\\ &= m\int\int\mathrm{div}\,[(y-y^*)\,\mathrm{grad}\,(y-y^*)]\,\mathrm{d}x_1\,\mathrm{d}x_2\\ &\quad -m\int\int[|\mathrm{grad}(y-y^*)|]^2\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad (151)\\ &< 0\end{aligned}$$

since the first integral vanishes by the Gauss integral theorem and the boundary condition. This shows the stability of the equilibrium solution y^* of (150).

3.4.6 A Migration Model

This section contains a brief restatement of a migration model based on the hypothesis that, in a competitive spatial labor market, migration is motivated by interlocal differences in (real) wage rates. Formally the equations describing equilibrium in a spatially extended commodity market remain applicable. Suppose, however, that unlike in the commodity model there exists a positive relationship between the size of the flow and the gradient of wages (cf. Section 2.4.3). This may be due to the fact that moving costs are difficult to know precisely, that they vary among households, and that they may contain a significant nonmonetary element. Then it is not unreasonable to postulate

$$\phi = m\cdot\mathrm{grad}\,w \qquad m > 0 \tag{152}$$

With m a positive constant this simple hypothesis states a proportional relationship between the size of a migration stream and the magnitude of the wage gradient.

Consider now a stationary economy in which the demand for labor in each location is constant over time, but of course different between locations. Assume that increments $-q(x_1, x_2)$ of populations are also given and independent of wages, but that they may depend on location. A stationary equilibrium is possible only if the overall increase of population, aggregated over the region as a whole, is zero. This hypothesis of zero aggregate population growth means that

$$\int\int_A q\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0 \tag{153}$$

Assume also that no immigration or emigration takes place from the region

$$\phi_n = 0 \qquad \text{on} \quad \partial A$$

The divergence law is, as before

$$\operatorname{div} \phi + q = 0$$

Combining this and (152), and assuming m to be uniform

$$\operatorname{div} \operatorname{grad} w = -\frac{q}{m} \qquad \text{in} \quad A \tag{154}$$

$$(\operatorname{grad} w)_n = 0 \qquad \text{on} \quad \partial A \tag{155}$$

This Poisson equation may be solved by means of a Green function

$$w(x_1, x_2) = \frac{1}{m} \cdot \int \int G(x_1, x_2, \xi_1, \xi_2) q(\xi_1, \xi_2) \mathrm{d}\xi_1 \mathrm{d}\xi_2 \tag{156}$$

In simple cases the Green function depends only on distance

$$r = [(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2]^{1/2}$$

In the formulation (156), wages appear as dependent on the population increases q, weighted by their distance from a given location (x_1, x_2). This is a formalization of the idea of "locational potential" propounded by Zipf (1941), Stewart (1958), and Warntz (1959). In particular, if we assume an unbounded region and postulate that the potential function remains regular, the Green function is given by the logarithmic potential

$$G = \frac{1}{2\pi} \log \frac{1}{r} \tag{157}$$

so that

$$w(x_1, x_2) = \frac{1}{2\pi m} \cdot \int \int q(\xi_1, \xi_2) \log [(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2]^{-1/2} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \tag{158}$$

The locational potential assumes here the form of a logarithmic potential function. Locations that absorb migrants ($q > 0$) have an increasing effect on wages, while locations that emit migrants have a decreasing effect, and in both cases this effect is proportional to the logarithm of the inverse of distance.

A three-dimensional interpretation of the flow relationships is required in order that the locational potential should take the form of Newton's potential. In that case the gravity law may be applied to the interaction of populations at a distance where population is considered in terms of flows rather than stocks (Beckmann 1957).

4 Long-Run Equilibrium of Trade and Production

4.1 INTRODUCTION

The equilibrium models examined in previous chapters were trade models for one, or occasionally, two commodities. Excess demand functions were taken as given. In this chapter we move towards traditional general equilibrium theory by explicitly formulating the production and consumption decisions that underlie these excess demand functions.

Thus, one decision unit of each kind – one firm and one household – will be associated with each location. This coexistence need not be taken too literally. The micro-structure of a location may well involve a physical division of space between households and firms in a realistic way. The areal densities of population and production we are interested in, however, are perceived on a more macroscopic level than that of a detailed city plan.

4.1.1 Firms

We assume that only one good is produced according to a neoclassical production function with traditional substitution among inputs. This good is assumed to be the single commodity consumed by the households, i.e. it is composite and includes housing. The same commodity is assumed to be invested as capital. The fact that it is also assumed to be used up in transportation in the von Thünen manner should not be thought of in terms of a horse eating the goods it is carrying. All that the one-commodity assumption amounts to is the specification of an identical production technology for consumer goods, capital goods, housing, and transportation. This does not seem too unrealistic as a first approximation.

We also assume that the same production technology is available at all locations. Local differences in production densities and choices among various input mixes are thus endogenously explained by the model. These depend on local prices, which in turn depend on proximities to input supplies and market places and on transportation costs. This is the classical land-use problem of deciding which type of production to establish at a certain location. More specifically, we ask what should be done at a certain place, and not where a certain activity should be carried out. The decision is thus of the von Thünen rather than the Weber type. The latter is relevant when we deal with colonization of previously empty space. When all space is already occupied, the landlord can only choose the use to which he will put his land, since he cannot transfer his piece of land to a place where the land rent is higher. Therefore the question of what to do where is more relevant than the question of where to do what.

Production depends upon three primary inputs: capital, labor, and land. We assume the production function to be linearly homogeneous, which implies that we can divide throughout by land input to obtain the areal density of production as a function of the areal densities of capital and labor used. This procedure makes land disappear as an explicit input; it only remains reciprocally in the areal densities. The reason for this is that housing and transportation services are *produced* (along with capital goods and consumer goods). These activities use up output, while the use of primary inputs such as land has already been accounted for elsewhere. Consequently production is the only activity which uses land.

Capital is invested from current output at each location. In a long-run equilibrium, capital stock is constant and we deal only with the replacement of used-up capital. Taking the depreciation rate of capital as a universal constant, capital cost in commodity terms equals this constant. Accordingly, an optimum capital stock is maintained if and only if the marginal productivity of capital everywhere equals this depreciation constant. It follows from this that capital incomes are entirely reinvested and that net profits are due to landlords as land rent. Accordingly, land rent is determined residually by the local profitability of land use in production.

4.1.2 Households

If the households at each location owned the firms located there, no interregional trade could occur, since trade would balance locally everywhere. Therefore we assume an ownership mapping that maps locations of owners to locations of property. This allows for ownership concentration and absentee ownership. As a result we also obtain the much more interesting aspects of trade and specialization among locations.

The households have to find the optimum use for their disposable income. Part of their income originates from land ownership (= profits) and constitutes nonlabor income. The rest of their income is obtained from labor. In other parts of this study we consider commuting and labor flows. To simplify, we disregard commuting at this stage of the analysis, and assume that all potential working time, except the share devoted to leisure and household work, is used as an input to the local firm.

Each household decides upon the quantity to be consumed (including housing) and the quantity of leisure (including household work) that are optimal in view of the local price and wage rate. The quantities are determined within the bounds of each household's income, which is composed of land rents and the value of all potential working time.

As in the case of the firms' decisions concerning how much capital to invest at each location, the households are assumed to decide an optimum size of local labor stock (family size, or whatever we choose to call it). Technically, this is achieved by assuming an individual utility function, depending on per capita consumption and per capita leisure, with each household maximizing the product of family size and individual utility as a measure of its total utility. This maximization is achieved by choosing an optimal family size (= labor force).

This construction is assumed to replace a locational decision by the households. Again, the question is what should be done at a certain location (i.e. which family size should be chosen), and not where should a certain activity be located (i.e. where the residence for a given family should be). Once again, the principle of where to do what, instead of what to do where, would perform poorly as a modeling instrument. The households would choose a combination of the highest real wage rate and the lowest price level, with the exact balance depending on the size of their nonlabor incomes. The optimal locations would then, at most, be on a set of nonzero linear measure. We have already noted that the firms would similarly choose locations of maximum land rent if the "where to do what" principle were applied. The result would be that both firms and households would crowd into sets of area measure zero. This demonstrates that determination of optimal locations in the Weberian manner is a poor modeling principle for investigating the structure of a space that is already occupied. It works for the colonization of empty space, but breaks down in our case where the von Thünen "locational" decision concerning the best activity at each location is preferable.

As applied to the household in particular, this principle may seem somewhat absurd at first, but it is just as logical as in the case of the firm.

4.1.3 Trade

Once the households have chosen the optimal labor stocks and the firms have chosen the optimal capital stocks, local balances for labor and capital are obviously established everywhere. Implicitly, we also have a local balance for the land market, since all land is used in production. This was assured by the definition of the areal densities.

Unlike the case for inputs, there is not necessarily any local balance for the output of the produced good. A certain quantity of the malleable commodity is produced at each location. This constitutes the supply. On the demand side, we have local household consumption (which, as we know, includes housing). Added to this are the need for reinvestment to replace used-up capital goods, in a certain proportion to the size of capital stock maintained, and the use of the commodity for transportation, in proportion to the flow volume. This last proportion may vary from one location to the next, depending on the variation in natural obstacles to road construction and transportation. (Linearity with respect to flow volume does not necessarily mean abstraction from congestion, as road maintenance costs are assumed to be included along with locomotion costs in the multiplicative factor.)

By subtracting the three items of demand (consumption, investment, and transportation) from supply (production of commodities), we obtain a local excess supply. It need not be zero, and generally it does vary from one location to another. It enters the commodity flow, if positive, and is withdrawn from it, if negative.

Globally, as a consistency condition for equilibrium, we could assume that excess demand should vanish for the whole region under study. This would mean insulation. Instead, we use the weaker assumption that trade with the exterior balances, i.e. that in aggregate the value of exports at sections of the boundary with an outflow of commodities equals the value of imports at sections with an inflow.

4.2 PRODUCTION

4.2.1 Constant Returns

According to our assumptions, production can be carried out according to a general neoclassical production function

$$Q = F(K, L, M) \tag{1}$$

which is homogeneous to the first degree in the three inputs: capital, labor, and land. Land is used as space for production and as a source of raw materials, which, through the application of capital and labor services, are converted into the finished product. Given such broad categories of inputs the assumption of linear homogeneity is not too unrealistic.

Increasing returns are often assumed in regional science, more often, in fact, than in general economics. This may be due more to a desire to arrive at certain conclusions concerning agglomeration of productive activities than to any conviction about the realism of the assumption itself. Moreover, it frequently appears that authors who suggest increasing returns, or externalities, or both, actually have accessibility to other important activities in mind. If the facility of communication is explicitly accounted for in the model, there does not seem to be any reason to add an assumption about external effects or increasing returns. All this would be more clear-cut in a model more disaggregated than the one introduced below. But, it is not unreasonable to assume that our single production function (which actually represents many different productive activities) is the analytical expression for a microcosm of all kinds of activities that require mutual accessibility. Then the scale of operations could be reduced or increased, in any proportion, and this would amount to nothing more than linear homogeneity. We do not argue that linear homogeneity cannot be criticized; we only imply that the particular reasons for assuming increasing returns in regional science applications may be weaker than is generally thought.

Accordingly, we can divide through by M, and obtain

$$Q/M = F(K/M, L/M, 1) \tag{2}$$

which demonstrates that the areal density of output depends only on the areal densities of capital and labor. We denote the areal densities by lowercase letters and write

$$q = f(k, l) \tag{3}$$

This function, of course, is no longer linearly homogeneous. It displays decreasing returns in the areal densities of capital and labor.

4.2.2 Marginal Conditions

We can now, as usual, write the marginal conditions for profit maximization. These state that the marginal value productivities of capital and labor should equal the prices of capital and labor services, i.e. the capital rent and the wage rate, respectively. Now, as the produced good itself is invested as capital, capital rent is proportional to the commodity price, where the proportionality factor is the constant rate of capital depreciation. Thus, dividing both the value productivity of capital and capital rent by commodity price, we conclude that the marginal productivity of capital must equal the depreciation rate. Denoting this latter constant by κ, we write

$$f_k(k, l) = \kappa \tag{4}$$

For labor we obtain the more familiar statement that its marginal productivity must equal the real wage rate. Formally

$$f_l(k, l) = w/p \tag{5}$$

Accordingly, we conclude that the real wage rate alone completely determines the areal capital and labor densities used in production, and hence also the resulting output density of commodities.

All the variables, the areal densities of capital, labor, and output, as well as prices and wages, are functions of the location coordinates x_1, x_2. We observe that they are not expressly included in the production function, which means that locational productivity differences are disregarded.

Now, according to Euler's theorem for homogeneous functions, the value of output is exhausted in the income shares of inputs. Denoting land rent by g, we can write

$$pQ = \kappa pK + wL + gM \tag{6}$$

or

$$g/p = q - \kappa k - (w/p)l \tag{7}$$

As we have seen the real wage rate determines output and inputs. Accordingly, it also determines the ratio of land rent to commodity price. We also note that there is no optimum condition for land. The original production function could have been used to obtain such a condition, but its fulfillment would be automatically guaranteed by Euler's identity. Land rent is thus determined residually from the profitability of land in production.

As for the producers' optima, we only have to add the assumption that the relevant second-order conditions are fulfilled. We also note that if the inputs are assumed to be substitutes, which is reasonable, then either a capital-intensive or a labor-intensive technology is chosen, according to whether the real wage rate is high or low.

We also conclude that, unless production is "regressive" in the Hicksian sense, production density will be higher, the lower the real wage rate; see Hicks (1946, p. 93).

4.3 CONSUMPTION

4.3.1 Ownership

We now know how the land rents are determined residually as net profits of the firms. As usual in general equilibrium economics, an ownership pattern should be specified to indicate which households own which firms. The most general procedure currently utilized in economics is to assume an (arbitrary) partitioning of the profits of each firm among all the households according to some shareholding pattern. As our case involves a nondenumerably infinite number of households, we would need a continuous density function defined throughout the region for the distribution of the profits of each firm. As the firms are also nondenumerably infinite in number, we would need a continuum of such density functions to specify the ownership pattern. This would by no means be impossible, but in order to avoid formalism we prefer a simpler construction where each firm has one and only one owner. The results are as rich as those obtained from the alternative construction, but the formalism is much less complex.

We can specify the entire ownership pattern by one single mapping that associates locations of households (denoted x_1, x_2) to locations of property (denoted ξ_1, ξ_2). The mapping

$$\xi_1 = \xi_1(x_1, x_2) \tag{8}$$

$$\xi_2 = \xi_2(x_1, x_2) \tag{9}$$

maps the region considered into itself. It is assumed continuously differentiable. In brief, we shall denote it $\xi = \xi(x)$.

Of course, this structure admits absentee ownership (with the sole exception of the necessary fixed point). What may be less obvious intuitively is that any degree of ownership concentration we would like to represent is admitted. Defining the Jacobian of the mapping as

$$J(x_1, x_2) = \partial(\xi_1, \xi_2)/\partial(x_1, x_2) \tag{10}$$

we know that the households in the small rectangle of area $\mathrm{d}x_1\,\mathrm{d}x_2$ own all the property contained in the small rectangle of area $\mathrm{d}\xi_1\,\mathrm{d}\xi_2 = J\mathrm{d}x_1\,\mathrm{d}x_2$. The households are wealthy or poor, depending on whether the value of J is high or low.

The nonlabor incomes of the households can now be defined by

$$y(x) = g(\xi(x))J(x) \tag{11}$$

As we recall, capital incomes are always reinvested, labor incomes are paid to the local workers (= residents), and what remains (land rent or net profits) are due to the owner. These land rents are first shifted to the locations of the owners and then corrected for the areal distortion introduced by the ownership mapping. This is accomplished through multiplication by the Jacobian.

An interesting fact is that, since the mapping $\xi(x)$ is a simple change of coordinates, we get

$$\int\int_A y(x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2 = \int\int_A g(\xi_1, \xi_2)\mathrm{d}\xi_1\,\mathrm{d}\xi_2 \tag{12}$$

Since the notation of the integration variables is immaterial, we conclude that the aggregate nonlabor incomes equal the aggregate land rents. This, of course, is due to the fact that $\xi(x)$ maps the region A into itself.

4.3.2 Household Decisions

We are now able to state the budget constraints of the households. Denoting consumption by q', leisure (including time for household work) by l', and all potential working time by L, we get

$$pq' + wl' = y + wL \tag{13}$$

In the following text, uppercase letters K and L are used to denote totals of capital and labor. In the case of capital we simply have $k = K$, as all capital is employed in production. Labor, however, is partitioned into working time and leisure, formally $L = l + l'$.

As implied above in the Introduction, we write the objective function as

$$LU(q'/L, l'/L) \tag{14}$$

Maximizing total utility with respect to the choice of consumption and leisure, subject to the budget constraint, we arrive at the conditions

$$\frac{\partial U}{\partial(q'/L)} = \lambda p \tag{15}$$

and

$$\frac{\partial U}{\partial(l'/L)} = \lambda w \tag{16}$$

Obviously, these first-order optimum conditions, along with the budget constraint, serve to determine the demand for consumer goods (including housing) and leisure (including household work), provided we know the price, wage rate, and nonlabor income of the household. Nonlabor income again depends on the price and wage rate, but at the location of the property, and on the ownership mapping. As the latter is taken as given, then once the prices and wage rates are given, the households' and firms' decisions are also given. By determining the part of potential working time reserved for leisure and household work, the household also determines the remaining share, which is the labor input supplied to the local firm.

This holds provided the total amount of potential labor L is known. Now, we assumed that the firms would choose to maintain a capital stock that would make marginal productivity equal to the rate of depreciation. Similarly, total potential working time is assumed to be subject to optimization.

4.3.3 Optimal Size of Local Population

According to our objective function $LU(q'/L, l'/L)$, an increase in L would have three effects. Obviously, more people would profit from any given individual utility, and this would increase the value of the objective function. On the other hand, individual utility would be reduced for any given quantities of consumption q' and leisure l', as they would have to be shared by more people. Third, we see from the budget constraint that the wealth of the household would be increased by total potential working time, thus enabling an increase in both consumption and leisure. Population size is optimal at a particular location if these three effects are in exact balance so that there are no incentives to increase or decrease population. We now establish the condition for optimal local population (= potential labor stock) and implicitly assume that the relevant second-order conditions are satisfied. (This is ensured by the reasonable assumption of "decreasing returns to scale" for the utility function, analogous to the assumptions concerning the production functions.)

Formally, the product LU is differentiated with respect to L, while ensuring that all three of the above effects are taken into account. By using the marginal optimality conditions for consumption and leisure we can write

$$\frac{\mathrm{d}(LU)}{\mathrm{d}L} = U - \frac{\lambda}{L}(pq' + wl') + \lambda\left(p\frac{\mathrm{d}q'}{\mathrm{d}L} + w\frac{\mathrm{d}l'}{\mathrm{d}L}\right) \tag{17}$$

where we recognize the three effects.

However, if we differentiate the budget constraint with respect to L (regarding price and wage rate as given), we get

$$p\frac{\mathrm{d}q'}{\mathrm{d}L} + w\frac{\mathrm{d}l'}{\mathrm{d}L} = w \tag{18}$$

which fits nicely into the last parenthetical term of the preceding expression. At the same time we can substitute $pq' + wl' = y + wL$ into the first parenthetical term from the budget constraint and, after cancelling equal terms of opposite sign, obtain

$$\frac{\mathrm{d}(LU)}{\mathrm{d}L} = U - \frac{\lambda y}{L} \tag{19}$$

Setting this equal to zero as a condition for optimum population, we have

$$LU = \lambda y \tag{20}$$

As λ is the Lagrangean multiplier of the budget constraint, it should be interpreted in terms of marginal gain from relaxing the constraint. In fact, calculation shows that

$$L\frac{\mathrm{d}U}{\mathrm{d}y} = \lambda \tag{21}$$

Substituting this into the optimum condition, the condition then reads

$$\frac{\mathrm{d}U}{\mathrm{d}y}\frac{y}{U} = 1 \tag{22}$$

which states that the elasticity of per capita utility with respect to income from property should be unity. By again assuming "decreasing returns to scale," we expect optimal population to be larger, the higher the nonlabor income.

4.4 TRADE AND EQUILIBRIUM

4.4.1 Introduction

As we have already indicated, the capital and labor markets are in local equilibrium. Strictly speaking, the capital market does not exist. The firms simply maintain capital stocks such that the marginal productivities equal the depreciation rate. So, the capital stocks $K = k$ are determined endogenously by the system.

As for labor, the total stock is divided in two parts, labor outside the household and leisure, formally $L = l + l'$. As l' is determined by the household and l by the firm, there is actually a local market for labor. The obvious variable for arranging the balance of supply and demand is the wage rate. So, if L is given, wages are then determined by supply and demand at each location.

The situation is complicated somewhat by the fact that L itself is subject to optimizing choice. In the long-run equilibrium, the local households adjust their sizes to their nonlabor incomes, in the way indicated in the preceding section. As these nonlabor incomes are land rents, whose size depends on the prices and wages at the property locations, the price–wage system becomes linked interlocally. But it is still correct to say that wages are determined locally on the labor market. Of course, these wages depend on the distribution of population. The latter becomes endogenous in the model, where the additional degree of freedom corresponds to the additional condition for optimal family size.

So far we have accounted for how the equilibria for inputs are achieved and how their total quantities are determined. We now turn to the output market. Unlike the case for inputs, there are generally no local equilibria for output. There is one component of supply at each location (local output). There are also three items of demand (consumption, reinvestment to replace used-up capital, and transportation of the commodity itself).

4.4.2 Local Excess Supply or Demand

In order to state the continuity equation for how the trade volume changes with local excess supply and excess demand, we begin by introducing the flow of trade in commodities. As usual throughout this study it is a vector field

$$\phi = (\phi_1(x_1, x_2), \phi_2(x_1, x_2)) \tag{23}$$

where the direction $\phi/|\phi| = (\cos\,\theta,\ \sin\,\theta)$ is defined as the actual direction of flow, and the norm $|\phi| = (\phi_1^2 + \phi_2^2)^{1/2}$ is defined as the volume of flow.

According to our introductory assumption a given proportion of the commodities traded is used up to provide for transportation. We denote this proportion

$$h = h(x_1, x_2) \tag{24}$$

In general it depends on location x, whereas it does not depend on the direction θ. The transportation cost is thus isotropic. It may be useful to recall that the assumption whereby the good itself is used up as it is transported merely implies that the good is produced according to the same production technology as is transportation. Later on we deal with cases where the technologies are different. It should also be kept in mind that the function $h(x)$ represents maintenance as well as locomotion costs, thus making it reasonable to assume that local transportation costs (in commodity terms) depend linearly on the volume of flow. Hence,

$$h|\phi| \tag{25}$$

is the transportation cost in commodity terms at each location. Multiplying by the commodity price, we convert this cost into monetary terms. Accordingly, total transportation costs are

$$T = \iint_A ph|\phi| \mathrm{d}x_1 \mathrm{d}x_2 \tag{26}$$

As q is now the quantity produced locally, q' is the quantity consumed, κk is the quantity reinvested as capital, and $h|\phi|$ is the quantity used up in transportation, excess supply equals $q - q' - \kappa k - h|\phi|$. Since this enters the flow we have the condition

$$\operatorname{div} \phi = q - q' - \kappa k - h|\phi| \tag{27}$$

which is a kind of local equilibrium condition for output. The market does not have to clear, but if it does not the excess must enter into the flow, or if negative, be withdrawn from it.

4.4.3 Optimum of Transportation

We are now in a position to consider the optimum of transportation. It is immaterial whether we assume that transportation costs are paid by consumers, producers, or some kind of specialized transportation enterprise. The outcome concerning optimal transportation is the same in all three cases. So, to be precise, let us suppose that specialized transportation enterprises buy everything that is produced at prevailing local prices and sell it to all the households and investing firms, again at prevailing local prices. We assume that these firms are numerous enough to act competitively. Their aggregate revenues are

$$\iint_A pq' \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{28}$$

from sales to consumers and

$$\iint_A \kappa pk \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{29}$$

from sales to investors, and their costs are

$$\iint_A pq \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{30}$$

due to purchases from the producers. The net profits are

$$\iint_A p(q' + \kappa k - q)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{31}$$

which, according to the equilibrium condition for the flow, equals the negative of

$$\iint_A p(\operatorname{div}\phi + h|\phi|)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{32}$$

As there are no externalities, the transportation enterprises act so as to maximize their aggregate profits, i.e. minimize the preceding expression. They minimize this (p and h are exogenous to them) by the choice of the flow ϕ. Now, minimizing with respect to ϕ is again a well-defined variational problem with

$$ph\frac{\phi}{|\phi|} = \operatorname{grad} p \tag{33}$$

as the appropriate Euler equation. As usual, this equation only determines the flow directions. But the divergence equation above serves to determine flow volume.

4.4.4 Some Implications

We have now derived all the equilibrium and optimum conditions. The last condition has the usual implications: that goods are shipped in the directions of the steepest price increase, and that in these directions prices increase by accumulated transportation costs.

From the last equation, after multiplying by ϕ, we readily obtain the useful relation

$$ph|\phi| = (\operatorname{grad} p)\cdot\phi \tag{34}$$

When this is substituted into the expression for aggregate profits for transporters (32), the latter expression is changed into the negative of

$$\iint_A (p \operatorname{div}\phi + \phi\cdot\operatorname{grad} p)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{35}$$

Now, we can substitute the elementary identity from vector analysis $\operatorname{div}(p\phi) = p \operatorname{div}\phi + \phi\cdot\operatorname{grad} p$, and apply Gauss's divergence theorem to obtain profits as

$$\iint_A \operatorname{div}(p\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\partial A} p(\phi)_n\,\mathrm{d}s = 0 \tag{36}$$

This last expression equals zero because $(\phi)_n$ denotes outflow of commodities in the direction normal to the boundary, $p(\phi)_n$ is value exports, and hence the curve integral is the net value export surplus for the whole region. Owing to our assumption of a trade balance with the exterior, this net surplus is zero.

Accordingly, we conclude that aggregate profits for the transportation enterprises equal the import surplus for the region, and that they are zero when trade with the exterior balances. This is as it should be in the presence of perfect competition and free entry in the long run. The fact that there are no net incomes from the transportation activity justifies our argument that it is immaterial whether producers, consumers, or specialized transporters are assumed to provide the transportation services.

4.4.5 Walras' Law

When the conclusion that aggregate profits from transportation are zero is applied to our initial expression ((32)), we have

$$\iint_A p(\operatorname{div} \phi + h|\phi|)\mathrm{d}x_1\mathrm{d}x_2 = 0 \tag{37}$$

which may be used to obtain

$$T = \iint_A ph|\phi|\mathrm{d}x_1\mathrm{d}x_2 = -\iint_A p \operatorname{div} \phi \,\mathrm{d}x_1\mathrm{d}x_2 \tag{38}$$

After substituting the divergence equation, $\operatorname{div} \phi = q - q' - \kappa k - h|\phi|$, into the cost expression and removing aggregate transportation costs from both sides, we obtain

$$\iint_A p(q - q' - \kappa k)\mathrm{d}x_1\mathrm{d}x_2 = 0 \tag{39}$$

This equation states that the aggregate value of all excess supplies, evaluated at local prices, is zero for the whole region. This last equation is derived from the equilibrium condition, which states that local excess supplies are entered into and local excess demands withdrawn from the flow, the condition that trade with the exterior balances, and the condition that transportation routes are optimally chosen. This was the only information used. In particular, the equations related to the individual producers and consumers were not used.

Let us now see what we can find out from these latter conditions. First, the budget constraints indicate that $pq' + wl' = y + wL$. According to the definition $L = l + l'$, the budget constraints can be written

$$pq' - wl = y \tag{40}$$

On the other hand, the fact that the revenues of the producing firms were exhausted in value shares of the inputs (or, if we prefer, that profits are due to the landlord as rent), implies that

$$pq - \kappa pk - wl = g \tag{41}$$

But we have defined $y = qJ$, and noted the fact that the area integral of nonlabor incomes equals the area integral of land rents. (This was simply due to the fact that multiplication by the Jacobian merely implied a change in coordinates.)

So, the integrals on the right-hand side of the last two expressions taken over the whole region must be equal, and the same must obviously apply to the left-hand side. Equating the integrals of the left-hand side and cancelling the integrals of wl, we obtain

$$\iint_A p(q - q' - \kappa k)\mathrm{d}x_1\,\mathrm{d}x_2 = 0 \tag{42}$$

which again states that the aggregate value of local excess supplies is zero. This relation was derived in a completely different way from the identical statement above. In fact, this new statement provides us with a sort of Walras' law. The fact that this is identical to the statement obtained from aggregate market clearing and trade balance relations provides a check on the consistency of the entire model construction.

4.5 HOW THE MODEL WORKS

4.5.1 Introduction

We are now in a position to see how the model functions. Its core is made up of two partial differential equations that must be solved in sequence, after which the rest of the equilibrium solution is easily determined. The first differential equation determines commodity prices and is derived from the "gradient law" (the optimality condition for the flow directions). The second determines flow volume and is derived from the "divergence law" (the condition that states how excess supply is entered into or excess demand withdrawn from the flow). Of course, appropriate boundary conditions are required in both cases. The first case contains "world market" prices, established by trade in the exterior and hence on the boundary. The second case involves trade volumes, likewise established by exterior trade and crossing the boundary. In both cases the boundary conditions combined with the corresponding differential equations serve to determine the spatial distribution in the interior.

It is logical to begin with the differential equation for commodity prices. This equation is easy to derive from the gradient law $ph\,\phi/|\phi| = \mathrm{grad}\,p$ by taking squares of both sides. As $\phi/|\phi|$ is a unit vector, its square equals (scalar) unity, and we get

$$(\mathrm{grad}\,p)^2/p^2 = h^2 \tag{43}$$

As $\mathrm{grad}\,p/p = \mathrm{grad}\,\ln p$, we can rewrite this equation in the form

$$(\partial \ln p/\partial x_1)^2 + (\partial \ln p/\partial x_2)^2 = h(x_1, x_2)^2 \tag{44}$$

where we have simply expressed the square of the gradient in terms of partial derivatives. As h is a known function of the location coordinates, we obviously have a simple differential equation for the natural logarithm of commodity price.

4.5.2 Huygens' Principle

Equations of this type are well known from optics. Our flow lines (the integral curves of the direction field $\phi/|\phi|$) correspond to light rays, our constant price contours to wave fronts, and our transportation cost (in commodity terms) to the refractive index. We thus have complete correspondence and can apply all that is already known about this widely studied differential equation.

Under fairly reasonable boundary conditions the equation always has a unique solution. The easiest way to explain this is by adopting the simple but ingenious method, due to Huygens, for constructing new wave fronts from known ones. We put one point of a compass on any point on the known price contour and draw a circle with a radius equal to the reciprocal of the transportation cost h. The radius obviously equals the distance commodities can be moved by using up one unit of them (as the cost is itself given in commodity terms). This procedure is repeated for as many nearby points on the known contour as necessary. The radius of the compass is continuously adjusted to the reciprocals of transportation costs at the centers of the circles, until there are enough circles to draw an envelope. This envelope for the family of circles is then the new price contour. As long as the price corresponding to the old contour is known, we can label the new one by its price. The price increases by the price of the commodity itself. This is because, by construction, the new price contour represents the farthest locations that can be reached from the old contour by using up the value of one commodity unit. As prices increase by transportation costs equal to the prices themselves, the rate of increase is exponential. This is also revealed by the fact that our differential equation is written in terms of the logarithm of prices.

This method may be used to construct any number of new price contours and thus to establish the whole price distribution in the interior. Of course, the procedure is approximate; its accuracy increases with the number of intermediate contours we construct by considering transportation that uses up not whole units, but smaller and smaller fractions of commodity units. It is intuitively understood that by going to the limit, the exact solution to the differential equation may, in principle, be obtained. This solution should exist and be unique with respect to the method of construction.

Once we have established a map of constant price contours, we can also determine a unique pattern of orthogonal trajectories. As goods are shipped in the directions of price gradients, these orthogonal trajectories are the flow lines. Hence the commodity price distribution and the flow directions over the whole region can be determined simultaneously.

4.5.3 Boundary Conditions

Let us now consider the boundary conditions in somewhat more detail. It is unlikely that commodity price would be constant on the boundary, and thus we cannot start from the boundary as a known constant price contour. In general prices may be expected to vary on the boundary. Let us assume, however, that the flow lines and price contours in the exterior are established in accordance with a transportation cost function $h(x_1, x_2)$ extended to the exterior, but merging smoothly with the function defined for the interior. The situation is then exactly the same as before and we can start from any price contour in the exterior that touches the boundary somewhere. This is the simplest assumption we can use and is by no means unrealistic. But such restrictive assumptions are not necessary. After all, the passage of light from one medium to another (regardless of the shapes of the boundaries and the wave fronts) is a well-understood phenomenon. Similarly, we could introduce any discontinuities we wish and possibly obtain interesting refraction phenomena for trade, while retaining the desirable features of existence and uniqueness.

4.5.4 Equation for Flow Volume

As the price and flow directions yield the most important information concerning regional structure, the less interesting differential equation for flow volumes can be treated more briefly. We know that the firms decide on capital and labor inputs and outputs as soon as prices and wages are given. These are also used to determine the local land rents. But these land rents are transferred by the ownership mapping to the households as nonlabor incomes. The households also decide everything – family size, working time, leisure, and consumption – once they know the price, wage rate, and nonlabor income. Due to the mapping from land rents (dependent on prices and wages) to nonlabor incomes, we conclude that all endogenous variables of the firms (k, l, q, and the resulting g) and all the endogenous variables of the households (l', l, and q') are determined by p and w. (The long-term stocks of capital $K = k$ and labor $L = l + l'$, of course, are then given as well.)

We see that these decisions by firms and households are independent, except for one aspect, i.e. that the l determined by the firm and the l' determined by the household must match. Our present position is that prices have been determined, but wages still represent a degree of freedom. It is natural to assume that wages are adjusted locally to reach a balance between labor supplied and labor demanded. It may then be concluded that all the variables dependent on the decisions made by firms and households depend only on commodity price if it is assumed that wages adjust by local labor market equilibria. But, as the prices are known after solving the corresponding differential equation, local excess supply (computed without regard to the commodities used up in transportation) is a known function of the space coordinates. Therefore we can define

$$q - q' - \kappa k = z(x_1, x_2) \tag{45}$$

Using the divergence equation we obtain

$$\operatorname{div} \phi + h(x_1, x_2)\, |\phi| = z(x_1, x_2) \tag{46}$$

which is easily converted into a partial differential equation for flow volume $|\phi|$.

4.5.5 Charpit's Method of Solution

In order to recognize the character of this differential equation, let us introduce some new symbols (which are used only temporarily in this context and may have other interpretations elsewhere in the text). First, we define the unit vector field

$$(P(x_1,x_2),Q(x_1,x_2)) = \phi/|\phi| \tag{47}$$

and note that it is known because

$$\phi/|\phi| = \text{grad}\ln p/|\text{grad}\ln p| \tag{48}$$

is known, once we have solved for the price distribution in the region. Next, we use the identity

$$\phi = |\phi|\frac{\phi}{|\phi|} = |\phi|(P,Q) \tag{49}$$

to obtain

$$\text{div}\,\phi = \text{grad}\,|\phi|(P,Q) + |\phi|\,\text{div}(P,Q) \tag{50}$$

Then, defining

$$R(x_1,x_2,|\phi|) = z-(h+\text{div}\,(P,Q))\,|\phi| \tag{51}$$

we can write the differential equation as

$$P\frac{\partial|\phi|}{\partial x_1} + Q\frac{\partial|\phi|}{\partial x_2} = R \tag{52}$$

We note that P and Q are known functions of the location coordinates and that R, in addition to its dependence on location coordinates, also depends on flow volume $|\phi|$. This last dependence is linear. Differential equations of this type can be solved using the well-known Charpit method, whereby the partial differential equation is replaced by two ordinary differential equations

$$\frac{\mathrm{d}|\phi|}{\mathrm{d}x_1} = \frac{R(x_1,x_2,|\phi|)}{P(x_1,x_2)} \tag{53}$$

and

$$\frac{\mathrm{d}|\phi|}{\mathrm{d}x_2} = \frac{R(x_1,x_2,|\phi|)}{Q(x_1,x_2)} \tag{54}$$

Once we have solved these, and found two parametric families of solution curves in $x_1, x_2, |\phi|$-space, we can easily construct the solution surface for $|\phi|$. This means that existence and uniqueness theorems for ordinary differential equations instead of the more tricky ones for partial differential equations apply.

4.6 THE SPATIAL STRUCTURE

4.6.1 Introduction

Having discussed determination of the trade flow volume and equilibrium in the model as a whole, we return to the basic differential equation for determining prices. The solution of this equation provides the essential information about regional structure, by determining the flow lines of commodity trade along with the price contours, which are important determinants of production technology and scale of production.

If the transportation cost function $h(x_1, x_2)$ and the price distribution p on the boundary ∂A are chosen appropriately, then the pattern of spatial price contours and orthogonal flow trajectories could have any character we might wish to represent. Thus the model is very rich in potential results, although it contains very little substantial information. We could represent spatial organization of, for example, the Christaller–Lösch type, but such regular geometrical cases would be merely illustrative.

This situation is by no means peculiar to spatial applications of economic theory, even though there is a greater desire to obtain more definite knowledge about visual two-dimensional patterns than about the equilibrium price constellations in the punctiform markets of general economic theory. In the theory of general (or partial) economic equilibrium, there is also a considerable lack of precise information, which would seem strange in other sciences such as physics, where formal mathematical language is used.

In order to remedy this situation, Samuelson (1947) introduced the ingenious "correspondence principle." By merely assuming the dynamic stability of equilibrium, he was able to supply qualitative information concerning the effects of various kinds of disturbances on the equilibrium prices. The equilibrium model immediately gained a great deal in information content. The premise was indeed nonrestrictive, since any equilibrium analysis would, in fact, be pointless unless the equilibrium were stable. As all kinds of disturbances are persistent in any real economy, they must be allowed for in model representations. If the equilibria were unstable, any such disturbance would either set the system in perpetual motion or move it to another equilibrium state that has a basin of attraction. Such a state would deserve closer study as its basin of attraction, or stability, would make it likely to be maintained for some time. The original equilibrium state would not deserve any attention because it could never survive in a changing world.

4.6.2 Structural Stability

We can use exactly the same philosophy in our model of a spatially extended economy. A prerequisite for such treatment is the concept of structural stability, which will have to be explained in detail. Intuitively, it means that the flow pattern and the map of constant price contours look qualitatively the same as before when small disturbances occur. The expression "qualitatively the same" implies that flow trajectories (or price contours) are deformed and displaced slightly, as are singularities (stagnation points of the flow and critical points of the price surface), which, moreover, retain their character (whether of sink, source, saddle point, etc.). In fact we could imagine the picture of constant price contours and their orthogonal trajectories as drawn on a perfectly elastic rubber sheet. The class of deformed pictures obtained by stretching this rubber sheet would then represent the class of qualitatively equivalent flows (and price contours).

The price–flow pattern would then be considered structurally stable if small changes (perturbations) in the differential equation left its qualitative features unchanged, i.e. if after a small perturbation it still belonged to the same equivalence class as described above. We should now specify the meaning of a small change or perturbation. This may be accomplished by defining a normed space of differential equations, although the idea should be intuitively clear. If the pattern of flow and price contours is deformed only slightly when the differential equation whose solution determines this pattern is changed slightly, then the pattern is referred to as structurally stable. If small changes in the differential equation (or its boundary conditions) lead to drastic changes, such as splitting or fusion of singularities, emergence or disappearance of singularities, reversals of flow directions, etc., then the pattern of flow and price geography would be regarded as structurally unstable. In these cases the new patterns could not be obtained by stretching the rubber sheet. We would have to cut it and glue new sides together to arrive at the new picture.

The idea we would now like to suggest is that a spatial price structure and a corresponding system of orthogonal flow trajectories would not in themselves deserve any attention, if they were not structurally stable. Constant change is again assumed in the economy we wish to model; perturbations of the (spatially) dynamic system are thus also admitted.

4.6.3 Topological Characterization

This poses the question of whether we can obtain (as Samuelson did) any qualitative information from this reasonable assumption of structural stability. Otherwise, not much is gained by introducing it. Fortunately there is a characterization theorem on structurally stable flows in the plane, which has surprisingly rich features. It indicates that a stable flow has only a finite number of isolated singular points of very few specified categories, and is laminar everywhere else. There are also global results on how the few singularities may be connected. This makes it possible to draw a very precise picture of the structurally stable flow and the spatial organization of the economy corresponding to such a flow. The characterization is again topological. Along with our basic picture we have to consider all the deformations obtained by the class of stretchings when the picture is drawn on an elastic rubber sheet.

The theorem which gives rise to these result was conjectured by Andronov and Pontryagin in 1937. Later on, it was rigorously proved and further developed in the work of Morse (1934), Smale (1967), and Peixoto (1973, 1977); for recent treatments see Peixoto (1977) or Hirsch and Smale (1974).

Before turning to the formalities, two aspects should be emphasized. *First*, we have actually obtained visual information concerning the pattern of flow and geographical organization of the economy that we regard as desirable. The information is topological, in contrast to the classical geometric information contained in the work of, for example, Christaller and Lösch. Topology, however, tells us more than might be imagined at first sight. For instance, it definitely contradicts the Christaller–Lösch paradigm of spatial organization, as their hexagonal pattern cannot be transformed into the structurally stable flow characterized below by any topological transformation. Moreover, all traditional market-area theories, from Launhardt and Weber onwards, are contradicted, as they imply boundaries of market areas at which trade stagnates. As already indicated, the accumulation of singular points is ruled out in structurally stable flows. Readers used to the Christaller–Lösch paradigm may find this difficult to accept. But it should be kept in mind that all traditional market-area theories (including that of Christaller and Lösch) actually assume constant transportation cost and consequently linear routes. Such models are thus linear in character, and it is well known that linear models can never display any phenomena of structural instability. Now, it is often pointed out that space is not homogeneous and uniformly traversible, as assumed in classical location theory. But if we believe this then we have to admit the nonlinearities and draw the proper conclusions. Since structural stability is not automatically guaranteed in nonlinear systems, but must be expressly assumed, and since it then rules out certain basic features of classical location theory, we should not try to fit the Christaller–Lösch model to nonhomogeneous space. Rather, the features that contradict structural stability should be discarded.

Second, classical location theorists appear to have exploited almost everything in the realm of Euclidean geometry with its regular shapes. If we wish to introduce a new, more general, theory of economic organization in continuous two-dimensional space (based on nonlinearities), we cannot expect a precise characterization in terms of regular shapes from Euclidean geometry. A topological characterization is the most we can hope to arrive at and is, in fact, the only way to regenerate this old field of research.

4.6.4 The System of Flow Lines

The dynamic system under discussion is described by the equation

$$h \frac{\phi}{|\phi|} = \operatorname{grad} \ln p \tag{55}$$

which can be written

$$\frac{\phi}{|\phi|} = \frac{\operatorname{grad} \ln p}{|\operatorname{grad} \ln p|} \tag{56}$$

This is actually a poor formulation for solving the equation. But since we do not intend to solve it, but rather to discuss the stability properties of the set of integral curves corresponding to the directions $\phi/|\phi|$, there is no harm in assuming p as known. In this way we can incorporate the changes in the system caused by changed price distributions on the boundary as perturbations, along with those caused by changes in the function $h(x_1, x_2)$.

As the flow is in the direction of the gradient of p, we can find some parameterization in which the equations for the flow lines can be written as

$$\frac{dx_i}{ds} = \frac{\partial p}{\partial x_i} = F^i(x_1, x_2) \qquad i = 1, 2 \tag{57}$$

This is a system of two ordinary differential equations in two variables. Normally, the solutions to such a system are well classified only if the right-hand side is linear in x_1, x_2. The singular points, where $F^i(x_1, x_2) = 0$ for $i = 1, 2$, are then known to be nodes, ordinary saddles, spirals, and centers. Except for the singularities, the flow lines are regular, i.e. topologically equivalent to parallel straight lines. This can be learned from any textbook on ordinary differential equations. Furthermore, if the equations are nonlinear, then they can behave in almost any way, with lines or areas of accumulating singularities, or singularities of a complicated composition of any number of elliptic and hyperbolic sectors.

Fortunately, the assumption of structural stability removes these complications and makes the solutions behave as if they arose for linear differential equations.

4.6.5 Perturbations

Along with the preceding system of differential equations, let us consider the following

$$\frac{\mathrm{d}x_i}{\mathrm{d}s} = G^i(x_1, x_2) \qquad i = 1,2 \tag{58}$$

which differs only slightly from it. This slight difference is defined by the conditions

$$|G^i - F^i| < \epsilon \qquad i = 1,2 \tag{59}$$

and

$$|\partial G^i/\partial x_j - \partial F^i/\partial x_j| < \epsilon \qquad i,j = 1,2 \tag{60}$$

The system G^i is called an ϵ-perturbation of the system F^i when the above inequalities are fulfilled. We have now introduced a C^1-topology in the space of differential equations by the metric ϵ. This C^1-topology is usually considered necessary because if the inequalities were only required to hold for the right-hand sides of the differential equations, but not for their derivatives, then any isolated singularity could be changed into a line of singularities (and *vice versa*) by an ϵ-perturbation. In other words, a C^0-topology would not provide structural stability to coincide with the intuitive concept suggested above. On the other hand, the requirement on derivatives higher than the first order makes the class of perturbations unnecessarily narrow (cf. Peixoto 1977).

Let us now consider the flow portrait for the F-system, along with the flow portrait for the slightly disturbed G-system. If there is a one-to-one continuous mapping between these solution spaces, such that trajectories are mapped on trajectories (of preserved direction), and singularities are mapped on singularities (of the same type), then the system is referred to as structurally stable. Of course, this must hold for any perturbed system that fulfills the inequalities. If there is no such homeomorphism the system is called structurally unstable.

It has been demonstrated that, in the space of differential equations, the subset of structurally stable ones is dense and open. So, the mathematical definition of structural stability really does capture what we have in mind since we are assured that infinitesimally close to any unstable differential equation there is a stable one, whereas the reverse does not hold.

4.6.6 A Characterization Theorem

The consequences of structural stability that have been demonstrated can be summarized in three points:

1. The flow is regular everywhere, except at a finite number of isolated singular points; regularity here means that only one trajectory passes through each point. Equivalently, we can say that it is topologically equivalent to a set of parallel straight lines.
2. The singular points are hyperbolic, i.e. they are nodes, saddles, or spirals. As we are dealing with gradient flows, it can easily be shown that the eigenvalues are real, and so spirals are ruled out. We are left with nodes (sources and sinks) and ordinary saddles.
3. Finally, there is the global result that no trajectory joins saddle points. At each saddle there are four incident trajectories, one pair directed inwards, and one pair outwards. No outward trajectory can thus be incident to another saddle, nor can it return after a loop to the same one in the inward direction.

4.6.7 Sinks and Sources

Let us now consider the interpretation of these conditions in terms of the organization of the spatial economy. To begin with the organization around a singular point, a source is a point from which all trajectories in a surrounding basin of repulsion diverge. They obviously form a set of radiating trajectories that are orthogonal to (more or less) circular, concentric price contours. As we have noted, commodity prices are important determinants of the scale and technology of production chosen. The economic organization, accordingly, is one of concentric rings of various activities, and the routes of transportation are radial (see Figure 4.1). The picture is thus very similar to the von Thünen case (or the case beloved of the proponents of the New Urban Economics).

The same is true for the case of the sink. There, all trajectories in a surrounding basin of attraction converge to the singular point, the price contours are again concentric closed curves, and the spatial organization is in terms of concentric rings. This case is similar to the previous one, the only difference being a reversal of flow directions. If we wish to interpret the cases of sources and sinks more specifically in terms of urban geography, we can say that the sources are productive, and the sinks consumptive, centers. We must, however, remember that our model is competitive. The consumption centers do not play the "monopsonistic" role of the center of the von Thünen model, or that of a central business district (CBD). Nor are the productive centers to be compared with the monopolistic firms of classical market area theory.

In our model production and consumption are activities going on everywhere, but at various rates, and nobody is in the position of a monopolist or a monopsonist. Sometimes,

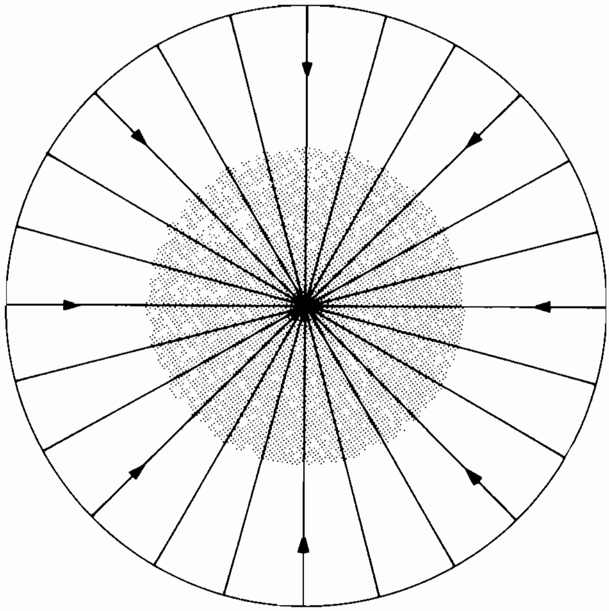

Figure 4.1. Flow and spatial organization around a node singularity.

it has been argued that spatial economy is intrinsically associated with monopoly and price discrimination. It is, of course, true that spatial (like temporal) distance is a perfect means of segregation of a market according to various demand elasticities, but there is no necessity to introduce monopoly merely because we are dealing with a spatial economy. A spatial competitive economy is as well defined as the case treated in spaceless economics, and as realistic or unrealistic as the latter. The only difference is that competition in a spatial economy does not make prices equal; it only confines the price differences to the intervals defined by transportation costs.

4.6.8 Saddle Points

Now that we have understood the character of the nodes (sources and sinks), let us discuss the remaining types of singularities, the ordinary saddles. For each ordinary saddle there are two pairs of incident trajectories, one ingoing and the other outgoing. Therefore this case is very different from the cases of sources and sinks, where an infinity of trajectories in a surrounding basin were incident. At a saddle, the surrounding space is separated into four sectors, each containing hyperbolic trajectories, attracted toward the singularity but missing it. The set of price contours, to which these hyperbolic trajectories are orthogonal, is itself a set of hyperbolas but rotated by an angle of about 45°. The various zones of economic activity are now contained between pairs of hyperbolas, in opposite sectors, and the organization of space is sectoral (rather than ring shaped) around a saddle singularity (see Figure 4.2).

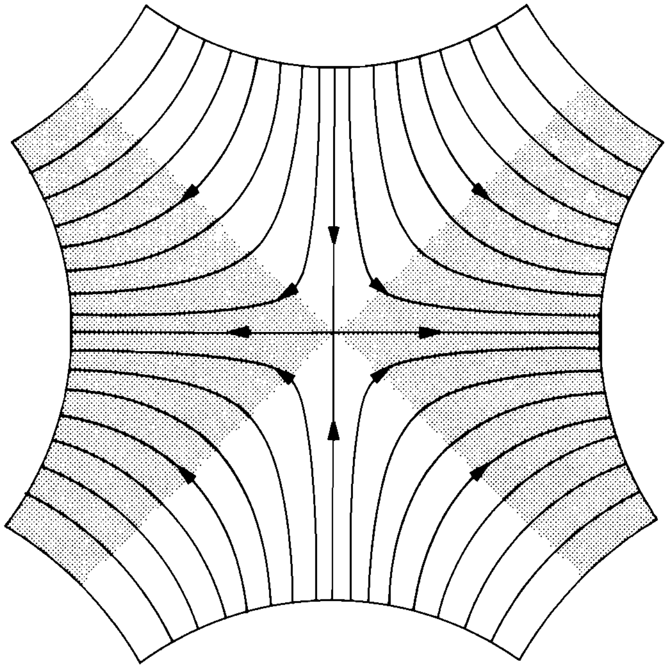

Figure 4.2. Flow and spatial organization around a saddle singularity.

Sectoral organization of space, where some activities are confined to North and South and other ones to East and West, is not completely unrealistic in urban geography, even if it does not appear so frequently in textbooks as the case of concentric rings. If we require a more precise interpretation of the saddle singularity, we should note the fact that the trajectories are curved towards it. Since all the routes of transportation are thus deflected from the straight line, transportation must be very favorable in the neighborhood of a saddle singularity. Recalling our former interpretation of sources and sinks as productive and consumptive centers, we could say that the saddle points are far from both types of centers and are, in fact, "condensation nuclei" of empty space with particularly favorable conditions for transportation.

As we have now interpreted the three types of singularities admitted in structurally stable flows, we recall that there are infinitely many of them, and that the flow is laminar everywhere else in space. Later on we will see how the principle that rules out saddle connections can be used to characterize the whole global picture of commodity flow and spatial organization.

4.6.9 Transversality

Before continuing, let us consider structural stability at singular points from a slightly different point of view, which will prove fruitful later when we are dealing with the subject of structural change. The trajectories we are dealing with can be obtained as a gradient field to a potential function (commodity price). Let us therefore consider a general potential function

$$\lambda = \lambda(x_1, x_2) \tag{61}$$

The singular points of its gradient flow field are defined by

$$\frac{\partial \lambda}{\partial x_1} = \frac{\partial \lambda}{\partial x_2} = 0 \tag{62}$$

Let us suppose that we are studying just one singularity in x_1, x_2, λ-space and that it is located at the origin, i.e. $x_1 = x_2 = \lambda = 0$ at the singularity. We can always attain this by a coordinate change, which is a simple translation, so that this assumption in no way affects the generality of the following discussion.

The question we now pose is the following: how likely is it that the Hessian of the potential is zero at the singular point we are studying? (In passing we note that the Hessian is the Jacobian of the system of differential equations defining the flow lines.) The formal condition for a vanishing Hessian is

$$\frac{\partial^2 \lambda}{\partial x_1^2} \frac{\partial^2 \lambda}{\partial x_2^2} = \left(\frac{\partial^2 \lambda}{\partial x_1 \, \partial x_2}\right)^2 \tag{63}$$

To answer the question we consider the combination of values of the three second-order partial derivatives as a point in three-dimensional space and observe that the condition for a vanishing Hessian defines a surface (a double cone) in this space. Our question is now transformed into the following: how likely is it that a point lies on a surface in three-dimensional space? One answer to this is given by the principle of transversality.

Consider pairs of various linear subspaces of ordinary Euclidean three-space. Two planes through the origin would be likely to intersect along a line, whereas one line and one plane through the origin are likely to intersect in a point. Anything else would have a vanishingly small probability of occurring. The likely modes of intersection are called transverse. We note that, for transverse intersections, the sum of the dimensions of the intersecting subspaces equals the sum of the dimensions of the intersection manifold and that of the surrounding space. (For the case of two planes $2 + 2 = 1 + 3$, for the case of a line and a plane $1 + 2 = 0 + 3$.) This principle is readily generalized to the much more interesting cases of affine subspaces, and of manifolds in general. In the latter case we transfer the dimension conditions to tangent planes and tangent lines to the surfaces and curves considered.

We are now prepared to answer the question. As the double cone defined by zero Hessian is a surface, we would need at least a curve for a transverse intersection. Of course, a necessary condition for transversality is that the sum of the dimensions of the intersecting manifolds at least adds up to the dimension of the surrounding space. Otherwise, the dimension condition stated can never be met. As our value combination is only a point (i.e. has zero dimension) it cannot meet the double cone (with dimension two) transversely in a surrounding three-dimensional space. Therefore, we conclude that a zero Hessian is ruled out by the principle of transversality.

In passing we should note that, if we were dealing not with one value combination of the second-order partial derivatives but with their development over time, then we would have to consider not a point but a curve, parameterized by time. Such a curve could meet the double cone transversely, and so we note that the Hessian could be zero at isolated moments of time. We could also consider a system dependent on several (mutually independent) parameters. In that case, the parameterized curve could become a surface, or even fill up the whole volume of the surrounding space, when we consider two, or three, such parameters. It should be noted that only in the latter case (with three parameters and a space-filling manifold) would it be likely that the apex of the double cone would be met. This is interesting because a "monkey saddle" flow (with six hyperbolic sectors), which will be seen to occur if we try to organize space by a hexagonal tiling (as do Christaller and Lösch), is a manifestly pathological case in view of transversality.

We will return to the concept of transversality when we deal with structural change. At present it is sufficient to note that transversality and structural stability are closely related concepts, and that transversality assures us that the Hessian is nonzero at our singular point.

4.6.10 Morse's Lemma

Now, if the Hessian is nonzero then we know that the potential surface can be changed into a so-called Morse saddle by a smooth change of coordinates. We can thus define a smooth mapping:

$$\xi_1 = \xi_1(x_1, x_2) \tag{64}$$

and

$$\xi_2 = \xi_2(x_1, x_2) \tag{65}$$

such that

$$\lambda = \pm \xi_1^2 \pm \xi_2^2 \tag{66}$$

in the neighborhood of the singular point. This is proved by Morse's lemma. A simple proof for the lemma is provided in Poston and Stewart (1978). We see that the potential surface in the new coordinates is either a circular paraboloid (turned right or upside down) or a hyperbolic paraboloid (a saddle). Thus, the results of transversality and of structural stability are the same. The difference is that transversality deals with the qualitative features of the potential surface close to critical points, whereas the results on structural stability (from the generic theory of differential equations) deal with the flows (in our case, gradient to a potential, but not necessarily so in general).

We can now use the standard cases of Morse saddle potentials to illustrate the structurally stable flows and price contours in the neighborhood of a singular point. Let us put

$$p = [\exp(\pm x_1^2 \pm x_2^2)]^{1/2} \tag{67}$$

Accordingly, we have

$$\text{grad} \ln p = (\pm x_1, \pm x_2) \tag{68}$$

and

$$|\text{grad} \ln p| = (x_1^2 + x_2^2)^{1/2} \tag{69}$$

Thus, if we assume

$$h(x_1, x_2) = (x_1^2 + x_2^2)^{1/2} \tag{70}$$

and

$$\phi/|\phi| = (\pm x_1, \pm x_2)/(x_1^2 + x_2^2)^{1/2} \tag{71}$$

then we clearly have the gradient condition, $h\phi/|\phi| = \text{grad} \ln p$, and the differential equation, $(\text{grad} \ln p)^2 = h^2$, reduced to identities.

Two positive signs refer to the case of a source, two negative signs to that of a sink, while one positive and one negative sign means we are dealing with a saddle. In fact, we can illustrate all the structurally stable flows around singular points by this single case. We refer back to the graphical illustrations in Figures 4.1 and 4.2. It should be noted that all these cases result from one single transportation cost function, where the local cost of transport is proportional to the distance from the singular point (at the origin). Transportation is favorable close to the singular points, and gets progressively less favorable with distance from these points. We have already noticed this in connection with the saddle singularity. Obviously, the situation is the same at nodes in our illustration.

The fact that different flow patterns result from the same differential equation

$$(\text{grad} \ln p)^2 = (x_1^2 + x_2^2) \tag{72}$$

should not be taken as a contradiction of our conclusion above that flow patterns and price potentials are unique. Of course, the solutions obtained for various sign combinations in our example are associated with very different boundary conditions. It is well known that the solutions of partial differential equations (unlike those of ordinary differential equations), not only leave certain parameters to be determined by the boundary conditions, but admit very different shapes of solution surfaces, depending on the exact character of the boundary conditions.

4.6.11 A Global Picture of Stable Flow

It is now time to turn to the question of how to construct our global picture of flows and price contours on the principle that no trajectories join saddle points. As trajectories in only four directions at a saddle are actually incident, this means that if we follow any one of these we finally either cross the boundary or end up at a singularity that must be a node. We note that two of the nodes must be sources and two of them sinks.

Let us then try to organize the flow in terms of the trajectories incident on the saddle points. As there are two pairs of such directions at each saddle, we should be able to organize the trajectories incident to saddles as a square grid. The singularities at the points of intersection can now be identified. Starting from any saddle point (and knowing that the incident trajectories must end up at nodes), we conclude that there must be a pair of sinks in the outgoing directions, and a pair of sources in the ingoing directions. We have thus determined the character of the nodes East, West, North, and South of the saddle point. However, we can do more than that. Consider the four points NE, SE, SW, and NW of the original saddle point. Each of these four singularities is associated with ingoing as well as outgoing trajectories. This leaves no other possibility than that these four singularities are again saddle points. And so we can continue, starting our discussion anew at each of the new saddle points, until we have oriented the whole square grid and determined the characters of each singularity. Thus we see that each saddle is connected to four nodes, and each node to four saddles.

Once we have this basic graph, it is easy to fill in the whole families of trajectories, not just the skeletons of those incident to saddle points, and to draw a set of orthogonal price contours. We recognize in Figure 4.3 the local pictures of spatial organization from Figures 4.1 and 4.2 around each singularity. But we also see that the circular price contours become squarer, the farther we move from the nodes. Returning to our identification of sources and sinks as centers of productive and consumptive character, respectively, we discover that the basic subdivision of space, in areas of different characters, is quadratic in shape. These basic squares each have a node as a center and four neighboring saddles as corners. They alternate in space (as regards their basically productive or consumptive character), and thus yield a chessboard pattern.

If we wish, we can also identify "market areas," by considering a different quadratic subdivision. Take any saddle point as a center, and four surrounding nodes as corners in a new subdivision of space. The result is that the original chessboard (defined by the reverse procedure) is slightly translated in space. The squares in this new subdivision all have four flow trajectories as sides. We note that no trajectory crosses such a square, so that trade is completely confined within each square of this kind. This comes pretty close to the traditional concept of isolated market areas. The only difference is that trade does not stagnate along the boundary as is the case with the traditional linear models. Trade follows the boundaries of our "market areas," but they are still completely self-contained with respect to trade. This, along with the fact that our centers have no exclusive character as sites of isolated firms (or "central business districts"), makes the difference.

Later on we will consider the apparent contradictions between our approach and traditional location theory more closely. Meanwhile, we conclude the present discussion by stressing that in the basic picture we can, of course, leave out any number of singularities, but that we must admit all the topological transformations of the basic pattern. This means that our picture, again, must be imagined as drawn on a perfectly elastic rubber sheet. It thus represents all the deformations that can be obtained by "stretching without tearing." Nevertheless, the qualitative characterization is surprisingly precise. The "correspondence principle" used here thus yields results at least as rich as in its original context.

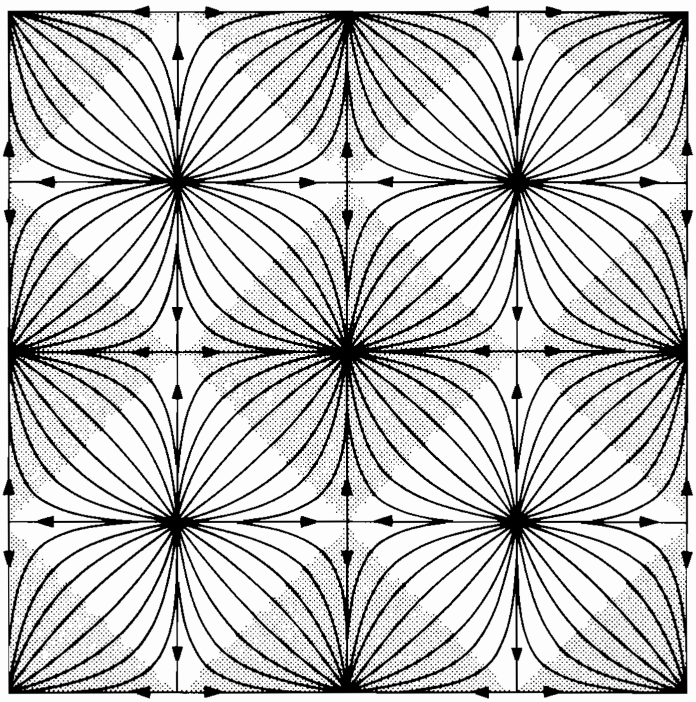

Figure 4.3. Global picture of a structurally stable flow.

4.7 CONTRADICTIONS WITH CLASSICAL LOCATION THEORY

Let us now return to the apparent contradictions between our topologically characterized spatial organization, and that of classical location theory. There are two major contradictions. *First*, as structural stability rules out the accumulation of singular points, there are no boundaries of market areas in the sense that trade stagnates along these boundaries. *Second*, as triangular or hexagonal flow patterns necessarily include monkey saddle singularities that are manifestly unstable, the Christaller–Lösch paradigm on tiling space with hexagons is ruled out.

4.7.1 Linearity of Classical Models

We must clearly consider these matters in some detail, since the ideas contradicted are deeply rooted in regional science. Consider the case of market areas as once defined by Launhardt and Weber. There are two monopolistic firms, located at some distance from each other and sharing the same two-dimensional space as their market areas. These market areas are separated by a boundary, which, by the famous and elegant "Launhardt funnel" construction, is found to be an intersection of cones. The routes are linear, and there is no question about the stagnation of trade along the boundary. This concept of a market area plays a major role in all regional economics.

It should, however, be noted that the assumption of linearity (homogeneous space with straight trajectories) is basic. Suppose that roads are built to connect the major sites, defined by the locations of our monopolistic firms in an otherwise homogeneous plane. These particularly good transportation facilities that now arise along the axis connecting the two firms cause the routes to be curved towards this axis, thus destroying the linearity of the model. In a continuous representation it is a relatively short step to see that the intersection of the boundary and the connection line becomes a saddle point. The trajectories become hyperbolic, and some trade (which may be as weak as we wish) develops along the boundaries.

Thus the concept of isolated market areas, in the sense that trade is completely confined within them, remains, as we have already noted. The difference is that there is now some (possibly very weak) flow of trade along the boundary itself. If there were no such trade, the situation would become unstable, because the model is no longer linear, and the accumulation of singular points along the boundary would be completely destroyed by any perturbation of the flow, however weak it might be.

4.7.2 Hexagonal Tiling

Let us consider the second problem, namely the case of a hexagonal tiling of space. A flow organized according to this principle would display either three major directions (separated by angles of 120°), or six directions (separated by angles of 60°). This would also be true for the incident directions at singular points.

Taking first the situation with three directions, we see that saddles could not fit this case. An ordinary saddle point requires four sectors, and if we try to join three hyperbolic sectors, we immediately find that the flow even becomes impossible to orient. Therefore the singular points must be nodes. Assembling the picture, we discover that it is possible to arrange a hexagon of adjacent nodes, with sources and sinks alternating. This is shown in Figure 4.4. But if we try to fit in more trajectories there is obviously a singularity missing in the middle of the hexagon, and this can only be a monkey saddle, consisting of six hyperbolic sectors.

Let us next consider the case where there are six incident directions at each node. The basic graph of the flow is then a regular triangular network. In trying to orient this network, we can start from a source. We note that only three of the surrounding singularities can be sinks. The remaining three singularities are associated with ingoing as well as outgoing trajectories. If we continue in this way, always trying to avoid monkey saddles, we again identify three of the singularities surrounding each sink as sources (one of these being our starting point). Doing this for all three sinks adds to the picture six sources, which are located on the sides of the larger hexagon shown in Figure 4.5. However, in completing so much of the orientation, we have in fact drawn three monkey saddles.

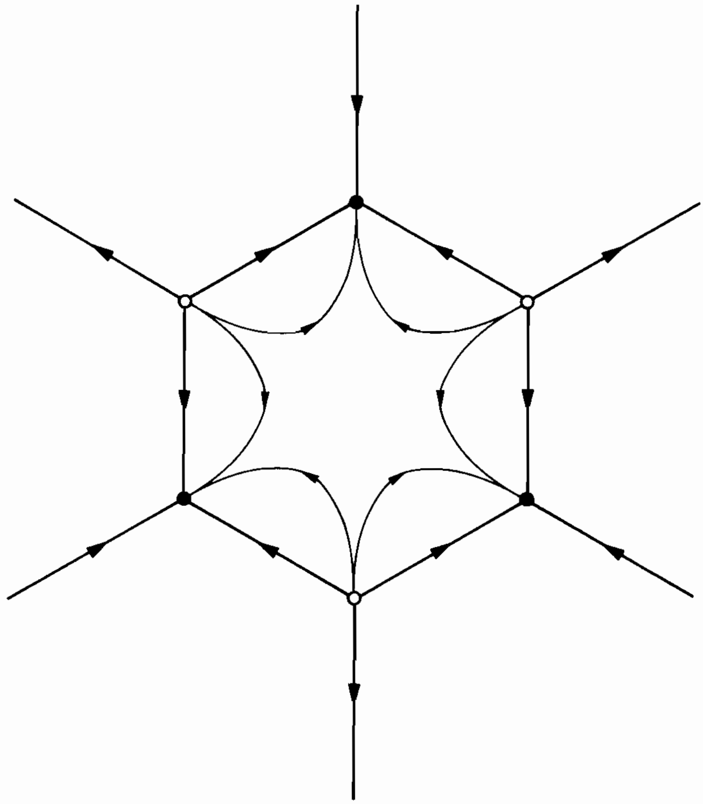

Figure 4.4. A missing monkey saddle.

Thus, our attempt to orient the graph so as to avoid monkey saddles actually leads to monkey saddles, and we conclude that they are closely related to any triangular or hexagonal type of flow. The resulting flow picture is shown in Figure 4.6.

As we have stated above, the monkey saddle case is highly structurally unstable, or, equivalently, extremely unlikely according to the principle of transversality. As the monkey saddle is one of the standard forms dealt with in catastrophe theory, we will review all the structural changes that it can undergo in Section 4.8, on structural change. But before this, we will consider accumulated singularities and monkey saddles from a more formal mathematical viewpoint. As we will see, the two are in fact quite closely related.

4.7.3 Formal Analysis

Recall our general potential function

$$\lambda = \lambda(x_1, x_2) \tag{73}$$

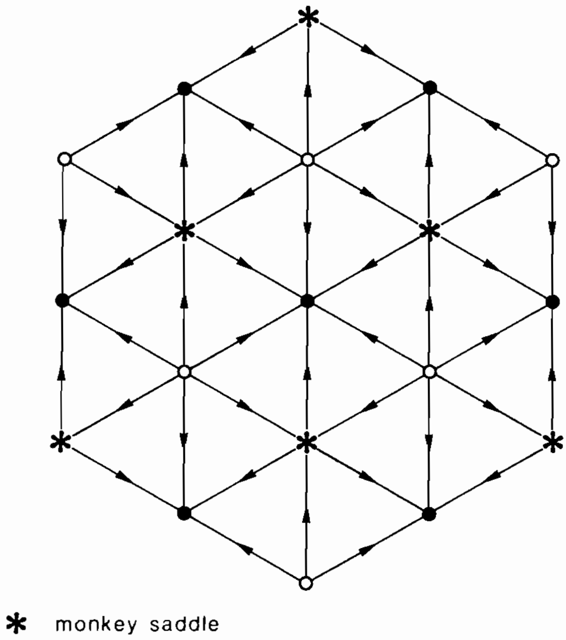

Figure 4.5. The occurrence of monkey saddles in a hexagonal lattice.

which, at a critical point, satisfies the conditions

$$\lambda_1 = \lambda_2 = 0 \tag{74}$$

where λ_1, λ_2 denote the partial derivatives $\partial\lambda/\partial x_1$ and $\partial\lambda/\partial x_2$. Now, suppose we have an accumulation of singularities. Then we can find a line, parameterized by s, along which these equalities hold identically. Differentiating, we get

$$\lambda_{11}\, dx_1/ds + \lambda_{12}\, dx_2/ds = 0 \tag{75}$$

and

$$\lambda_{21}\, dx_1/ds + \lambda_{22}\, dx_2/ds = 0 \tag{76}$$

From elementary linear algebra we know that these equations have a nontrivial solution only when

$$\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} = 0 \tag{77}$$

i.e. when the Hessian is zero. A nonzero Hessian certainly rules out the accumulation of singularities.

Next, let us consider the case of the monkey saddle. The monky saddle point is a critical point of the potential. The potential surface has a horizontal tangent plane, and intersects it along lines pointing in three different directions. These lines divide the

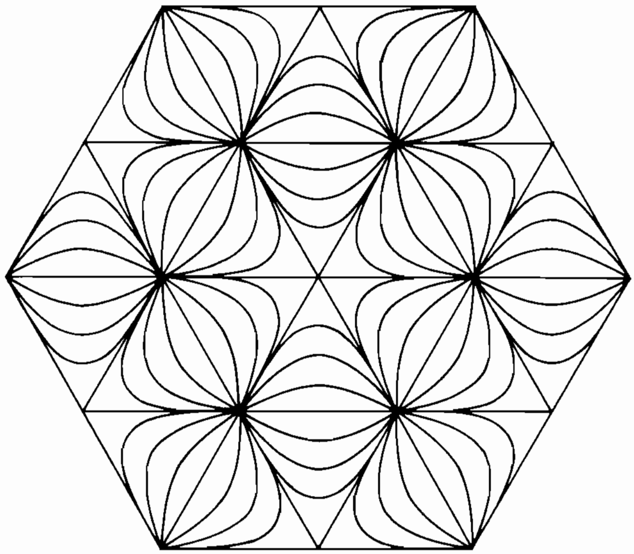

Figure 4.6. Flows associated with hexagonal or triangular subdivisions of space.

tangent plane into six sectors, such that the potential surface alternately lies above and below it. Our conclusion is that there are three directions in which the surface cuts the tangent plane. The tangent plane being horizontal, there are thus three directions in which $\lambda(x_1, x_2) = \text{constant}$.

Differentiating this equation of constant potential twice we get

$$\lambda_{11}(dx_1)^2 + 2\lambda_{12}(dx_1)(dx_2) + \lambda_{22}(dx_2)^2 = 0 \tag{78}$$

Assume for the moment that λ_{22} is not zero. We can then divide through by λ_{22} and obtain a quadratic equation in dx_2/dx_1, which can be solved to yield

$$\frac{dx_2}{dx_1} = -\frac{\lambda_{12}}{\lambda_{22}} \pm \frac{(\lambda_{12}^2 - \lambda_{11}\lambda_{22})^{1/2}}{\lambda_{22}} \tag{79}$$

Obviously if the Hessian (which appears in the square root) is nonzero we get two different directions dx_2/dx_1 in which the potential surface is constant. If the Hessian is zero, then we only have one direction, provided that λ_{12} is nonzero. We note in passing that this fits the case of a line of singularities. If the tangent plane "cuts" the surface along one line only there must be a case of tangency.

Suppose instead that λ_{22} is zero but that λ_{11} is not. Then, again, we can proceed as before, obtaining

$$\frac{dx_1}{dx_2} = -\frac{\lambda_{12}}{\lambda_{11}} \pm \frac{(\lambda_{12}^2 - \lambda_{11}\lambda_{22})^{1/2}}{\lambda_{11}} \tag{80}$$

with the same conclusion. There is no particular significance in λ_{11} or λ_{22} being zero: this just means that one of the directions of constant potential is parallel to one of the axes. There is still nothing unduly alarming if both λ_{11} and λ_{22} are zero, but λ_{12} is not: then both directions are parallel to the axes.

Therefore we conclude that, if the Hessian is nonzero, we get two directions of constant potential. If the Hessian is zero, but not all the partial derivatives are zero, then we have a line of singularities. So, in order to have three directions of constant potential we conclude that all the partial derivatives must be zero. This means that, in the space of values of the three second-order partial derivatives that we discussed in connection with the concept of transversality, the apex of the double cone must be met. (See pp. 138–139.) We are therefore dealing with a manifestly unlikely and unstable phenomenon when we consider monkey saddles.

The reader may wonder about the implicit assumption above that the term in parentheses would be positive, unless zero. However, if it were negative then the critical point would be elliptic (and not hyperbolic) in the terminology of classical differential geometry. This would imply a quadratic form

$$\lambda_{11}(dx_1)^2 - 2\lambda_{12}(dx_1)(dx_2) + \lambda_{22}(dx_2)^2 \tag{81}$$

that is definite (positive or negative), and we would, in fact, be dealing with simple maxima or minima of the potential. The singularities would then be nodes, not saddles, and there would be no intersections at all between the potential and its tangent plane. This also shows up in the fact that the directions of "intersection" dx_2/dx_1 or dx_1/dx_2 become complex numbers. Therefore, we need not consider further the cases involving imaginary square roots.

Therefore we conclude that structural stability and transversality contradict both the classical market area concept and the Christaller–Lösch hexagonal organization of space. If we think that structural stability is a reasonable assumption, then we must make the corresponding change of paradigm. We should stress once again that our arguments say nothing about these models in their original *linear* contexts. A linear system can never be structurally unstable. Only when we wish to transfer these paradigms to the more general (and realistic) assumption of nonlinearity do the contradictions arise.

4.8 STRUCTURAL CHANGE

4.8.1 Introduction

So far we have employed the concept of structural stability as our main modeling instrument, both because it was a reasonably nonrestrictive assumption and because it yielded such rich results in terms of characterization. Structurally unstable patterns were disregarded because they would, at most, have a temporary existence in the course of evolution. Structures existing before and after such temporary transitions would have easily recognizable qualitative features and would only undergo smooth changes for most of the time.

However, even if the structurally unstable patterns do not interest us as such (because they exist, at most, momentarily), they have an indirect significance because the structures before and after the passage through instability are very different. If we could somehow classify such possible transitions we might learn something about the possible sudden changes of structure that can occur in the process of evolution.

Of course, any changes would be possible if we did not constrain the system in some way, and the idea of classification would become meaningless. However, we might begin by assuming that the system is dependent on a certain number of external parameters that can evolve independently of each other. The potential surfaces would then be replaced by parameterized families of potential surfaces. If we assume transversality (or structural stability), no longer for each potential but for families of potentials, then we move into the realm of catastrophe theory as developed by Rene Thom. This theory, which is one of the major intellectual achievements of our age, yields a clear classification of structural transitions. Again, the classification is topological. The canonical forms of the surface families, again, represent all their topological equivalents. We do not venture to give any intuitive interpretation of Thom's classification theorem. The reader is referred to Poston and Stewart (1978) for an "easy" introduction. Here we just record the main results.

4.8.2 Further Discussion

As we are dealing with two-dimensional space, the catastrophes of particular interest are the umbilics. In passing, it should be mentioned that although regional scientists have used a lot of catastrophe theory, they always deal with the one-dimensional catastrophes. Perhaps this is another result of the fact that continuous two-dimensional space is almost always forgotten in regional modeling.

Due to the so-called "splitting lemma," we can always dissociate one of the coordinates from structural change when the number of parameters does not exceed two. With three parameters (representing the causes of the perturbations we wish to consider), we get the simplest, truly two-dimensional catastrophes: the elliptic and hyperbolic umbilics. As these include the monkey saddle case, and since three parameters make it possible to study a wide range of exogenous change, we will limit discussion to these cases.

The canonical form (or the "universal unfolding") of the elliptic umbilic is

$$\lambda = x_1^3 - 3x_1x_2^2 + a(x_1^2 + x_2^2) + bx_1 + cx_2 \tag{82}$$

and that of the hyperbolic umbilic is

$$\lambda = x_1^3 + x_2^3 + 2ax_1x_2 + bx_1 + cx_2 \tag{83}$$

The three exogenous parameters are denoted a,b,c while x_1, x_2 are, as usual, the space coordinates. To depict the potential λ as a function of two coordinates and three parameters, we would need a space of six dimensions. For this reason we have divided each of the Figures 4.7 and 4.8 into upper and lower parts. The upper part is a three-dimensional representation of the bifurcation manifold in parameter space. Each time the value combination a,b,c crosses this manifold there is a sudden change of structure. The lower parts of the diagrams are various flow pictures, corresponding to various value

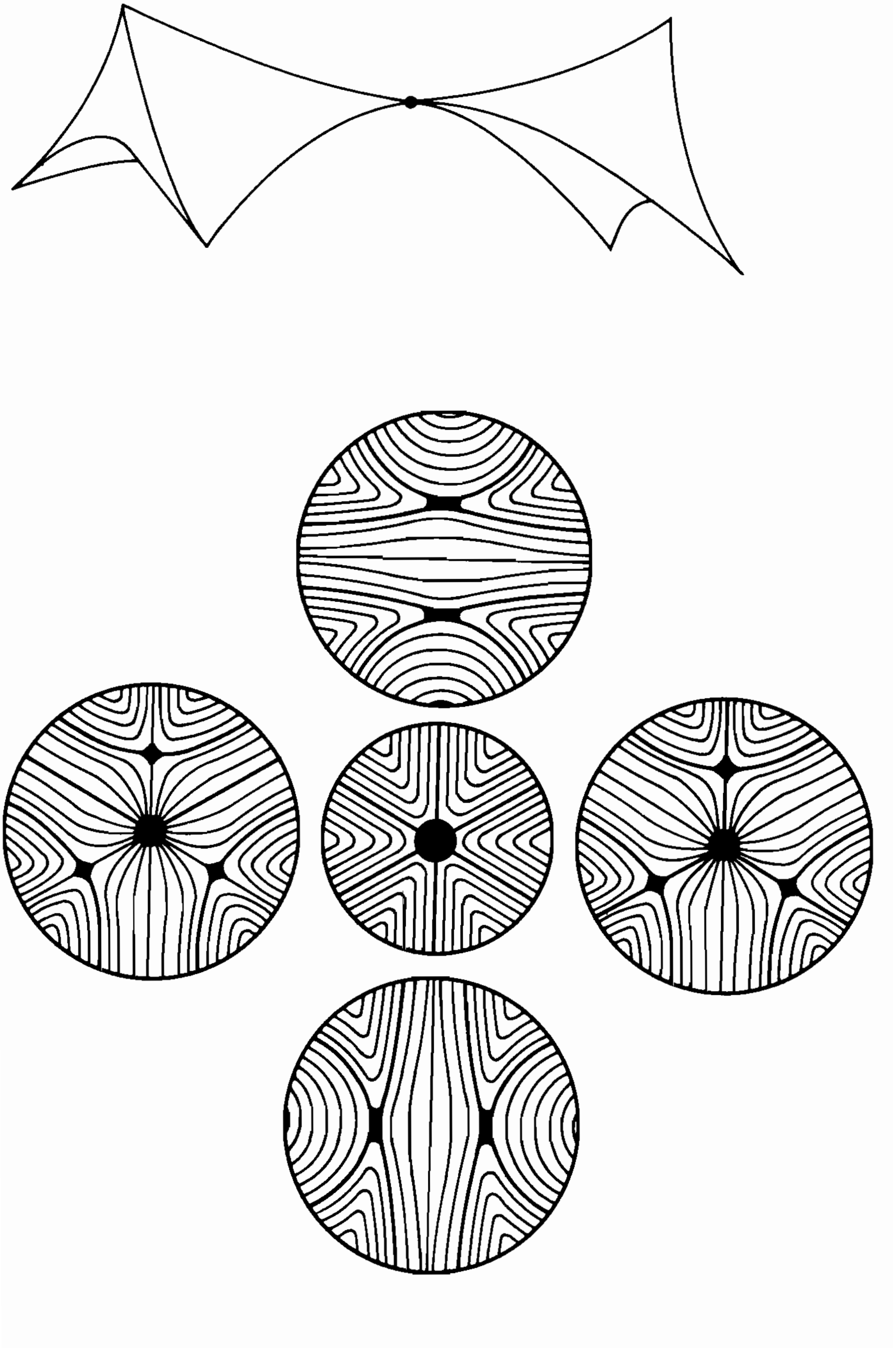

Figure 4.7. The elliptic umbilic case.

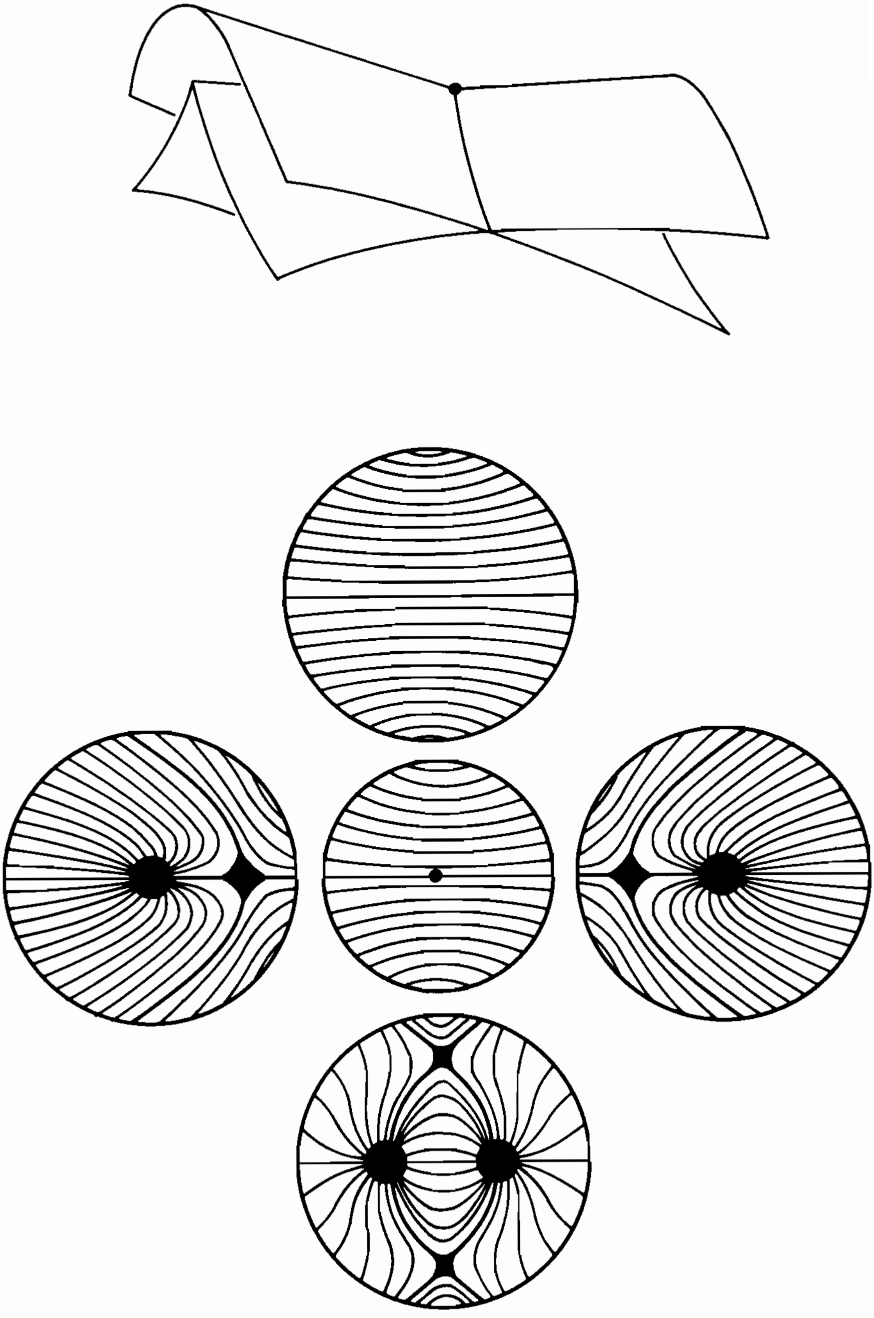

Figure 4.8. The hyperbolic umbilic case.

combinations of the parameters. In the middle is the flow picture in the particular (and most unstable) case where $a = b = c = 0$. This is, in fact, the monkey saddle flow in the case of the elliptic umbilic, whereas in the case of the hyperbolic umbilic the corresponding flow is laminar with the exception of an isolated stagnation point.

Depending on the exact course of parameter change, any of the transitions, from one flow picture to another, is possible. We thus see how singularities fuse and split, emerge and disappear, and how the entire pictures seem to be subject to sudden rotations when the origin of parameter space is encountered.

Due to the classification theorem, we know that all structural change phenomena that are likely to occur (with three parameters causing external change) have been depicted. Of course, the characterization is topological, as always, and in contrast to our ideas on stable structure, the characterization of structural change is local rather than global.

In conclusion, it should be borne in mind that the considerations of structural stability and structural change developed above are equally applicable to the planning models in the following chapters, even if we do not explicitly repeat the argument in each case.

5 *Planning Models*

5.1 A LONG-RUN MODEL WITH COSTLESS RELOCATION OF RESOURCES

5.1.1 Introduction

The model presented here is designed to handle the following planning problem. There exists a geographical region of given shape and extent. We consider a number of different productive activities, represented by linearly homogeneous production functions, allowing smooth substitution among inputs. In order to emphasize the advantages of geographical specialization, even in the absence of localized input supplies, we assume that the same production functions apply at all locations.

There is a local utility function, dependent on the quantities of produced goods available for consumption. The goal is to maximize the total utility obtained by aggregation with respect to all locations. The maximum is obtained by means of appropriate distribution of given aggregates of capital and labor among locations and among productive activities. The third classical input, land, is immobile and hence only the division of land among various activities at each location is considered.

Local consumption may differ from local production for any good, so that commodity flows and production of transportation services have to be specified. Transportation, of course, also uses up inputs. More specifically, we assume that transportation services, very specific in type, are produced by a Leontief technology without substitution and that only capital and labor, but not land, are used. This is fairly realistic if we consider transportation costs in terms of fuel, drivers' services, and wear and tear on vehicles. The inputs embodied in the existing network of roads are not taken into explicit consideration, as the planning of a new network is an even more long-run undertaking than the planning of an optimal spatial distribution of production activities.

It should be emphasized that housing is included in the productive activities under consideration. A "flow" of housing, which might seem at first meaningless, simply means that workers live at locations other than those of their occupation. Whether the commodities are physically moved to the consumer or the consumer moves in order to consume housing or public services is of no importance. We can consider a movement of either consumers or services, provided that the costs are accounted for correctly.

The main outcome of the analysis is a principle of geographical specialization as opposed to the possibility of producing everything locally without any interregional trade. This specialization occurs even in the absence of comparative advantages, since the same productive possibilities are available everywhere. It should be noted that the main conclusions are independent of which particular utility function we postulate.

The mathematical paradigm is that of a continuous two-dimensional space where areal densities of consumption, production, and inputs are taken into account. These areal densities for land, of course, are fractions that, at any location, add up to a given constant, which is at most unity for all space-consuming activities. All these areal densities are assumed to be smooth functions of the space coordinates. In the same way, the flows of goods are regarded as continuous flows in the plane. They take paths that minimize transportation costs between any pair of locations. The structure of roads is represented by a location-dependent, but direction-independent, need for capital and labor. Transportation cost is given by the line integral of the costs for inputs at all the locations traversed by a particular route. The optimal paths are thus obtained by solution of Euler equations for well-defined variational problems.

The continuous flow concept also implies that, if we know the optimal flow directions, the local changes in flow volumes can be linked to the local excess supplies.

5.1.2 The Model

Let x_1, x_2 denote the space coordinates. We are dealing with a region A of two-dimensional Euclidean space, bounded by a simple smooth curve ∂A. Unless otherwise indicated, all the variables introduced are functions of x_1, x_2. Surface integrals are taken over all of A and line integrals along the entire boundary ∂A, again unless counterindicated.

There are n different commodities (goods or services, including housing but not transportation). If the quantities of these commodities available for consumption at a given location x_1, x_2 are $q_1, q_2, \ldots, q_n$, then the local utility is $U(q_1, q_2, \ldots, q_n, x_1, x_2)$ and the total utility to be maximized is

$$\iint U(q_1, q_2, \ldots, q_n, x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{1}$$

Explicit inclusion of the space coordinates makes it possible to assign different weights to consumption in various locations. To begin with, however, we simplify the expression by deleting these x_1, x_2 and occasionally we let the utility function take the form $\Sigma\epsilon_i \ln q_i$. Unless otherwise indicated, all summations run over $i = 1, 2, \ldots, n$.

Let k_i, l_i, m_i denote the areal densities of capital, labor, and land used in the ith productive process at a given location x_1, x_2. The linearly homogeneous production functions are then

$$f^i(k_i, l_i, m_i) \equiv m_i f^i(k_i/m_i, l_i/m_i, 1) \tag{2}$$

Unless otherwise indicated, expressions written for some index i are assumed to hold for all $i = 1, 2, \ldots, n$. Since the space coordinates are not explicitly included we assume that the same production possibilities are available everywhere. For the sake of example we let

$$f^i = A_i k_i^{\alpha_i} l_i^{\beta_i} m_i^{\gamma_i}$$

where the indices $\alpha_i + \beta_i + \gamma_i = 1$. Local excess supplies are

$$f^i(k_i, l_i, m_i) - q_i \tag{3}$$

which enter into the commodity flows or, if negative, are withdrawn from them. We denote the commodity flows by ϕ_i. These flows are vector fields, i.e. the ϕ_i are two-dimensional vectors whose components are functions of the space coordinates x_1, x_2. Of course, a vector field has both direction and magnitude. The direction is simply the actual direction of the flow considered and the magnitude is the quantity of commodities shipped in the flow.

According to one of the basic theorems in vector analysis, Gauss's divergence theorem, the divergence of a vector field represents the source density of an incompressible flow such as the transportation of commodities. The source density, in this case, of course, is local excess supply so that, in view of (3), we may write

$$\operatorname{div} \phi_i = f^i(k_i, l_i, m_i) - q_i \tag{4}$$

Mathematically, the divergence of a vector field equals the partial derivative of its first component with respect to the first space coordinate plus the partial derivative of its second component with respect to the second space coordinate. Thus (4) are partial differential equations for the magnitudes $|\phi_i|$ of the vectors as soon as the flow directions $\phi_i/|\phi_i|$ and excess supplies on the right-hand side are known. Later on, we shall return to the determination of the flow directions.

As stated in the Introduction, the transportation of goods uses up capital and labor inputs, say $\kappa_i|\phi_i|$ and $\lambda_i|\phi_i|$, respectively. The κ_i and λ_i are given functions of the space coordinates and reflect the structure of fixed transportation capacity provided by the existing road network. The linear dependence on flow magnitudes means that we abstract from congestion. This simplifies the analysis a great deal. A nonlinear dependence on $|\phi_i|$ is not difficult to handle, but the interference of the different flows augments the degree of analytical complication disproportionately to the increase in realism it achieves.

If there are given aggregate resources of capital and labor, denoted K and L, we arrive at the following constraints

$$\iint \sum (k_i + \kappa_i |\phi_i|) \mathrm{d}x_1 \mathrm{d}x_2 = K \tag{5}$$

$$\iint \sum (l_i + \lambda_i |\phi_i|) \mathrm{d}x_1 \mathrm{d}x_2 = L \tag{6}$$

Production uses up k_i units of capital and l_i units of labor; transportation uses $\kappa_i|\phi_i|$ units of capital and $\lambda_i|\phi_i|$ units of labor. Summing over all commodities and integrating over all locations yields the total usage of these inputs.

As mentioned in the Introduction, it is assumed that in planning we are completely free to move capital and labor among locations and among activities. As for land, this may only be transferred between activities. So

$$\sum m_i = m \tag{7}$$

where m is a positive, at most unitary, location-dependent number. In general it is less than unity, since some space has already been used up in constructing the given fixed transportation capacity or is otherwise unavailable for further exploitation.

We thus have a well-defined optimization problem, namely to maximize (1) subject to the constraints (4), (5), (6), and (7) by choosing the appropriate scalar fields k_i, l_i, m_i, and q_i and the vector fields ϕ_i. This will be accomplished using a Lagrangean method. We associate Lagrange multipliers p_i with (4), r with (5), w with (6), and g with (7). For the time being they are only undetermined multipliers, but the notation indicates that they turn out to be shadow prices for goods, capital rent, wage rate, and land rent, respectively. They can also be interpreted as equilibrium prices in a competitive system with individually optimizing agents.

5.1.3 Optimum for Production

We now derive the optimum conditions, starting with those obtained by maximizing with respect to k_i, l_i, and m_i

$$p_i f_k^i(k_i, l_i, m_i) = r \tag{8}$$

$$p_i f_l^i(k_i, l_i, m_i) = w \tag{9}$$

and

$$p_i f_m^i(k_i, l_i, m_i) = g \tag{10}$$

These can be recognized as the common marginal conditions for profit-maximizing firms. With production functions homogeneous of degree one, the marginal productivity functions f_k^i for capital, f_l^i for labor, and f_m^i for land become homogeneous of degree zero. So, taking the first two marginal conditions alone we obtain the system

$$f_k^i(k_i/m_i, l_i/m_i, 1) = r/p_i \tag{11}$$

and

$$f_l^i(k_i/m_i, l_i/m_i, 1) = w/p_i \tag{12}$$

This system (11)–(12) is indeed smoothly invertible as the Jacobian is nonzero due to second-order conditions for profit maximization. By the inverse function theorem, we get

$$k_i/m_i = F_k^i(r/p_i, w/p_i) \tag{13}$$

and

$$l_i/m_i = F_l^i(r/p_i, w/p_i) \tag{14}$$

As (10) can take the form

$$f_m^i(k_i/m_i, l_i/m_i, 1) = g/p_i \tag{15}$$

by substituting from (13) and (14) we obtain

$$g/p_i = f^i_m(F^i_k(r/p_i, w/p_i), F^i_l(r/p_i, w/p_i), 1) \tag{16}$$

which links product price to the three input prices. The conclusion is that if capital rent, wages, and land rent are given, (16) determines the prices of all goods produced at all locations, provided that production takes place. This is an important conclusion to be used later on.

The equivalence of the optimum conditions introduced and the profit-maximizing conditions for an individual firm at a given location will become obvious from the following considerations. Assume that a firm has to maximize its profits by choosing an appropriate mix of productive activities. Capital and labor services are freely available at the local prices r and w, whereas the firm owns a fixed amount of land m available for all its activities. For capital and labor, the optimum conditions at given product prices p_i are (8)–(9) or equivalently (11)–(12). We can then invert the system to obtain (13)–(14). Substituting (13)–(14) into the production function and using (2), the profits of the firm are

$$\sum\{p_i f^i(F^i_k(r/p_i, w/p_i), F^i_l(r/p_i, w/p_i), 1) - rF^i_k(r/p_i, w/p_i) - wF^i_l(r/p_i, w/p_i)\}m_i \tag{17}$$

This is to be maximized subject to the constraint (7) on the total quantity of land available. In view of the fact that both the maximand and the constraint are linear, the solution is to put $m_i = m$ for that i which maximizes

$$p_i f^i - rF^i_k - wF^i_l \tag{18}$$

and $m_i = 0$ for the other activities. If several activities are to be profitable, (18) must be equal for all of them. This common value could be called g, which represents the profits imputed to the land-owning firms as land rent. If all activities take place, we get

$$p_i f^i - rF^i_k - wF^i_l = g \tag{19}$$

for all i. In view of Euler's theorem for homogeneous functions

$$f^i = f^i_k k_i + f^i_l l_i + f^i_m m_i \tag{20}$$

and using (7), (11)–(12), and (13)–(14), we see that (19) is exactly the same as (16). This establishes the local equivalence between profit maximization and overall planning.

Conditions (8)–(9) also contain additional information. In view of the fact that (5)–(6) are integral constraints, the associated Lagrange multipliers r and w must be constant with respect to the space coordinates. This means that the efficiency conditions for allocating capital and labor in space require capital rent and wage rate to be constant with respect to location. This is not true for land rent g as it is a Lagrange multiplier for the constraint (7), which is local, i.e. not in integral form.

The conclusions we can draw from all this are that (16) determines all the p_i for which production is to take place, and that the variations in production opportunity prices in space are determined solely by variations in land rent. Capital rent and wages are spatially invariant due to distributive efficiency requirements.

5.1.4 Optimum for Flows

Next, we turn to the optimum conditions for the commodity flows, i.e. the maximization of (1) with respect to the ϕ_i, given the constraints (4)–(7). The flows appear in two ways in the constraints, as $|\phi_i|$ in (5)–(6) and as div ϕ_i in (4). The Lagrange multipliers associated with these constraints are the p_i, r, and w. The optimum conditions expressed as Euler equations are

$$(r\kappa_i + w\lambda_i)\phi_i/|\phi_i| = \operatorname{grad} p_i \tag{21}$$

These conditions mean that the flow directions $\phi_i/|\phi_i|$ agree with the directions, grad p_i, of the steepest increase in p_i, and that along the flow lines the p_i increase at a rate of $(r\kappa_i + w\lambda_i)$. We recall that κ_i and λ_i were the local requirements of capital and labor for the transportation of a unit of the ith commodity. Accordingly $(r\kappa_i + w\lambda_i)$ is the local cost for transportation. As the p_i were interpreted as product prices, (21) simply states that each commodity flow takes the direction of the steepest increase in its price and that prices in this direction increase by transportation costs. This makes good economic sense.

In the preceding section we concluded that an efficient distribution of capital and labor in the region requires capital rent and wage rate to be independent of location. In passing, we may note that this can be interpreted in market equilibrium terms, i.e. that when capital and labor are free to move, they seek the place of production where the reward is the highest. In the absence of relocation costs this equalizes factor prices in space. The consequence of this and the fact that κ_i and λ_i are given functions of the space coordinates is that the increases in prices along the optimal routes are also given functions of the space coordinates. In fact, from (21) we obtain

$$|\operatorname{grad} p_i| = r\kappa_i + w\lambda_i \tag{22}$$

These are partial differential equations for the prices p_i with the right-hand sides as given functions of the space coordinates.

5.1.5 Specialization

We are now in a position to prove a general specialization theorem. From (16) we see that with r and w given, g and p_i are related by continuous one-to-one mappings as long as the Jacobians of the systems (11)–(12) are nonzero, which is assumed in traditional economic theory. We could write (16) as

$$p_i = p_i(g) \tag{23}$$

from which we obviously get $|\operatorname{grad} p_i| = p_i'(g)|\operatorname{grad} g|$.

In (22), the right-hand side designates given functions of the space coordinates, say $r\kappa_i + w\lambda_i = \theta_i h(x_1, x_2)$. The θ_i can be interpreted as characteristic constants for each good. It is a reasonable simplification to assume that if the shipping of one good costs twice as much as the shipping of another good at one location, the same relation will hold everywhere in the region. Hence, equating the two expressions for $|\text{grad } p_i|$, we get

$$\theta_i h(x_1, x_2) = p_i'(g) \, |\text{grad } g| \tag{24}$$

These conditions can hold for several commodities, say the ith and jth, only if the ratios $p_i'(g)/p_j'(g)$ take the constant value θ_i/θ_j everywhere. But there is no reason whatsoever why the $p_i'(g)$ functions should be linearly dependent. After all, they were obtained from (16), which depended on the various independent production functions.

We thus conclude that, with respect to goods that are transported, only one commodity will be produced at each point in the region. The continuity of the production function and a nonzero Jacobian for system (11)–(12) guarantee that this specialization will not only apply to sets of measure zero such as isolated points or curves, but will split the region into a finite collection of subregions of nonzero areas with specialized activity in each. The land rent in each of these regions will be determined by the local revenue shares for these specialized activities.

The reader should note the affinity between our conclusion and von Thünen's theory, where specialization in concentric rings occurs, despite the fact that there are no localized productivity differences. In general economic theory, trade is supposed to occur only when there are at least comparative localized advantages, due to immobility of inputs. There are no such advantages in our model. Nevertheless, specialization does occur. The reason is that when numerous outputs are ultimately produced from a few primary inputs, output prices are tied to the few input prices. In order for production of all the outputs to be equally profitable, their prices must covary spatially in a very specific way. The result is a specialization pattern that is inherent in two-dimensional space itself. It is then not surprising that this point is missed in trade theory, since general economics lacks the spatial dimension.

5.1.6 Independence of Utility Functions

Before continuing, it should be noted that the optimality conditions for production and transportation are independent of the utility function (1). Hence, regardless of how we evaluate the availability of the various commodities in different locations, the following conclusions apply. Labor and capital should seek the locations of best reward, which, under conditions of free mobility, equalize capital rent and wages over space. Production everywhere should be arranged as if land-owning firms tried to maximize their profits, which must equal local land rents. Commodity flows should take the directions in which prices increase most steeply, and the price increases in these directions should equal local transportation costs. The result is such that, if there are commodity flows, then there should be specialization in the production of only one commodity at each location.

These conclusions are derived from consideration of a planning problem constrained by available resources. But the result could be interpreted equally well in terms of rationally behaving individual workers, capitalists, landowning producers, and transporters in a state of general equilibrium. In particular the conclusions are independent of which social utility function $U(q_1, q_2, \ldots, q_n, x_1, x_2)$ is used. The only optimality conditions in which this function plays a role are the

$$\partial U/\partial q_i = p_i \tag{25}$$

which state that marginal utility should equal price everywhere. Conditions (25) pose a set of additional constraints on the model that relates local commodity prices to local consumption of goods.

A similar result is obtained by considering the behavior of individual consumers who dispose of their incomes so as to maximize their individual utility functions. The demand functions thus obtained are similar in structure to the inverted system (25), but caution is advised in the planning case where local budget constraints might not be fulfilled automatically. If we still wish to admit customer autonomy, we might have to consider an interregional income-transfer policy as a means of fulfilling the planning objectives. This, however, is the only point where a contradiction between planning and market equilibrium could arise.

5.1.7 Macro Relations

We now establish a number of macro relations within the model. Observe that, according to a general formula in vector analysis

$$\operatorname{div}(p_i\phi_i) = (\operatorname{grad} p_i)\phi_i + p_i \operatorname{div} \phi_i \tag{26}$$

holds identically for any scalar field p_i and any vector field ϕ_i. From Gauss's divergence theorem it follows that

$$\iint \operatorname{div}(p_i\phi_i)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int p_i(\phi_i)_n\,\mathrm{d}s \tag{27}$$

The left-hand double integral is taken over all of the region, whereas the right-hand line integral is taken along the boundary of the region. The $(\phi_i)_n$ are the components of the vector fields ϕ_i normal to the boundary. Hence, the $p_i(\phi_i)_n$ have the simple interpretations of value exports or imports, depending on sign, across the boundary. The line integrals take care of all flows across the entire boundary and hence the right-hand side of (27) equals net exports from the region. Let us therefore define

$$X_i - M_i = \int p_i(\phi_i)_n\,\mathrm{d}s \tag{28}$$

Next we note that, due to (21)

$$(\operatorname{grad} p_i)\phi_i = (r\kappa_i + w\lambda_i)\,|\phi_i| \tag{29}$$

The right-hand expression is the product of local transportation costs (as evaluated by the input requirements and the local factor costs) and the quantities of commodities shipped. Taking the double integral of (29), we arrive at the total transportation costs, denoted T_i. Thus

$$\iint (\operatorname{grad} p_i)\phi_i \,\mathrm{d}x_1\,\mathrm{d}x_2 = T_i \tag{30}$$

On the other hand, (4) and the well-known fact that with linearly homogeneous production functions all revenues are distributed as factor shares, i.e. $p_i f^i = rk_i + wl_i + gm_i$, yield

$$p_i \operatorname{div} \phi_i = rk_i + wl_i + gm_i - p_i q_i \tag{31}$$

Denoting, in aggregate for a given branch, capital incomes by R_i, wages by W_i, the profits of landlords by G_i, and the value of consumption at local prices by C_i, we get

$$\iint p_i \operatorname{div} \phi_i \,\mathrm{d}x_1\,\mathrm{d}x_2 = R_i + W_i + G_i - C_i \tag{32}$$

Now, integrating both sides of (26) and substituting from (27)–(28), (30), and (32), we obtain

$$X_i - M_i = T_i + (R_i + W_i + G_i) - C_i \tag{33}$$

i.e. net value exports for each branch equal factor incomes plus transportation costs minus consumption.

If we now sum over all the various branches, we can define $X - M = \Sigma(X_i - M_i)$, $T = \Sigma T_i$, $G = \Sigma G_i$, and $C = \Sigma C_i$. But with regard to capital and labor income it should be recalled that not all of these inputs are accounted for in (33). Due to (5) and (6), some quantities are used in transportation. We have not yet accounted for the incomes of the transporters. Hence, $\Sigma(R_i + W_i) = R + W - T$. The result is then

$$X - M = R + W + G - C \tag{34}$$

which simply means that, in value terms, net exports equal factor incomes minus consumption.

In a regional economy with zero balance of payments, where $X = M$ so that net imports of some goods are bought by net exports of other goods, we conclude that the sum of aggregate factor incomes is the value of aggregate consumption. This is not a trivial conclusion because both incomes and consumption are evaluated at local prices.

The result establishes an aggregate budget constraint for the economy and hence the model is consistent with consumer autonomy and locally fulfilled budget constraints. Consistency, however, does not guarantee local fulfillment of budget constraints for states we wish to consider. But it does establish that if a socially desirable spatial organization of the region does not lead to local fulfillment of budget constraints, we can always design an appropriate, completely internal income-transfer policy that makes budget constraints hold locally and permits free choice for consumers.

5.1.8 Examples

We now describe two examples of possible spatial organization patterns based on the model outlined above. Assume first that the fixed transportation capacity is equally distributed in space so that all the κ_i and λ_i are constants. Owing to the constancy of r and w, the local transportation costs $\theta_i = (r\kappa_i + w\lambda_i)$ are also invariant in space.

Let all $\phi_i = \pm \text{ grad } \rho = \pm(x_1/\rho, x_2/\rho)$, where $\rho = (x_1^2 + x_2^2)^{1/2}$. If $p_i = \bar{p}_i + \theta_i|\rho - \rho_i|$, then equation (21) is fulfilled. The flows all become radial and the constant price contours become concentric circles. This suggests a production specialization structure in concentric rings, as in the familiar von Thünen case. The difference is that there is not a single "central business district" in the center to which all commodities flow. Rather, the whole region is supplied by commodities produced in each ring. This case is illustrated in Figure 5.1 where, for illustrative purposes, we show a four-commodity model with activities called public services (S), industry (I), housing (H), and agriculture (A).

In the second example we assume that fixed transportation capacity is not equally distributed in space but is concentrated in the central parts of the region. Suppose that all the κ_i and λ_i are proportional to ρ, where again $\rho = (x_1^2 + x_2^2)^{1/2}$. Thus we can write local transportation cost as $(r\kappa_i + w\lambda_i) = \theta_i\rho$, where the θ_i are again constants. We can now set all $\phi_i = \pm \text{ grad} \frac{1}{2}(x_1^2 - x_2^2) = \pm (x_1, -x_2)$. If we let $p_i = \bar{p}_i + \theta_i|\frac{1}{2}(x_1^2 - x_2^2) - \bar{\sigma}_i|$ then (21) is again fulfilled. The flow lines integrate to hyperbolas, $xy = \text{constant}$, and the constant price lines are hyperbolas, $(x^2 - y^2) = \text{constant}$, rotated by 45° in comparison to the paths (see Figure 5.2).

These illustrations agree perfectly with the optimality conditions stated. They are not chosen at random but represent spatial organization around singularities of the only types admitted under the assumption of structural stability; cf. Chapter 4. Structural stability considerations apply to the planning case, in the same way as the considerations in this chapter on specialization apply to the equilibrium case.

5.1.9 Intermediate Goods

Intermediate products were not taken into account in the preceding analysis. In particular, it is interesting to know whether the specialization theorem still holds even when it implies that an output could be shipped to another place to produce something that is reimported rather than locally produced. In fact, the theorem does still hold, as is shown now for the case of a Cobb–Douglas technology. Let

$$f^i = A_i k_i^{\alpha_i} l_i^{\beta_i} m_i^{\gamma_i} \prod_j (f_i^j)^{\epsilon_{ij}} \tag{35}$$

where f_i^j denotes the quantity of output j used as input in the production of output i. The product in (35) is taken over all indices j from 1 to n. Linear homogeneity now means that

$$\alpha_i + \beta_i + \gamma_i + \sum_j \epsilon_{ij} = 1 \tag{36}$$

The optimum conditions corresponding to (8)–(10) are obviously

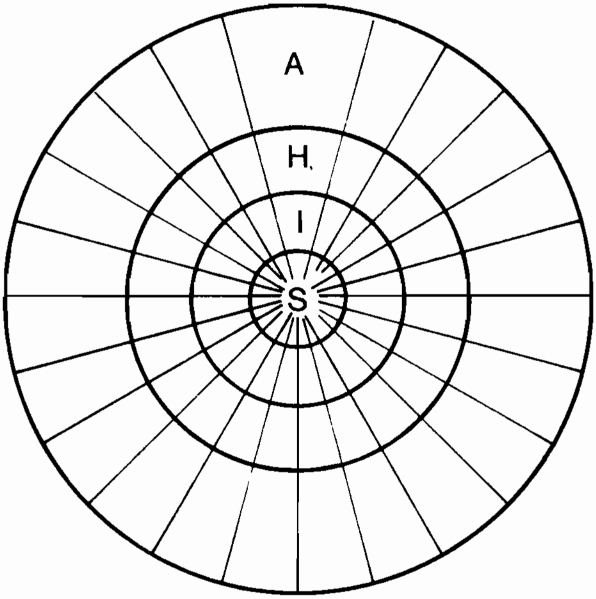

Figure 5.1. Organization in rings of a four-activity economy.

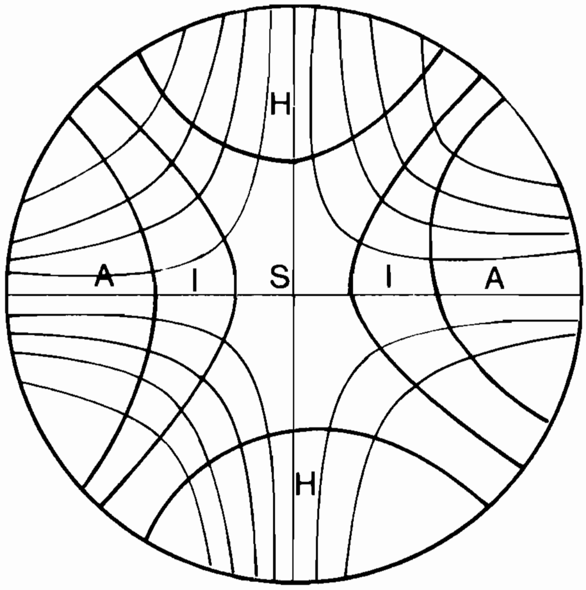

Figure 5.2. Organization in sectors of a four-activity economy.

$$\alpha_i p_i f^i / k_i = r \tag{37}$$

$$\beta_i p_i f^i / l_i = w \tag{38}$$

$$\gamma_i p_i f^i / m_i = g \tag{39}$$

and

$$\epsilon_{ij} p_i f^i / f_i^j = p_j \tag{40}$$

We can substitute back from (37)–(40) into (35) and, in view of (36), obtain

$$A_i \alpha_i^{\alpha_i} \beta_i^{\beta_i} \gamma_i^{\gamma_i} \prod_j (\epsilon_{ij})^{\epsilon_{ij}} p_i = r^{\alpha_i} w^{\beta_i} g^{\gamma_i} \prod_j (p_j)^{\epsilon_{ij}} \tag{41}$$

Taking logarithms we get a set of n linear equations in the logarithms of the $(n + 3)$ prices. Regarding r, w, and g as given, we can solve for the logarithms of all the p_i, as the matrix of the system $[\epsilon_{ij} - \delta_{ij}]$ (where δ_{ij} is the Kronecker delta) is nonsingular. Accordingly, the $\ln p_i$ are obtained as explicit linear expressions of $\ln r$, $\ln w$, and $\ln g$. After taking exponentials and substituting, we transform (41) into the explicit form

$$p_i = B_i r^{\hat{\alpha}_i} w^{\hat{\beta}_i} g^{\hat{\gamma}_i} \tag{42}$$

where the B_i, as well as the exponents, are constants that can be calculated from the original constants in (35).

Consider now a proportional change in r, w, g, and all the p_i. The solution to (37)–(40), whatever it is, is obviously unchanged so that (35) is still fulfilled. This demonstrates that (42) must hold for proportional changes in all prices, that is

$$\hat{\alpha}_i + \hat{\beta}_i + \hat{\gamma}_i = 1 \tag{43}$$

We can thus substitute back from (37)–(39), disregarding (40) altogether, and obtain

$$f^i = \hat{A}_i k_i^{\hat{\alpha}_i} l_i^{\hat{\beta}_i} m_i^{\hat{\gamma}_i} \tag{44}$$

due to (43). Now these are Cobb–Douglas production functions in the primary inputs only and they are linearly homogeneous in them.

Accordingly, as (44) fulfills the condition (2), the entire chain of reasoning with regard to specialization still holds. This, of course, does not preclude the possibility that if there is a certain hierarchy, so that goods produced at a certain stage are never used in the production of any of their inputs, then the flows should simply take a one-way route to higher levels. When interdependence is more complicated, however, it is possible that goods are reimported at a later stage of refinement.

5.1.10 Local and Global Optima

It should be noted that the optimality conditions stated so far are local in character. Determination of the global optimum is a matter the outcome of which is likely to change with the boundary conditions.

Our specialization theorem states that at every location in the region there is complete specialization in the production of traded goods. But if the utility function does not include the space coordinates as explicit arguments, i.e. if a certain consumption is equally valued at all locations, then local production and no trade constitute a solution that fulfills all local optimality conditions. And, since the goods are not traded, the specialization theorem does not exclude this possibility. For certain cases the solution is probably a global optimum, as the given input quantities are used most efficiently when no portion of them is "wasted" in moving commodities around.

This can be illustrated by simplifying the model. As trade rather than specialization is at issue, we can discuss a one-commodity economy. The production function for this commodity can be rendered in a Cobb–Douglas form and the utility function can be assumed to be logarithmic and without explicit dependence on the space coordinates. We do not specify any transportation technology, but assume in the traditional von Thünen way that the product may be used up in transportation. The unit of distance is normalized so that the cost of moving one unit of goods one distance unit uses up exactly one unit of the goods.

We thus have the following problem. Maximize

$$\iint \ln q \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{45}$$

subject to

$$K = \iint k \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{46}$$

$$L = \iint l \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{47}$$

and

$$k^{\alpha} l^{\beta} - q - |\phi| = \operatorname{div} \phi \tag{48}$$

The optimality conditions are then

$$1/q = p \tag{49}$$

$$\alpha \, k^{\alpha} l^{\beta}/k = r/p \tag{50}$$

$$\beta \, k^{\alpha} l^{\beta}/l = w/p \tag{51}$$

and

$$\phi/|\phi| = \operatorname{grad} \ln p \tag{52}$$

In these conditions r and w are independent of the location coordinates, whereas p is not. The conditions state that:

(i) Local marginal utility is everywhere equal to the opportunity costs for goods in the flow.

(ii) Production is everywhere so arranged that marginal value products of the inputs equal their local opportunity costs. With constant r and w, these opportunity costs are equal in space and there is no incentive to relocate inputs.

(iii) The flow of traded goods is in the direction of the steepest price increase and the rate of increase in this direction is exponential, since moving one unit of goods uses up one unit of its own value.

We see that (49)–(51) determine inputs k and l, output $k^{\alpha} l^{\beta}$, and consumption q, once r, w, and p are known. As r and w take constant values, determined by the constraints (46)–(47), we see that p completely determines the spatial densities.

So, let us choose any function $p(x_1, x_2)$ such that $|\text{grad} \ln p| = 1$. Then (52) is fulfilled and $\phi/|\phi| = (\cos\theta, \sin\theta)$ is a known unit vector field. As $\text{div}\, \phi = \text{grad}\, |\phi| \cdot (\cos\theta, \sin\theta) + |\phi|\, \text{div}\, (\cos\theta, \sin\theta)$, (48) becomes a partial differential equation in the flow intensity $|\phi|$. Solution of this differential equation solves the whole problem. Hence we have seen that any price structure such that

$$|\text{grad} \ln p| = 1 \tag{53}$$

holds can represent a sensible local optimum. This can be illustrated by two different solutions.

First, set $\rho = (x_1^2 + x_2^2)^{1/2}$ and $\phi/|\phi| = \text{grad}\, \rho$. This flow obviously satisfies (52) for $p = e^{\rho}$. Assuming now that $\alpha = \beta = r = w = 1/4$, we get from (49)–(51) that $k^{\alpha} l^{\beta} = p$, and $q = 1/p$. Thus $k^{\alpha} l^{\beta} - q = e^{\rho} - e^{-\rho} = 2 \sinh \rho$. This result can be substituted into (48). But $\text{div}\, \phi = \text{grad}\, |\phi| \cdot \text{grad}\, \rho + |\phi|\, \text{div grad}\, \rho$. Using polar coordinates, $x_1 = \rho \cos\omega$ and $x_2 = \rho \sin\omega$, we easily get $\text{grad}\, |\phi| \cdot \text{grad}\, \rho = \mathrm{d}|\phi|/\mathrm{d}\rho$. Moreover, $\text{div grad}\, \rho = 1/\rho$. Thus (48) becomes an ordinary linear differential equation

$$\frac{\mathrm{d}|\phi|}{\mathrm{d}\rho} + \left(1 + \frac{1}{\rho}\right)|\phi| = 2 \sinh \rho \tag{54}$$

Given a simple boundary constraint, the equation is readily solved. The spatial organization associated with this solution is one where goods flow radially outwards and price increases at an exponential rate in this direction, whereas consumption decreases outwards and production increases outwards. As excess supply is zero at the origin and decreases outwards, the case is impossible unless there is a singularity with net outflow at the origin.

Second, we easily see that by setting $k, l, k^{\alpha} l^{\beta} = q = 1/p$ constant, and $|\phi|$ identically equal to zero, all the equations are fulfilled. Thus, this case of no trade and local production is another local optimum. It is difficult to tell which of the two cases is a global optimum.

The reader might ask whether there are always just two local optima: one with trade and one without. In fact, it is easy to find cases with more than two local optima. Let us change the model (45)–(48) by assuming that the cost of movement is not the same everywhere in the region, but that it increases in proportion to the distance from the origin, so that communications are best in the center and become worse at the periphery. Thus, we assume that $\rho|\phi|$ units of the goods are used up in moving one unit of goods one distance unit. Then (48) is changed to

$$k^\alpha l^\beta - q - \rho|\phi| = \operatorname{div}\phi \tag{55}$$

Only (52) in the optimality conditions is changed by this and now takes the form

$$\rho\phi/|\phi| = \operatorname{grad}\ln p \tag{56}$$

In accordance with this, (53) is changed to

$$|\operatorname{grad}\ln p| = \rho = (x_1^2 + x_2^2)^{1/2} \tag{57}$$

We can now easily find at least three different solutions to (55), namely p = constant and $p = \exp((\pm x_1^2 \pm x_2^2)/2)$. The latter are actually four cases, but after discarding reversals of flow directions we are left with two qualitatively different flows, one radial and one saddle. It is immediately apparent that the nontrade, the radial, and the saddle flows are all local optima. Again it is difficult to tell which one is a global optimum without considering the boundary conditions.

This multiplicity of local optima did not occur in our equilibrium model in the preceding chapter, as a price–flow distribution on the boundary was taken as given from world market conditions. To the extent it seems reasonable to use an analogous boundary condition in the planning problem, the arbitrariness will be removed. This may be reasonable, as acceptance of the trade conditions, determined by trade outside the region studied, may lead to maximum benefit from trade with the exterior.

5.1.11 Boundary Constraints

Let us consider these constraints from a more formal point of view. From (55) we see that

$$pk^\alpha l^\beta - pq = p\operatorname{div}\phi + p\rho|\phi| \tag{58}$$

But, from (56), $p\rho|\phi| = \operatorname{grad} p\cdot\phi$. Substituting this and using the identity $\operatorname{div}(p\phi) = \operatorname{grad} p\cdot\phi + p\operatorname{div}\phi$, we get

$$\iint (pk^\alpha l^\beta - pq)\,dx_1\,dx_2 = \iint \operatorname{div}(p\phi)\,dx_1\,dx_2 \tag{59}$$

From Gauss's theorem, the right-hand side equals the curve integral $\int p(\phi)_n$. This, however, is zero in two cases: when $(\phi)_n$ vanishes identically on the boundary, and when it does not vanish but trade with the exterior balances. Obviously, we only need to be concerned with these two cases of either complete insulation or of balancing interregional trade.

Setting the right-hand side of (59) equal to zero yields

$$\iint pk^\alpha l^\beta\,dx_1\,dx_2 = \iint pq\,dx_1\,dx_2 \tag{60}$$

Accordingly the aggregate value of output equals the aggregate value of consumption. Now, the optimality condition (49) states that marginal utility equals product price.

With our logarithmic utility function we have $pq = 1$ in the whole region. As the integrand in the right-hand side of (60) is unitary, we conclude that the integral equals the area of the region. Denoting this (the total quantity of land) by M, we get

$$\iint pk^{\alpha} l^{\beta} \, dx_1 \, dx_2 = M \tag{61}$$

Let us next substitute from (50)–(51) into the production function and solve for

$$k^{\alpha} l^{\beta} = \left(\frac{\alpha}{r}\right)^{\alpha/\gamma} \left(\frac{\beta}{w}\right)^{\beta/\gamma} p^{(\alpha+\beta)/\gamma} \tag{62}$$

where $\gamma = 1 - \alpha - \beta$. As r and w are spatial constants, local output is proportional to a power function of the price p. We can also solve for k and l from (50)–(51) and integrate to obtain

$$\frac{\alpha}{r} \iint pk^{\alpha} l^{\beta} \, dx_1 \, dx_2 = K \tag{63}$$

and

$$\frac{\beta}{w} \iint pk^{\alpha} l^{\beta} \, dx_1 \, dx_2 = L \tag{64}$$

Substituting from (61) into (63)–(64) we get $\alpha/r = K/M$ and $\beta/w = L/M$, which can be substituted into (62). The result is

$$k^{\alpha} l^{\beta} = \left(\frac{K}{M}\right)^{\alpha/\gamma} \left(\frac{L}{M}\right)^{\beta/\gamma} p^{(\alpha+\beta)/\gamma} \tag{65}$$

Local output is thus a Cobb–Douglas function of the average areal densities of capital and labor, multiplied by the aforementioned power function of local price.

These relations must hold in any case where either there is no trade with the exterior or exterior trade balances.

We note that output $k^{\alpha} l^{\beta}$ is an increasing function of price. This function is given and identical in all cases that may be regarded as candidates for a global optimum. From (49), on the other hand, we know that consumption q is a decreasing function of price. So, excess supply

$$z = k^{\alpha} l^{\beta} - q \tag{66}$$

is certainly an increasing function of p. Thus, considering two different cases, distinguished by subscripts i and j, we conclude that

$$(p_i - p_j)(z_i - z_j) \geqslant 0 \tag{67}$$

must hold at all locations.

Let us now treat two alternative price–flow patterns that fulfill the optimality conditions. Consider the value flows

$$p_i k_j^\alpha l_j^\beta - p_i q_j = p_i \operatorname{div} \phi^j + p_i \rho |\phi^j| \tag{68}$$

It is true that

$$|\phi^j| \geqslant (\phi^i/|\phi^i|)\cdot\phi^j \tag{69}$$

since projection of the vector ϕ^j on the direction $\phi^i/|\phi^i|$ results, at most, in the norm $|\phi^j|$. So, using the optimality condition (56) for the flow ϕ^i, (69) yields

$$p_i\rho|\phi^j| \geqslant \operatorname{grad} p_i\cdot\phi^j \tag{70}$$

If we substitute from (70) into (68), we see that the right-hand side must at least equal $\operatorname{div}(p_i\phi^j)$. Using the notation z_j for excess supply from (66) we thus get

$$p_i z_j \geqslant \operatorname{div}(p_i\phi^j) \tag{71}$$

and by integration and use of Gauss's theorem

$$\iint p_i z_j \,\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant \int p_i(\phi^j)_n \,\mathrm{d}s \tag{72}$$

with equality when $i = j$, as seen from (56).

With regard to the right-hand side of (72), we conclude that it is zero if $i = j$, as already seen. This results from the trade balance condition. We also conclude that it is zero if both cases considered involve trade across the boundary, as $p_i = p_j$ are then determined by the "world market" on the boundary, and the trade balance condition requires the integrals $\int p_i(\phi^i)_n$ and $\int p_j(\phi^j)_n$ to be zero. The same is true when both cases represent insulation, as $(\phi^i)_n$ and $(\phi^j)_n$ are then identically zero.

The only situation where the right-hand side of (72) can be nonzero is when case i represents insulation and case j represents balancing trade. The nonzero flow across the boundary in the case of trade is then evaluated at the prices characteristic of insulation. We have no reason to expect that an integral such as this should be zero. As for the situation where both cases represent insulation or balancing trade, the right-hand side of (72) is zero regardless of how we permute i and j. Recalling that (72) holds as equalities when $i = j$, we get

$$\iint (p_i - p_j)(z_i - z_j)\,\mathrm{d}x_1\,\mathrm{d}x_2 \leqslant 0 \tag{73}$$

The only way a nonpositive integral can be obtained from a nonnegative integrand according to (67) is through an integrand that is identically zero, i.e.

$$(p_i - p_j)(z_i - z_j) \equiv 0 \tag{74}$$

As our excess supply function is strictly increasing we conclude that $p_i = p_j$ and $z_i = z_j$ must hold identically. Thus, any two solutions fulfilling the optimum conditions, together with the boundary condition, which states either that there is no trade with the exterior or that trade balances, are identical. So, the solution is unique. More specifically, there is a unique solution with trade and a unique solution with insulation.

Let us return to the situation where one case, say i, represents insulation and the other,

j, represents trade. Then one of the right-hand side integrals of (72) does not have to be zero and accordingly, the zero in (73) is replaced by the expression

$$-\int p_i(\phi^j)_n \, \mathrm{d}s \tag{75}$$

If this line integral should turn out to be strictly negative, then we are in trouble, as (73) does not hold, and the discussion leading to uniqueness would no longer be valid.

How likely is it that the curve integral in (75) is negative? Negativity obviously means that insulation prices p_i are lower than world market prices p_j where the flow ϕ^j leaves the region, and higher where it enters. On the other hand, world market prices are low where the flow enters and high where it leaves the region. This is because the flow of trade adjusts to the direction of increasing prices. We conclude that spatial price differences in the case of insulation must be smaller than the differences in world market prices.

But the price differences in the case of insulation are obtained as accumulated transportation costs. Therefore, since world market price differences between various points on the boundary are greater than the costs of transportation between them, it seems that a profit can be obtained from arbitrage across the region. This profit can be converted into increased consumption in the region.

As a result, it appears as if planning authorities should open up trade with the exterior when boundary price differences exceed transportation costs. This was the only case that caused trouble with respect to the uniqueness proof. We may conclude that it holds when the planning authorities give due consideration to trade opportunities with the exterior that benefit interior consumption.

5.2 RELOCATION COSTS FOR CAPITAL AND LABOR

5.2.1 Introduction

Let us now return to the problem of planning the use of capital and labor in a region, but relax the assumption that relocations of capital and labor are costless. We still have initially given quantities of capital and labor. Now, however, not only aggregates, but spatial distributions of these aggregates are also given. The future distributions can differ from these initial distributions in two ways. First, capital wears out and if it is not completely replaced by new equipment the stock of capital will change, whereas labor stock normally changes at the net reproduction rate. Second, labor and capital can actually be transferred in space by the application of transportation services.

The relevant assumptions should be stated more precisely. Suppose we consider only one produced commodity, and that this commodity can either be used as consumers' goods or, equally well, be invested as capital stock.

Capital stock wears out exponentially at a given depreciation rate. Accordingly, local production, minus local consumption, minus local capital depreciation, minus local net capital accumulation, is the quantity entered into the flow of capital goods, or, if negative, withdrawn from it. As we are focussing on capital flows and accumulation, we disregard flows of consumers' goods. If we wished to include them, there would be no difficulty in so doing because the model does not distinguish between consumers' goods and capital goods.

Labor stock accumulates at a given net reproduction rate. The quantity entered into the flow of labor or, if negative, withdrawn from it, is local labor reproduction, minus local accumulation of labor. Again, we disregard short-run phenomena such as commuting, and focus on migration and labor accumulation.

Production is thus determined by the local labor and capital stocks, per unit land area, or rather by what remains of them after the fixed-coefficient transportation technology has withdrawn what is needed for the transportation of capital goods and migrants.

5.2.2 Analysis

The goal function is now a utility index dependent on local consumption, aggregated over both space and time. Accordingly, we maximize

$$\iiint U(q)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}t \tag{76}$$

By introducing the space and time coordinates as arguments in the utility function we can account for spatial and temporal discounting. Of course, q denotes consumption. The rate of consumption and the utility index are continuous and differentiable functions of the space and time coordinates.

The production technology is again represented by a neoclassical production function

$$f(k, l) \tag{77}$$

where k is capital stock used in production of goods and l is labor stock used for the same purpose.

Transportation services are again produced by a Leontief technology of fixed coefficients. We assume that each unit of flow uses up κ units of capital and λ units of labor. In order to simplify the notation, we normalize the units of measurement of capital and labor so that transportation costs for one unit are the same for both flows.

Denoting the flow of capital by ϕ and the flow of labor by ψ, we know that $\kappa(|\phi| + |\psi|)$ units of capital and $\lambda(|\phi| + |\psi|)$ units of labor are withdrawn from the local stocks for production of transportation services. What remains is used in production of goods (for consumption and investment). Thus, denoting the local stocks of capital and labor by K and L, respectively, we have

$$k = K - \kappa(|\phi| + |\psi|) \tag{78}$$

and

$$l = L - \lambda(|\phi| + |\psi|) \tag{79}$$

Suppose capital wears out in proportion to the accumulated stock at the rate α.

We then have a replacement requirement of αK. Denoting net capital accumulation by $\dot{K}$, where the dot represents a derivative with respect to time, we see that the quantity $\alpha K + \dot{K}$ is withdrawn for investments. As the quantity q is withdrawn for consumption, the difference $f(k, l) - q - \alpha K - \dot{K}$ is added to the flow at each location or, if negative, withdrawn from it. Accordingly

$$\operatorname{div} \phi = f(k, l) - q - \alpha K - \dot{K} \tag{80}$$

For labor the stock increases at the net reproduction rate, denoted β. Thus, the local increase in labor due to reproduction is βL, whereas the local accumulation of labor is denoted $\dot{L}$. The difference enters the migration flow or, if negative, is withdrawn from it. Formally

$$\operatorname{div} \psi = \beta L - \dot{L} \tag{81}$$

The optimization is now a well-defined problem. We seek the maximum of (76) subject to the constraints (78)–(81). As a preliminary step we substitute for k and l from (78)–(79) into (80). In this way we dispose of two constraints and the two substituted variables. Only the constraints (80)–(81) remain (with the substitutions already made). We have to choose consumption q, capital stock K, labor stock L, and the flows of capital and migrants, ϕ and ψ, respectively. We want to find optimal function forms defined over space and time. In other words, this is a variational problem whose solution is obtained in terms of Euler equations. We associate Lagrangean multipliers p and w with (80) and (81), respectively. It should be noted that the Lagrangean multipliers are not constants, but change over space and time, due to the fact that the constraints are in local, not aggregate, form.

We can now state the Euler equations for optimality. For consumption, we obtain

$$U'(q) = p \tag{82}$$

For production, we obtain the two conditions

$$pf_k = \alpha p - \dot{p} \tag{83}$$

and

$$pf_l = -\beta w - \dot{w} \tag{84}$$

For transportation, we obtain

$$p(\kappa f_k + \lambda f_l) \frac{\phi}{|\phi|} = \operatorname{grad} p \tag{85}$$

and

$$p(\kappa f_k + \lambda f_l) \frac{\psi}{|\psi|} = \operatorname{grad} w \tag{86}$$

These conditions are arrived at as solutions to a planning problem. But again, it is not too difficult to find interpretations of the conditions in terms of market equilibrium. We note that the left-hand sides of (83)–(84) are marginal value productivities of capital and labor. Hence, we expect the right-hand sides to be input costs. Owing to the fact that the goods produced are also capital goods, p is also related to the price of capital stock. Compared to the static optimum conditions, we might be surprised to find, not the input prices, but their time derivatives (and reversed signs).

However, if we assume that the firms do not maximize their momentary profits, but rather their accumulated profits over a time interval, then it is obvious that a future decrease in input prices should be an incentive to postpone accumulation of capital stock. Thus the negative of the time derivative of price is a reasonable measure of temporary input cost. Likewise, capital depreciation and consequent capital replacement requirement are obvious cost items.

As for labor, net reproduction plays the same role as capital depreciation, but the sign is reversed. This may seem a bit odd at first, but in fact a natural increase in the local labor stock makes it possible to avoid wage raises in the future to attract more immigrants. The firm is to some extent protected by transportation costs from surrounding competitors. Any local surplus of labor may be assumed to emigrate only if the wage difference is greater than the migration cost. Likewise, in order to attract immigrants, the local wage rate ought to be higher than in the surroundings and again greater than migration costs.

So, in terms of dynamic optimum, the conditions (83)–(84) are not irrelevant in a market economy setting. These conditions are fairly close to those found in the recent theory of "microeconomic foundations of macroeconomics," where firms are supposed to plan their stocks of inputs by designing an appropriate dynamic price policy.

Conditions (85)–(86) are even easier to interpret in market equilibrium terms. Each unit of flow uses up κ units of capital and λ units of labor. The marginal productivities indicate the sacrifice in terms of goods not produced due to this withdrawal of inputs. The opportunity cost in terms of commodities is $(\kappa f_k + \lambda f_l)$. If we multiply by commodity price p, we obtain the monetary opportunity cost $p(\kappa f_k + \lambda f_l)$. This, of course, is the local cost of transportation, so that it is natural to find it in the left-hand sides of (85)–(86). Accordingly, these conditions again tell the familiar story that flows take the direction of the steepest price increases, and that the price increases in these directions equal transportation costs. Thus (85)–(86) are conditions of efficient trade and spatial equilibrium.

We can take squares of both sides of the vector equations (85)–(86) and equate them. The unit flow fields then multiply up to unit scalars and we obtain the equations

$$p^2(\kappa f_k + \lambda f_l)^2 = (\operatorname{grad} p)^2 = (\operatorname{grad} w)^2 \tag{87}$$

Next, we can substitute for the marginal productivities from (83)–(84), so that (87) becomes

$$(\alpha\kappa p - \beta\lambda w - \kappa\dot{p} - \lambda\dot{w})^2 = (\operatorname{grad} p)^2 = (\operatorname{grad} w)^2 \tag{88}$$

This is a pair of differential equations in the price of commodities and the wage rate. Through solving it, we establish the developments of prices and wages in space and over time.

By substituting the solutions for prices and wages into the right-hand sides of (83)–(84) we can then solve for the capital k, and labor l, used in production, Next, (77) gives the resulting output. But as capital and labor stocks used in production are known, we see that total capital and labor stocks, K and L, depend only on flow volumes. Accordingly, the right-hand sides of (80)–(81) depend only on flow volumes.

On the other hand, (85)–(86) indicate that the directions of the flows are gradient to price and wage. Knowing p and w, we also know the unit flow fields $\phi/|\phi|$ and $\psi/|\psi|$. But we also know that $\operatorname{div} \phi = \operatorname{grad} |\phi| \cdot \phi/|\phi| + |\phi| \operatorname{div}(\phi/|\phi|)$, and likewise for ψ. Thus, the left-hand sides of (80)–(81) also depend only on flow volumes and their gradients. This means that (80)–(81) supply us with another pair of partial differential equations. After solving them for flow volumes, we know all the variables of the model.

Thus, the differential equations (88) contain the initial information from which everything else can be calculated. As soon as we know prices and wage rates, we know the essential structural facts in terms of the flow lines and the corresponding potentials, including their development over time. This means that the solution to (88) is extremely interesting in the context of this model.

In fact, these equations are easy to discuss if we introduce an artifice to separate the spatial and temporal aspects of price–wage changes. Let us define a new scale for time and space by putting $t = \delta\tau$, $x_1 = \epsilon\xi_1$, and $x_2 = \epsilon\xi_2$. We note that this coordinate change does not distort space; it only introduces a linear change in scale. By letting ϵ approach zero the scale is magnified so that, in the limit, we are dealing with a single point. Likewise, by letting δ approach zero, the time scale is magnified so that in the limit we are dealing only with conditions at a certain point in time.

Let us now change system (88) so that we let the time derivatives be referred to the τ coordinate, whereas the gradients are stated in terms of the ξ_1, ξ_2 coordinates. As a result of this

$$(\alpha\kappa p - \beta\lambda w - \delta\kappa\dot{p} - \delta\lambda\dot{w})^2 = (\epsilon \operatorname{grad} p)^2 = (\epsilon \operatorname{grad} w)^2 \tag{89}$$

If we now let $\epsilon \to 0, \delta = 1$, then

$$\kappa\dot{p} + \kappa\dot{w} = \alpha\kappa p - \beta\lambda w \tag{90}$$

whereas if we let $\delta \to 0, \epsilon = 1$, we get

$$(\operatorname{grad} p)^2 = (\operatorname{grad} w)^2 = (\alpha\kappa p - \beta\lambda w)^2 \tag{91}$$

Equation (90) is a pair of dependent linear differential equations in p and w. It is very easy to solve because we have ordinary linear differential equations with constant coefficients. The only special fact to be taken into account is that, with respect to (90), one of the functions can be chosen arbitrarily. Likewise, equations (91) are easy to deal with in terms of the qualitative features of two-dimensional flows. These equations separate the spatial aspect so that we can study flow patterns, as in the stationary cases. In the same way, equation (90) separates the temporal aspects, so that we can study the price–wage dynamics in a point economy without spatial extension, as in traditional economic theory.

5.3 DIGRESSION ON THE WEBER AND von THÜNEN PRINCIPLES OF LOCATION AND LAND USE

5.3.1 Introduction

There are two ways of formulating the location problem. One, due to Weber (1909), is to ask: "Where should a certain activity be established?" The other, due to von Thünen (1826), is to ask: "Which activity is the best one to establish at a given location?" In Chapter 1 we suggested that the Weberian approach applies to cases where production is not land-consuming, whereas the von Thünen model is appropriate for cases of land-consuming activities. Suppose, however, that we wish to compare the performance of the two approaches in the case of land-consuming activities by putting the Weberian solution in a form where land rent is taken into account along with transportation costs.

We are going to establish a number of facts concerning the two alternative approaches. It will be demonstrated that the von Thünen problem emerges in a natural way from the question of how located scarce resources should be used in order that the total welfare of the inhabitants of an extended region be maximized. Accordingly, the social optimality of the Weberian solution depends on whether it coincides with the von Thünen solution or not. It is, therefore, interesting to discover that the Weberian solution is weaker. Its conditions are implied by the von Thünen conditions, but the reverse does not hold.

5.3.2 The Frame for Comparison

To compare the two approaches we need a setting sufficiently general to do justice to both. We note that both von Thünen and Weber are concerned with a set of potential production sites, that is, the continuous two-dimensional plane itself. Both have a located market place. Weber assumes a discrete set of located input supplies, whereas von Thünen deals with the input of land of constant fertility dispersed everywhere. Both seem to have a technology of constant coefficients in mind that allows one to proceed directly to the cost functions.

A continuous space model, where input supplies and output demand can be lumped in certain locations or more or less dispersed everywhere, and where production can take place anywhere, seems to provide a sufficiently general frame for the comparison (also cf. Section 6.1.3).

Instead of the classical fixed-coefficient technology we assume one with substitutability. For illustrative purposes we choose a Cobb–Douglas function. This choice is not crucial, as the analysis will work with any linearly homogeneous production function. But linear homogeneity is essential. Land is one of the inputs. By dividing through both sides of the production function by the input of land we wish to obtain the areal density of output as a well-defined function of the areal densities of the remaining inputs. This only works if the original production function is linearly homogeneous.

The question is one of whether linear homogeneity is too restrictive an assumption or not. The answer depends on the general purpose of the analysis. In spatial economics increasing returns are very often assumed. The purpose of this is apparently to explain agglomeration. Realistically speaking, if there were only decreasing returns no production would ever be established, and if there were no decreasing returns nothing would limit production once established. In practice, of course, the size of each producing unit is determined at the margin between increasing and decreasing returns.

Formally, with free entry, the production of each unit is established at the level of minimum average costs. Then, at any given set of prices, all changes of inputs and outputs are multiplicative due to the entry and exit of producing units. Once we are concerned, not with the individual firm, but with production conditions at a given location, linear homogeneity does not seem unreasonable.

5.3.3 Production

Assume a production function

$$Q = A \prod_{i=0}^{n} V_i^{\alpha_i} \tag{92}$$

where the V_i are inputs and Q is output. We reserve V_0 for land, but otherwise we do not specify the inputs. As a condition of linear homogeneity

$$\sum_{i=0}^{n} \alpha_i = 1 \tag{93}$$

We now divide through by V_0, obtaining

$$q = A \sum_{i=1}^{n} v_i^{\alpha_i} \tag{94}$$

The lowercase letters represent the areal densities of output and inputs. By definition we always have $v_0 = 1$.

5.3.4 Resources and Trade

The use of inputs for production v_i need not be equal to the local availabilities w_i. Local excess resources are transported to areas of local shortage. The flows of commodities transported are construed as vector fields

$$\phi^i = (\phi_1^i(x_1, x_2), \phi_2^i(x_1, x_2)) \tag{95}$$

that associate a flow vector with each location x_1, x_2. The unit tangents

$$\phi^i/|\phi^i| = (\cos\theta_i, \sin\theta_i) \tag{96}$$

indicate flow directions, and the Euclidean norms

$$|\phi^i| = \{(\phi_1^i)^2 + (\phi_2^i)^2\}^{1/2} \tag{97}$$

measure flow volumes. From Gauss's divergence theorem, the divergences of the flows, defined as

$$\operatorname{div} \phi^i = \partial\phi_1^i/\partial x_1 + \partial\phi_2^i/\partial x_2 \tag{98}$$

denote the local changes in the flows due to local sources, as in hydrodynamics. In our case the local sources are excess resources not used in local production, and so

$$\operatorname{div} \phi^i = w_i - v_i \tag{99}$$

This spatial transfer of resources is not costless. By assumption the costs of transportation are given in commodity terms. Hence, at each location a quantity $\lambda_i|\phi^i|$ of the produced good is used up in transportation of the ith input. The local transportation cost functions can be taken as spatial invariants, or as any given functions $\lambda_i(x_1, x_2)$ of the space coordinates. In the latter case the ratio of any two such functions should be a spatial invariant, provided the same transportation system is used for the different flows. To complete the picture, we should add that, land being immobile, there is no flow for $i = 0$. Accordingly, $w_0 = v_0 = 1$ everywhere.

As productive resources are moved from their original locations to the sites of production, the finished goods are moved from the production locations to the final consumers. Accordingly, there is a flow of produced commodities

$$\psi = (\psi_1(x_1, x_2), \psi_2(x_1, x_2)) \tag{100}$$

of direction

$$\psi/|\psi| = (\cos\tau, \sin\tau) \tag{101}$$

and volume

$$|\psi| = \{\psi_1^2 + \psi_2^2\}^{1/2} \tag{102}$$

The cost of shipment of produced commodities at each location is

$$\mu|\psi|$$

If we denote the final consumption by q', we can write the divergence–source relationship for the produced commodity

$$\operatorname{div} \psi = q - q' - \sum_{i=1}^{n} \lambda_i|\phi^i| - \mu|\psi| \tag{103}$$

The local change of flow is local output, minus local consumption, minus local transportation cost, in commodity terms for all the flows.

5.3.5 Social Optimum

We now only need an objective function in order to arrive at a well-defined optimization problem that can be handled by calculus of variations methods. To this end assume a local utility function

$$u(q', x_1, x_2) \tag{104}$$

that depends on local consumption. We also include the space coordinates explicitly so that we can put different weights on consumption at different locations, reflecting, for instance, different population densities.

We aggregate over locations and maximize

$$\iint_R u(q', x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{105}$$

subject to constraint (94) on available production technology and constraints (99) and (103) on resource availability and on the necessary consumption of resources for transportation.

We associate Lagrangean multipliers r_i with the constraints (99) and p with (103). Unlike the case in ordinary constrained optimization, the Lagrangean multipliers are not constant. They depend on location. This is because our constraints are in local, not integral, form. The notation indicates that the r_i and p are efficiency prices for inputs and outputs as is to be expected from duality in ordinary mathematical programming.

Maximization is with respect to input and output densities and flows, v_i, q, ϕ^i, and ψ. The Euler equations (necessary conditions for maximum welfare) are

$$\mathrm{d}u/\mathrm{d}q = p \tag{106}$$

$$\alpha_i q/v_i = r_i/p \qquad i = 1,2,\ldots,\mathrm{n} \tag{107}$$

$$p\lambda_i \phi^i/|\phi^i| = \operatorname{grad} r_i \qquad i = 1,2,\ldots,n \tag{108}$$

and

$$p\mu\psi/|\psi| = \operatorname{grad} p \tag{109}$$

Condition (106) states that the price of output be everywhere equal to local marginal utility. Its main interest lies in the fact that this is the only optimality condition that involves the utility function at all. Accordingly, the rest of the conditions, (107) for production, and (108)–(109) for trade and transportation, are efficiency conditions of Paretian type.

5.3.6 The von Thünen Conditions

Conditions (107) state that marginal value productivities of all the inputs, except land, equal their efficiency prices. Accordingly, they are conditions for optimal choice of production program by profit-maximizing firms at each location. In other words the conditions (107) provide an answer to the von Thünen question of which process to choose at any given location.

From (107), using (93), we also find that net profits per unit land area are

$$pq - \sum_{i=1}^{n} r_i v_i = \alpha_0 pq$$

These profits are due to the landowners as land rent. Denoting land rent by r_0, and recalling that, by definition, $v_0 = 1$, we can equate the preceding expression to $r_0 v_0$. Thus

$$\alpha_0 q/v_0 = r_0/p \tag{110}$$

This resembles (107), but it is not an optimum condition. It is a definition of land rent as net profits from the productive activity that is most profitable at the location concerned. Substituting from (110) and from (107) into the production function (94), we find a necessary condition among all the prices, including land rent

$$A \prod_{i=0}^{n} \alpha_i^{\alpha_i} p = \prod_{i=0}^{n} r_i^{\alpha_i} \tag{111}$$

5.3.7 Optimal Trade

Even though the conditions for optimum transportation and trade (108)–(109) are not our primary concern here, we should at least give them a brief interpretation. They tell us two things. Firstly, flows of inputs and outputs are in the directions of the gradients of their efficiency prices. As the latter express local scarcities, the conditions tell us that all resources are shipped in the directions where their valuations increase most. Secondly, the conditions tell us that in these directions the prices increase at the rate of transportation costs. As conditions for efficient trade and transportation these clearly have an intuitive appeal.

We can now see how the picture of production and trade fits together. Taking the gradient of both sides of (111)

$$\operatorname{grad} p/p = \sum_{i=0}^{n} \alpha_i \operatorname{grad} r_i/r_i$$

If we now substitute from (108)–(109) and use (96) and (101) we get

$$\mu = \sum_{i=0}^{n} \alpha_i \lambda_i (p/r_i) \cos(\theta_i - \tau) \tag{112}$$

To be exact, we have not yet defined θ_0 and λ_0, but if we let the former be the direction of steepest increase of land rent and the latter equal the ratio of the rate of increase in this direction to product price, then (112) indicates the limits for the spatial variation of land rent.

5.3.8 The Weber Condition

Returning to our main concern, we conclude that so far we have established that optimal resource allocation in space entails the von Thünen conditions, whereas there is no obvious relation to the Weberian approach.

Let us therefore examine how the von Thünen and Weber problems are related to each other. To this end we first note that the relation

$$pq - \sum_{i=0}^{n} r_i v_i = 0 \tag{113}$$

always holds, irrespective of any optimality, once we assume that net profits are due to landowners as rent. Taking the gradient of this identity we get

$$\left(q \operatorname{grad} p - \sum_{i=0}^{n} v_i \operatorname{grad} r_i\right) + \left(p \operatorname{grad} q - \sum_{i=0}^{n} r_i \operatorname{grad} v_i\right) = 0 \tag{114}$$

However, from (94), for any production activity that satisfies the assumed technology, we obtain

$$\operatorname{grad} q = \sum_{i=0}^{n} \alpha_i (q/v_i) \operatorname{grad} v_i \tag{115}$$

Strictly, the right-hand side can be derived only with the lower summation limit 1. As $v_0 = 1$ and hence grad $v_0 = 0$, the lower limit can be extended to 0.

We next substitute from (115) into (114) and obtain

$$\left(q \operatorname{grad} p - \sum_{i=0}^{n} v_i \operatorname{grad} r_i\right) + \sum_{i=0}^{n} (\alpha_i pq - r_i v_i) \operatorname{grad} \ln v_i = 0 \tag{116}$$

This condition allows us to compare the Weber and von Thünen approaches. The first term in parentheses in (116) is an expression for the spatial rate of change of profits as we move the given production process $(q, v_1, v_2, \ldots, v_n)$ to another location. Only when this term vanishes is there no incentive to relocate the process. Accordingly, we can interpret

$$\left(q \operatorname{grad} p - \sum_{i=0}^{n} v_i \operatorname{grad} r_i\right) = 0 \tag{117}$$

as the Weber condition for optimal location. In order to understand this, we recall from conditions (108)–(109) for optimum trade and transportation that the spatial rate of change of input and output prices equals the cost of transportation. This cost is measured along the minimum-cost routes that solve the differential equations (108)–(109). Should transportation cost be spatially invariant, then these routes are straight lines. In the case where inputs are available at certain specified locations only and the output has to be sold at another specified location, we see that (117) then expresses exactly the balance of forces in the classical Weberian location optimum.

There is still one difference. Weber does not include rent as a cost for *not* using land for some alternative purpose. However, it seems fairer to the Weberian approach to allow for a variable land rent in the model. We obtain the Weber case exactly if land rent is assumed to be spatially invariant.

5.3.9 von Thünen Implies Weber

We have finally interpreted both Weber and von Thünen within the framework of a single model. The von Thünen conditions are (107), together with definition (110), whereas the Weber condition is (117). Recall that (116) holds everywhere, irrespective of any optimality. It hence applies independently of which conditions we use. If we now substitute from (107) and (110) into the second summation of (116), we find that it vanishes and hence that (117) is implied. We therefore conclude that *the von Thünen conditions for the choice of an optimal process at each location imply the Weber conditions for the choice of an optimal location for each process.*

5.3.10 Weber Does Not Imply von Thünen

Does the reverse hold too? To answer this question we construct a counter-example where (117) holds, so that the second summation in (116) is also equal to zero, but where the parenthetic term within the summation does not vanish. Suppose all the v_i, except that for land, v_0, change in proportion everywhere. Accordingly, all the logarithmic gradients are equal, except of course that for v_0, which vanishes. Next, choose any set of constants β, such that

$$\sum_{i=0}^{n} \beta_i = 1 \tag{118}$$

but

$$\beta_i \neq \alpha_i \qquad \text{for some } i \tag{119}$$

Then put

$$\beta_i q / v_i = r_i / p \qquad i = 0, 1, \ldots, n \tag{120}$$

Substitution from (118) and (120) into the second summation of (116) makes it vanish, and hence (117) does hold. However the conditions (107), due to (119) and (120), do not hold.

Hence, the Weber condition for optimum location of each process is strictly weaker than the von Thünen condition for optimal choice of process at each location. As the von Thünen conditions correspond to social optimum in the spatial allocation of available resources, we conclude that the Weber condition is too weak to ensure this.

5.3.11 Conclusion

Sometimes one encounters in spatial economics the combined question: *which* activity should be established, and *where* should it be located? As we have seen, the first question alone, asked for each separate location, works sufficiently well as an organizing principle for the spatial economy. Combining the questions in the way indicated, we only get the answer that the best thing to do is to establish the most profitable activity at the location where land rent is highest. As land, if anything, is immobile, this cannot work as an organizing principle for extended space.

6 *Some Long-Run Location Theory*

6.1 CLASSICAL PROBLEMS

6.1.1 The Procurement Problem

Classical location theory is a long-run theory; its main purpose is to explain where facilities are located and where various economic activities take place on a permanent basis. It requires a long period for plants to be located and for an industry to reach locational equilibrium. Moreover, only in the long run can products be offered at constant unit cost for any quantity in a given location – the short or medium run would always impose capacity limitations on output. But constant unit cost has so far been the standard case in location theory.

As a starting point we consider the seemingly simple problem of procuring at minimum total cost quantities $q(\underline{x})$ of a commodity for consumption when this commodity can be produced without limit at constant unit costs $c(\underline{x})$ in various locations $\underline{x}$. The result is a natural extension of the continuous transportation problem in which production cost was uniform among locations. Now, however, it is necessary to consider the sum of both transportation and production costs; for total production cost is no longer simply proportional to total output when production cost depends on location.

Let production per unit area, i.e. production intensity, be denoted by $z = z(\underline{x})$. The divergence law then assumes the form

$$\operatorname{div} \phi = z - q \tag{1}$$

or stated in words, flow divergence equals local production minus consumption. Consider now the problem of minimizing total costs

$$\min_{\phi,\, z} \int\!\!\int [k\,|\phi| + cz]\, \mathrm{d}x_1\, \mathrm{d}x_2$$

subject to (1). As before we obtain the gradient law

$$\operatorname{grad} \lambda = k\frac{\phi}{|\phi|} \qquad \phi \neq 0 \tag{2}$$

The efficiency conditions for z are obtained as in a discrete linear program

$$\left.\begin{aligned} z &= 0 \qquad \text{for} \quad \lambda < c \\ z &\geqslant 0 \qquad \text{for} \quad \lambda = c \end{aligned}\right\} \tag{3}$$

Thus production takes place where and only where the product value equals the cost of production. In no case can production cost be less than the product's value: that would imply unbounded profit opportunity, which is inconsistent with competitive equilibrium.

When production takes place at constant unit cost $c(\underline{x})$, subject to no capacity constraints, then the potential function λ is independent of the consumption program. It depends in fact only on production locations, and more precisely on the spatial pattern $c(\underline{x})$ of production cost. This case has been implicit in much of the earlier location theory, which is a theory of the supply side only. This is transparent in the special case when production costs do not vary very much

$$|\text{grad}\, c| \leqslant k \tag{4}$$

In this case the solution is in fact

$$\lambda \equiv c \tag{5}$$

and production is for local use only. There is no integrated spatial market, but only a set of independent local markets where the product price equals local production cost.

It has been assumed so far that production is possible in all locations. The impossibility of production could be signalled by a very large value of c. At the transition from possible to impossible production locations the gradient condition (4) would be violated, so that flows may now occur.

To see that the price structure λ is independent of the consumption program in general, we consider the dual problem. The Lagrangean integral of the primal problem is

$$\begin{aligned}\iint L \,\mathrm{d}x_1\,\mathrm{d}x_2 &= \iint -k|\phi| - cz + \lambda(z - q - \text{div}\,\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2\\ &= \iint -k|\phi| + (\lambda - c)z - \text{div}(\lambda\phi) - \lambda q + \phi\,\text{grad}\,\lambda\,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

By the Gauss integral theorm

$$-\iint \text{div}(\lambda\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2 = -\int \lambda\phi_n\,\mathrm{d}s$$

and this vanishes by the boundary condition. Now

$$\begin{aligned}\iint L\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \iint -\lambda q + (\lambda - c)z + \phi\cdot\left(\text{grad}\,\lambda - k\frac{\phi}{|\phi|}\right)\mathrm{d}x_1\,\mathrm{d}x_2\\ &\leqslant -\iint \lambda q\,\mathrm{d}x_1\,\mathrm{d}x_2\end{aligned}$$

for all $\lambda \leqslant c$ and all

$$|\text{grad}\,\lambda| \leqslant k$$

The dual problem is therefore

$$\max_{\lambda} \iint \lambda q \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

subject to

$$\lambda \leqslant c \tag{6}$$

and

$$|\mathrm{grad}\, \lambda| \leqslant k \tag{7}$$

Constraints (6) and (7) do not contain q and the efficiency condition (1) does not contain λ. We therefore conclude that λ is independent of q. The intuitive reason is that each location can procure its supply from the cheapest source, regardless of what other locations may demand, for costs, and hence prices, are independent of quantities.

When condition (4) is violated a singular solution will result such that the demand of a two-dimensional region is produced in a one-dimensional set of locations. To illustrate this consider a circular region in which production cost depends on distance r from the center

$$c = c(r)$$

When

$$\frac{\mathrm{d}c}{\mathrm{d}r} < -k$$

unit production cost falls more rapidly than the rate of transportation cost. It is then economical to shift all production to the perimeter of the circle. The areal density of production on the circular perimeter is then infinite. Conversely, when

$$\frac{\mathrm{d}c}{\mathrm{d}r} > k$$

production is most economical at the center, once more resulting in an infinite production density, but this time in a point location. The procurement problem may thus involve singular solutions. Production will no longer be dispersed at finite areal densities but will be localized along lines or in isolated points. Such lines have been referred to as belts of production by Lösch (1940).

A standard problem in location theory concerns the location of production activities that require as inputs no localized resources. The necessary raw materials (known as ubiquities) are assumed to be available everywhere at zero or constant cost. Labor is always required. When wages are uniform, the only locational factor is distance from points of consumption. This distance is minimized when production takes place right at the point of consumption. In this connection one speaks of "market orientation." This concept is usually extended to the case where production is localized, i.e. it is concentrated in a few points inside the consumption area in order to reap economies of scale (see Section 6.2.4).

The only factor that can cause the processing of ubiquities to be located away from consumption points is cheap labor. Differences in wages rates must exceed differences in commodity price, i.e. the levels of transportation costs for commodities. This is possible when the cost of migration exceeds the costs of transporting goods (see below). As Adam Smith once remarked (but this was more true in his time than at present), "After all that has been said of the levity and inconstancy of human nature, it appears evidently from experience that a man is of all luggage the most difficult to be transported" (Adam Smith, *The Wealth of Nations*, ed. Edwin Cannan, Methuen, London, 1950, p. 84).

It is now apparent that this so-called "labor orientation" phenomenon is a special case of the procurement problem that is obtained when production costs are equal to a constant plus labor costs. It is also assumed in labor orientation that labor is available in unlimited amounts at wage rates depending only on the location. The model thus describes the location of a labor-oriented industry. It is clearly not market oriented since prices are determined independently of demand. Of course the locational distribution of production levels z will depend on the locational distribution of demand q, while the spatial pattern of prices is independent of demand.

6.1.2 One Mobile Input: Resource versus Market Orientation

Now let production involve one mobile input, typically a material. This input is found in so-called resource deposits at certain locations. The question is, will the material be processed at the site of the deposits, at the market, or in transit?

We may assume that production transforms one unit of resource into one unit of the product. Demand is for products only. The resource can be extracted at constant unit costs $c(\underline{x})$ in unlimited amounts. (For problems associated with the exhaustion of deposits see Section 6.2.1.) Costs of labor or other ubiquitous inputs are assumed constant and can be ignored. Labeling product as 1 and resource as 0, we have the problem formulation

$$\min_{z_0, \phi_0, \phi_1} \iint c z_0 + k_0 |\phi_0| + k_1 |\phi_1| \mathrm{d}x_1 \, \mathrm{d}x_2$$

subject to

$$\operatorname{div} \phi_1 = z_1 - q_1 \tag{8}$$

$$\operatorname{div} \phi_0 = z_0 - z_1 \tag{9}$$

Consider the Lagrangean integral

$$\iint -c z_0 - k_0 |\phi_0| - k_1 |\phi_1| + \lambda(z_1 - q_1 - \operatorname{div} \phi_1) + \mu(z_0 - z_1 - \operatorname{div} \phi_0) \mathrm{d}x_1 \, \mathrm{d}x_2$$

The efficiency conditions are

$$\operatorname{grad} \lambda = k_1 \frac{\phi_1}{|\phi_1|} \qquad \phi_1 \neq 0 \tag{10}$$

$$\operatorname{grad} \mu = k_0 \frac{\phi_0}{|\phi_0|} \qquad \phi_0 \neq 0 \tag{11}$$

$$\left.\begin{array}{lll} z_0 = 0 & \text{wherever} & \mu < c \\ z_0 \geqslant 0 & \text{wherever} & \mu = c \end{array}\right\} \tag{12}$$

$$\left.\begin{array}{lll} z_1 = 0 & \text{wherever} & \lambda < \mu \\ z_1 \geqslant 0 & \text{wherever} & \lambda = \mu \end{array}\right\} \tag{13}$$

Suppose first that k_0 and k_1 are constant everywhere and that both resource and product move along the same paths from deposit to consumption location. Then the price of the good with the higher transportation cost must rise faster. If this is the product, then its value relative to the resource is lower at the deposit than at the consumption location. Hence the last equation, (13), implies that production takes place at consumption sites and that only the resource is transported. The opposite holds when the resource is costlier to ship. We have therefore the following principle governing market versus resource orientation: *Processing is market oriented when the resource is easier to transport, and is resource oriented when the product is easier to transport.*

Suppose now that there is a discontinuous change in the levels of k_0 and k_1 such that in some region $k_0 < k_1$ while in the adjacent region $k_1 < k_0$. This may be due to a change of mode of transportation, say from water to rail transportation. Then production will be attracted to the switch point and so-called "processing in transit" will then take place (Hoover 1948, pp. 38–40).

6.1.3 The Launhardt–Weber Model

It is instructive to examine how the classical Weber problem (actually treated by Launhardt before Weber), which is the epitome of the discrete-point approach, appears when treated in a continuous framework. We consider an industry using two material inputs in fixed proportion to produce a single mobile output for a single (point) market. (On this topic see also Section 5.3.)

A particularly interesting case is that in which the sum of the weights of any two of these three commodities exceeds the weight of the third. (In modern steel production, for example, the weights of coal, ore, and steel produced are almost equal, so that this would meet the requirements.) For then it is possible – depending on the location of the input sources and the output market – that the optimal location of the processing plant falls into neither source location nor market, but in between. As is shown in standard works on location theory, the condition that determines the production location is then that the three weights, "pulling" in their respective directions, should constitute an equilibrium of forces (Beckmann 1968, p. 16). We assume, of course, that the given locations and the given weights permit this type of solution.

To establish a correspondence between the classical Weber problem and a continuous model, we interpret the raw material sources as small regions of limited area. In the case of mineral deposits this can only add realism to the model. Also, the market is not a point but a small region. One way to introduce the source locations into the model is by means of a resource cost function that has a finite (possibly zero) value c in the deposit area and is infinite (or finite and very large) outside. Market demand is described as usual by $q(\underline{x})$. Let the subscripts 0, 1, 2 refer to the product and to the two inputs, respectively. The object is to

$$\min_{\substack{\phi \\ z_i \geqslant 0}} \iint \sum_{i=0}^{2} [k_i|\phi_i| + c_i z_i] \, dx_1 \, dx_2$$

subject to

$$\operatorname{div} \phi_0 = z_0 - q \tag{14}$$

$$\operatorname{div} \phi_i = z_i - b_i z_0 \qquad i = 1,2 \tag{15}$$

The first constraint states that net outflow of product equals production minus consumption. The second condition states that net outflow of a material input equals the amount mined z_i minus the amount used in production, where b_i is the input coefficient.

Consider the Lagrangean

$$L = -\sum_{i=0}^{2} [k_i|\phi_i| + c_i z_i] + \sum_{i=1}^{2} \lambda_i(z_i - b_i z_0 - \operatorname{div} \phi_i) + \lambda_0 (z_0 - q - \operatorname{div} \phi_0) \tag{16}$$

The efficiency conditions are then

$$k_i \frac{\phi_i}{|\phi_i|} = \operatorname{grad} \lambda_i \qquad \phi_i \neq 0 \qquad i = 0,1,2 \tag{10}$$

$$\left.\begin{aligned} z_0 &= 0 && \text{wherever} \quad \lambda_0 < \sum_i b_i \lambda_i + c_0 \\ z_0 &\geqslant 0 && \text{wherever} \quad \lambda_0 = \sum_i b_i \lambda_i + c_0 \end{aligned}\right\} \tag{17}$$

$$\left.\begin{aligned} z_i &= 0 && \text{wherever} \quad \lambda_i < c_i \\ z_i &\geqslant 0 && \text{wherever} \quad \lambda_i = c_i \end{aligned}\right\} \tag{18}$$

Condition (17) requires us to find the production location where the output value equals the sum of the input values, while everywhere else it must be less. Clearly at a production location all three flows must be nonvanishing. What is the change in net profits that results when a production location is moved in the direction ν? This is given by the directional derivative D_ν of the profit function

$$D_\nu \left(\lambda_0 - \sum_{i=1}^{2} b_i \lambda_i - c_0\right) = \nu \cdot \operatorname{grad} \left(\lambda_0 - \sum_{i=1}^{2} b_i \lambda_i - c_0\right) \tag{19}$$

$$= \nu \cdot \left(k_0 \frac{\phi_0}{|\phi_0|} - \sum_{i=1}^{2} k_i b_i \frac{\phi_i}{|\phi_i|}\right) \tag{20}$$

using (10) and observing that costs c_i are constant.

At a profit maximum this directional derivative must be nonpositive for all ν. It follows that the vector in parentheses must vanish. Hence

$$k_0 \frac{\phi_0}{|\phi_0|} = \sum_{i=1}^{2} k_i b_i \frac{\phi_i}{|\phi_i|} \tag{21}$$

But this is precisely the condition that there is an equilibrium of forces between the weights of the resources pulling with strength $k_i b_i$ in the directions of the resource locations and the pull k_0 of the product in the direction of a consumption location.

We may also investigate the set of production locations that emerges when demand is widely dispersed across the entire region. For given resource locations, the locus of optimal production points is determined as follows. It is a circular arc through the two resource deposits such that the two rays from the resource deposits form a constant angle with the direction of flow to the market location. All markets lying on such a ray are served by the same production location. The angles are determined in the same way as in the previous particular solution for a single market point, because they depend only on the relative weights that must be moved. In this way all points lying between lines A and B (Figure 6.1) will be served by production locations lying on this circular arc. The same applies for market locations lying in the opposite directions between lines D and C. The circular arcs containing all production locations are symmetric. All markets lying between lines A and D are served by a production location at resource deposit 1, while all market locations lying between lines B and C are served by a production location at resource deposit 2.

All markets that lie on the same flow line of the product are therefore served by the same production location. The production density on the two arcs and in the two resource deposits will therefore be infinite again: once more the solution becomes singular. The interior of the area spanned by the two circular arcs contains locations of production for local consumption: production and market location coincide. Here the areal density is finite.

Generalizations to more than two inputs or to alternative sources of supply will not be developed here. The point was simply to show how a continuous flow model can handle the classical Weber problem, and in fact extend it.

6.1.4 Indivisibilities

The purest case of allocation of indivisible resources in spatial economics is the following. Suppose that production requires no land input (or rather a negligible amount) but requires one unit of an indivisible resource. With one unit of this indivisible resource any quantity of the product may be produced for export from one location, the location being considered as a single, dimensionless point. How many units of this indivisible resource should be used and where should they be located?

Consider first how to locate a single facility (or unit of the indivisible resource). The

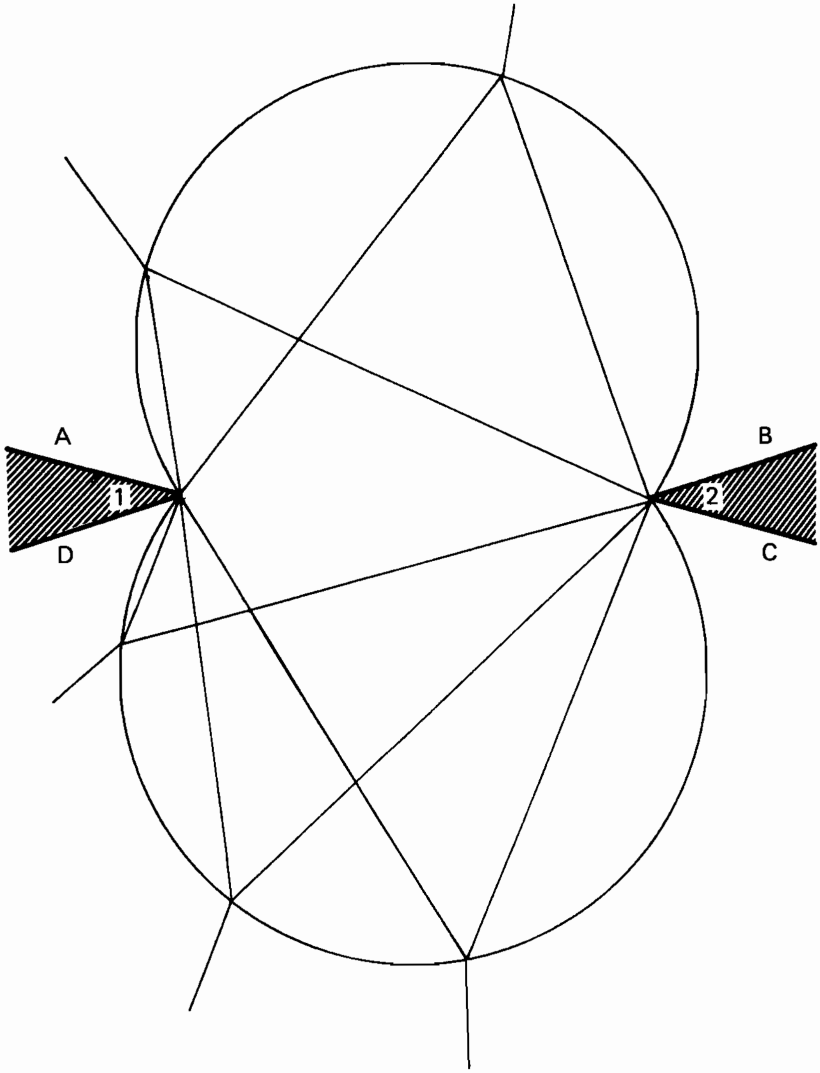

Figure 6.1. Solution of the Launhardt–Weber problem for different market locations.

object is to minimize total transportation cost. Formally the problem is the continuous transportation model of Section 2.1, with this modification

$$q(x_1, x_2) \geqslant 0 \qquad \text{throughout } A$$

A single singular source, represented by a point (x_1^0, x_2^0) surrounded by a circle of infinitesimal radius, may still be chosen, as shown in Figure 6.2. The problem is now as follows

$$\min_{\underline{x}^0,\, \phi} \iint k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

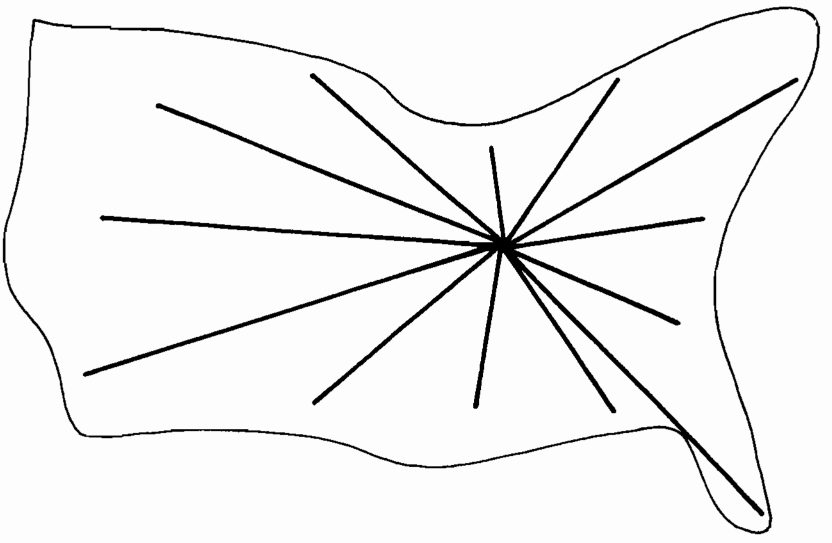

Figure 6.2. Location of a single facility.

subject to

$$\operatorname{div} \phi + q = 0 \qquad \text{in } A \tag{22}$$

$$\phi_n = 0 \qquad \text{on the outer boundary of } A \tag{23}$$

$$\int_{\underline{x}^0} \phi_n \, ds = \iint_A -q \, dx_1 dx_2 \qquad \text{at the singular point } \underline{x}^0 \tag{24}$$

Consider its dual, as discussed in Section 2.2.9. In the present problem it assumes the form

$$\min_{\underline{x}^0} \max_{\lambda} \iint \lambda \, q \, dx_1 dx_2 \tag{25}$$

subject to

$$|\operatorname{grad} \lambda| \leqslant k \tag{26}$$

Assume in particular that

$$k(\underline{x}) \equiv k \tag{27}$$

so that the flow lines are straight lines. Let $r(\underline{x}, \underline{x}^0)$ denote the straight-line distance between $\underline{x}$ and $\underline{x}^0$. Now A is punctured at $\underline{x}^0$ and therefore constraint (22) does not apply there. Its Lagrangean multiplier must be zero. Thus

$$\lambda(\underline{x}^0) = 0 \tag{28}$$

Then

$$\lambda(\underline{x}) = k \cdot r(\underline{x}, \underline{x}^0) \tag{29}$$

The dual becomes

$$\min_{\underline{x}^0} \int\int r(\underline{x}, \underline{x}^0) q \,\mathrm{d}x_1 \mathrm{d}x_2 \tag{30}$$

in agreement with the ordinary formulation of the facility location problem when costs other than transportation are constant.

An interesting conclusion may now be drawn from the primal formulation. Since $k(\underline{x}) \equiv k$ it may be set equal to unity. The flow field consists of the radius vectors issuing in all directions from $\underline{x}^0$. Suppose that flow vectors at $\underline{x}^0$ are replaced by strings to which weights are attached equal in strength to the flow vectors. Now the combined sum of weights times path lengths will be a minimum if and only if these vectors, considered as forces, are in equilibrium at $\underline{x}^0$. This is a continuous analog of the Varignon machine for the Weber problem (Beckmann 1968, p. 18). Hence the optimum facility location $\underline{x}^0$ must represent an equilibrium of flow when flows are interpreted as forces.

With the Manhattan metric this becomes the *Principle of the Median*, for it must now be true in the North–South and East–West directions separately: as many customers must pull from the North as from the South, and as many from the East as from the West. In fact, this principle still applies to each of several facilities separately, but the visualization and implementation of the principle become more difficult.

The solution of the indivisibility problem is well understood in the case of a homogeneous unbounded two-dimensional plane. It can then be reduced to finding that spacing of points which minimizes average costs per unit area when the market area surrounding any one point is considered. It is not hard to see that these market areas constitute a regular net of hexagons. (Cf. Section 4.7.2.) (When demand depends on price it is not obvious that deliveries will be made to all points of the hexagon – corners may be rounded off in the case of linear demand curves (Beckmann 1971). This is the classic Löschian equilibrium problem.

6.2 SOME RECENT DEVELOPMENTS

6.2.1 Optimum Utilization of Exhaustible Resources

The following is essentially based on Beckmann (1982b). At location (x_1, x_2) let there be a total $A(x_1, x_2)$ of the resource in the ground, but let the maximum feasible rate of extraction be $a(x_1, x_2)$. Thus if $z(x_1, x_2, t)$ is the rate of extraction at location (x_1, x_2) at time t one has two restrictions

$$z \leqslant a \tag{31}$$

$$\int_0^t z(s)\,\mathrm{d}s \leqslant A \tag{32}$$

With a planning horizon T one may restrict oneself to $t \leqslant T$. The technology assumed is one of fixed coefficients. Thus the cost of extraction should depend only on location

$$c = c(x_1, x_2) \tag{33}$$

Presumably, the exhaustible resource will be converted at some point into consumable products, which then generate utility. We may sidestep this conversion and feed the resource directly into the utility functions

$$u = u(x_1, x_2, q) \tag{34}$$

where q is the rate of consumption (in the form of final products) of the exhaustible resource at location x_1, x_2.

The cost of transportation will be treated as exogenous. A transportation cost

$$k(x_1, x_2)|\phi(x_1, x_2)| \tag{35}$$

is associated with a flow ϕ to location x_1, x_2. The welfare function to be optimized is thus a sum of utilities minus costs. Therefore utility must also be measured in money units

$$\begin{aligned} W = \int_0^T \int_A \int & u(x_1, x_2, q(x_1, x_2)) - k(x_1, x_2)|\phi(x_1, x_2)| \\ & - c(x_1, x_2)\, z(x_1, x_2)\, dx_1\, dx_2\, dt \end{aligned} \tag{36}$$

In addition to the constraints on resource extraction we have the source–sink equation

$$\operatorname{div} \phi(x_1, x_2) = z(x_1, x_2) - q(x_1, x_2) \tag{37}$$

Consider the Lagrangean

$$\begin{aligned} & \int_0^T \int_A \int u - k|\phi| - cz + \lambda [z - q - \operatorname{div} \phi] + \mu [a - z] \\ & + \nu(t) \left[A - \int_0^t z(s)\, ds \right] dx_1\, dx_2\, dt = \iiint L\, dx_1\, dx_2\, dt \text{ (say)} \end{aligned} \tag{38}$$

where all variables except A and a depend on location (x_1, x_2) and on time t. For simplicity, we have not discounted future utilities and costs. The efficiency conditions are as follows

$$\left. \begin{aligned} q(x_1, x_2) = 0 & \longleftrightarrow \frac{\partial u}{\partial q} < \lambda(x_1, x_2) \\ q(x_1, x_2) \geqslant 0 & \longleftrightarrow \frac{\partial u}{\partial q} = \lambda(x_1, x_2) \end{aligned} \right\} \tag{39}$$

Obviously λ is the price of the resource at location (x_1, x_2) and time t. (If the resource is not shipped directly but only after some transformation, then λ is the value of the resource content of this transformed good at location (x_1, x_2) and time t.) The condition states that consumption should not take place if marginal utility, even at zero consumption, falls short of the resource price. Otherwise marginal utility should be made equal to price. With Cobb–Douglas or logarithmic utility functions, where marginal utility is infinite at zero consumption, some of the resource must be consumed everywhere and at all times. Next

$$\left.\begin{aligned} z = 0 &\longleftrightarrow \lambda < c + \mu + \int_t^T \nu(s)\,\mathrm{d}s \\ z \geqslant 0 &\longleftrightarrow \lambda = c + \mu + \int_t^T \nu(s)\,\mathrm{d}s \end{aligned}\right\} \tag{40}$$

This condition specifies where extraction should take place at a given time. It should not take place where the market price λ falls short of extraction cost c plus capacity rent μ plus exhaustion rents

$$\int^T \nu(s)\,\mathrm{d}s$$

Notice that

$$\left.\begin{aligned} \mu = 0 &\longleftrightarrow z < a \\ \mu \geqslant 0 &\longleftrightarrow z = a \end{aligned}\right\} \tag{41}$$

and

$$\left.\begin{aligned} \nu = 0 &\longleftrightarrow \int_0^t z\,\mathrm{d}s < A \\ \nu \geqslant 0 &\longleftrightarrow \int_0^t z\,\mathrm{d}s = A \end{aligned}\right\} \tag{42}$$

Thus rent μ is incurred only when the location is worked to capacity a, and the rent ν falls due only after exhaustion of the site. Finally

$$\left.\begin{aligned} k\frac{\phi}{|\phi|} &= \operatorname{grad}\lambda \qquad \text{where} \quad \phi \neq 0 \\ k &\geqslant |\operatorname{grad}\lambda| \qquad \text{where} \quad \phi = 0 \end{aligned}\right\} \tag{43}$$

Equation (43) states the familiar equilibrium condition for interlocal trade in a two-dimensional spatial market.

This general model will now be illustrated in terms of a simple one-dimensional example. Let consumption be concentrated in a single location, placed at point zero, and let the resource be available in an interval extending from r_0 rightwards to infinity (see Figure 6.3). Further, let

$$a(r) = a$$

$$A(r) = A$$

the extraction cost

$$c(r) = c$$

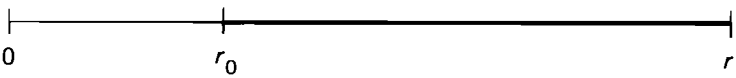

Figure 6.3. Location of an exhaustible resource.

and the transportation cost

$$k(r) = k$$

all be constant. Assume the utility function to be logarithmic

$$u = \log q \tag{44}$$

The welfare function is then

$$\int_0^T \int_0^\infty [\log q - k|\phi| - cz] \, dr dt \tag{45}$$

Let $r_1(t)$ be the largest distance from which the resource is shipped to the consumption point initially. Then (40) implies

$$\lambda(r_1, t) = c$$

and (43) implies

$$\lambda(0, t) = kr_1 + c$$

From (39)

$$c + kr_1 = \lambda(0, t) = \frac{\partial u}{\partial p} = \frac{1}{q}$$

so that

$$q = \frac{1}{c + kr_1} \tag{46}$$

Since the marginal utility of consumption equals the marginal cost of production, the rate of consumption thus depends on the distance from which the resource must be brought, and as this distance increases with time, the rate of consumption necessarily falls.

Consider now production. At time zero the resource must be extracted in an interval (r_0, r_1) such that supply equals demand

$$q = \frac{1}{c + kr_1} = a \cdot (r_1 - r_0) \tag{47}$$

From this

$$r_1 = \frac{1}{2}\left(r_0 - \frac{c}{k}\right) + \left[\frac{1}{4}\left(r_0 - \frac{c}{k}\right)^2 + \frac{1}{ak} + \frac{c}{k} r_0\right]^{1/2} \tag{48}$$

The initial period during which resource extraction takes place in the interval (r_0, r_1)

lasts A/a units of time. After that, the operation shifts to an interval (r_1, r_2). Clearly, each successive termination point r_{n+1} may be calculated from the previous one r_n in the same manner, for (46) and (47) must again apply, yielding

$$r_{n+1} = \frac{1}{2}\left(r_n - \frac{c}{k}\right) + \left[\frac{1}{4}\left(r_n - \frac{c}{k}\right)^2 + \frac{1}{ak} + \frac{c}{k}r_n\right]^{1/2} \tag{49}$$

Although the setting of this problem is essentially continuous, the solution consists of a sequence of discrete shifts of operations.

This result clearly also applies in a two-dimensional context. Thus the interval can be replaced by a succession of concentric rings surrounding a von Thünen city as the mining operations are successively shifted farther away. Each interval of operations has the same duration A/a. But the discrete nature of the selection of mining sites still applies and continues to apply, even when the resource does not occur in all places beyond distance r_0 but only in selected areas, and also when consumption is not concentrated in one place but is dispersed in a continuous manner. The details must be worked out in every specific case, but the qualitative nature of the solution applies in general.

6.2.2 Water Resource Management

The continuous model of transportation offers a natural framework for modeling the production, distribution, and consumption of water in a region. Such regions need not be self-sufficient – as a watershed is – but may be importers and/or exporters of water to neighboring regions or to the ocean.

In the absence of man-made water channels or pumping activities, etc., there is the natural water flow ϕ^0 generated by the naturally existing source–sink distributions q^0 and satisfying

$$\operatorname{div}\phi^0 + q^0 = 0 \qquad \text{in} \quad A \tag{50}$$

$$\phi_n^0 = g^0 \qquad \text{on} \quad \partial A \tag{51}$$

Any change through the production z or consumption q of water will lead to a superimposed flow pattern ψ such that

$$z - q = \operatorname{div}\psi \qquad \text{in} \quad A \tag{52}$$

$$\psi_n = b \qquad \text{on} \quad \partial A \tag{53}$$

where b is the desired change in net exports of water. Actual physical water flow will then be

$$\phi = \phi^0 + \psi \tag{54}$$

but the cost of moving water will be a function only of the additional superimposed flow ψ. The overall objective may be, for example, to minimize production and moving costs

$$\min \iint c(x_1, x_2)z + k(\psi, x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{55}$$

subject to (52) and (53). For the program z, q, b to be feasible, one must have

$$\iint_A (-q_0 + z - q)\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\partial A} (g + b)\mathrm{d}s \tag{56}$$

or, since the natural water flow may be assumed as feasible

$$\iint_A (z - q)\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\partial A} b\ \mathrm{d}s \tag{57}$$

A broader objective would be to maximize the net utility of water consumption

$$\max \iint u(q, x_1, x_2) - c(x_1, x_2, z) - k(\psi, x_1, x_2)\mathrm{d}x_1\,\mathrm{d}x_2 \tag{58}$$

subject to (52) and (53).

The cost function k will not be isotropic. Rather the costs of pumping water in the direction opposite to the natural flow direction – the topographical gradient – will be a maximum, while the cost will be a minimum and may even be zero in the direction of the topographical gradient.

For the natural flow ψ^0, the potential function λ^0 will correspond to the topographical level or altitude. Now the potential function $\lambda(x_1, x_2)$ for the ψ field represents the economic value of water in the various locations (x_1, x_2), on the assumption that in the natural flow field ψ^0 the value of water was equal everywhere.

The principal advantage of the continuous flow field approach to changes ψ in water flow is that it does not require assumptions or previous commitments about a network of water channels. Rather, this approach determines the principal directions along which water should be directed, and thus *this* analysis should precede any analysis based on assumed hardware locations. A convenient starting point would be to calculate the value of water (the potential function) that would result if the present pattern of production and consumption were achieved by an efficient distribution system on the basis of given transportation cost functions $k(\psi, x_1, x_2)$. Any discrepancy between values so calculated and costs actually incurred would indicate wastes or inefficiencies of one sort or another.

6.2.3 Investment in Transportation Facilities*

Suppose a (small) region is to be developed for the production of a single commodity, and that production and transportation will be undertaken in a market framework by hiring capital and labor at competitive prices. Land is free, i.e. owned by the developer. There is a capacity limit $a(x_1, x_2)$ for production. Variable inputs are proportional to output z and their cost is $h(x_1, x_2)$.

The utility of consumption of the product at location (x_1, x_2) is $u(q, x_1, x_2)$, to be expressed in money terms. In addition to production, transportation must also be

*This subject is treated more fully in Puu (1979b).

financed. Let the operating cost of the transportation system depend on capacity c according to the following simple law

$$k(\phi, c) = k(x_1, x_2) \cdot \left[\frac{|\phi|}{c}\right]^{\eta} \qquad \eta > 1 \tag{59}$$

and let the construction cost (appropriately distributed in time) and maintenance cost per unit of capacity c be $m(x_1, x_2)$. Then the total net benefits of the development project are:

$$\iint u(q, x_1, x_2) - k(x_1, x_2)\left[\frac{|\phi|}{c}\right]^{\eta} - h(x_1, x_2)z - m(x_1, x_2)c \; \mathrm{d}x_1 \mathrm{d}x_2 \tag{60}$$

and this is to be maximized with respect to the policy variables

$$q(x_1, x_2), \phi(x_1, x_2), z(x_1, x_2), \text{and } c(x_1, x_2) \tag{61}$$

subject to constraints

$$z \leqslant a(x_1, x_2) \tag{62}$$

and

$$\operatorname{div} \phi = z - q \tag{63}$$

The efficiency conditions are

$$\left.\begin{aligned} q &= 0 \longleftrightarrow u' < \lambda \\ q &\geqslant 0 \longleftrightarrow u' = \lambda \end{aligned}\right\} \tag{64}$$

Consumption should take place only where its marginal utility can be made equal to the price $\lambda(x_1, x_2)$

$$\left.\begin{aligned} z &= 0 \longleftrightarrow \lambda < h + \gamma \\ z &\geqslant 0 \longleftrightarrow \lambda = h + \gamma \end{aligned}\right\} \tag{65}$$

Here γ is a rent on land (or on resource capacity a)

$$\left.\begin{aligned} \gamma &= 0 \longleftrightarrow z < a \\ \gamma &\geqslant 0 \longleftrightarrow z = a \end{aligned}\right\} \tag{66}$$

Condition (65) specifies where production should take place: only at sites that can yield a nonnegative rent γ. In fact, where γ is zero, the limits of the development are reached.

$$\frac{k\eta|\phi|^{\eta}}{c^{1+\eta}} = m \tag{67}$$

The marginal cost m of transportation investment should equal its marginal benefit. This may be solved for c

$$c = \left[\frac{\eta}{m} \cdot k|\phi|^{\eta}\right]^{1/(1+\eta)} \tag{68}$$

In the special case where $\eta = 1$

$$c = \left(\frac{k}{m}|\phi|\right)^{1/2} \tag{69}$$

Finally we have the following efficiency condition for flow

$$\frac{\eta k}{c^{\eta}} |\phi|^{\eta-2} \phi = \text{grad } \lambda \tag{70}$$

Here we may substitute for c

$$\eta k \left[\frac{\eta}{m} k|\phi|^{\eta}\right]^{-\eta/(1+\eta)} |\phi|^{\eta-2} \phi = \text{grad } \lambda$$

or

$$(\eta k)^{1/(1+\eta)} m^{\eta/(1+\eta)} |\phi|^{-(2+\eta)/(1+\eta)} \phi = \text{grad } \lambda \tag{71}$$

In the special case where $\eta = 1$

$$(km)^{1/2} |\phi|^{3/2} = \text{grad } \lambda \tag{72}$$

In this way optimal flows and investments are determined simultaneously.

The following result is easily derived along the same lines. Let the unit cost of transportation be given by

$$k = k\left(\frac{|\phi|}{c}, \underline{x}\right)$$

and let congestion toll charges $k' \cdot (|\phi|/c)$ be charged (where k' denotes the partial derivative with respect to $(|\phi|/c)$ as suggested in Section 2.4.4. Then the optimal local expenditure on road capacity equals the local income from congestion tolls

$$mc = k' \frac{|\phi|}{c} \cdot |\phi|$$

6.2.4 Increasing Returns in Production

The problem to be considered next is best introduced by looking at a discrete set of locations, some of which are producer locations and some of which are consumer locations or markets. When production takes place under constant or diminishing returns, one can associate with each location an excess supply function in terms of local price, and competitive market equilibrium can be visualized, defined, and described. In fact let p_i be price at location i, $q_i(p_i)$ excess demand, x_{ij} commodity shipment from i to j, and h_{ij} transportation cost per unit shipment. The equilibrium conditions are the two equations

$$q_i(p_i) + \sum_j (x_{ij} - x_{ji}) = 0 \tag{73}$$

and

$$\left.\begin{array}{l} x_{ij} = 0 \longleftrightarrow p_j - p_i < k_{ij} \\ x_{ij} \geqslant 0 \longleftrightarrow p_j - p_i = k_{ij} \end{array}\right\} \tag{74}$$

It can be shown that these equilibrium conditions may be obtained through the maximization of a suitable consumers' and producers' surplus integral with the transportation cost introduced explicitly (Samuelson 1952). It is then a straightforward matter to study the comparative statics of this equilibrium system.

Introducing increasing returns changes the nature of the suppliers and the markets. At each location there will be only one supplier (if increasing returns continue for ever, as we shall assume here). As before, each market is served by only one or a small number of supplier locations. If we assume profit-maximizing behavior in terms of Edgeworth–Bertrand strategies of mutual price cutting, one firm will survive in each market, i.e. the lowest-cost supplier, whose costs are determined by his scale. The price actually charged is determined by the cost of the second cheapest supplier and equilibrium occurs at a price below that cost. This equilibrium need not be unique and hence will depend on initial conditions and on dynamics of the price-setting and entry–exit behavior assumed for each of the participating firms.

Fascinating simulation games may be played and a multiplicity of outcomes observed. These will show some degree of regularity and predictability only when transportation costs are large relative to the extent of the returns to scale. For then, the final result is a large number of firms occupying locations at fairly regular distances and each commanding their own territory with safety.

A technical change, a shift in demand, or even a change in the ownership of firms may upset such an equilibrium, particularly if the equilibrium was only locally stable, i.e. stable against small changes. Just such a change of ownership occurred in the American brewing industry through the takeover of some large brewers by conglomerates. Backed by larger financial resources, Budweiser and Miller have engaged in more aggressive marketing strategies. This has reduced the market share and even threatened the existence of such former giants as Schlitz and it has eliminated or jeopardized many local brewing companies. Schlitz's misfortunes are in part attributable to a change in product design (taste) that did not find favor with consumers. This is one illustration of the dynamics of an industry characterized by increasing returns and significant transportation and communication costs. It appears that monopolistic competition in the national beer market of the United States is going to be superseded by massive oligopoly.

A deeper understanding of the spatial side of these issues requires an explicit introduction of spatial coordinates. This in turn means an analysis in continuous space.

It has become fashionable to contrast two fundamental approaches to the location of economic activities, associated respectively with the names of von Thünen and Weber (cf. Section 1.1). In its boldest form, the Weberian type of model locates discrete activities at discrete points ignoring space requirements but emphasizing the economics of distance, while keeping in mind cost factors that may vary with location. The von Thünen approach asks to what economic activity a given piece of land should be allocated. It emphasizes land requirements as well as distance to market and pays due attention to the dependence of other cost factors on location. In a complete analysis of spatial equilibrium there is room for both approaches. In partial analysis, the Weberian model is better adapted to production activities that have increasing returns to scale and are thus localized, i.e. concentrated in a few places, even in an otherwise perfectly homogeneous economic environment. The von Thünen model seems more "natural" when activities with constant or diminishing returns to scale are considered.

Suppose, however, we try treating economies of scale in a von Thünen setting. Can it be done, what does it mean, and might it be useful? This is the problem addressed in this section. To simplify the analysis we keep all other complications to a minimum. Thus we consider a single commodity offered in a market whose participants act as if they are in perfect competition: as price takers. As before, let the density of local demand be a given function of the local (delivered) commodity price $p(x_1, x_2)$

$$q(p, x_1, x_2) \tag{75}$$

In other words consumers are acting as price takers.

To gain some experience we first treat production under constant returns to scale. Let the production function involve only two factors, land M and labor L, both available at given prices $g(x_1, x_2)$ and $w(x_1, x_2)$, respectively. The production function is assumed to be Cobb–Douglas

$$Z = b(x_1, x_2)L^{\alpha}M^{\beta} \qquad \alpha + \beta = 1 \tag{76}$$

In terms of the densities

$$z = Z/A = \text{output per unit area}$$

$$l = L/A = \text{labor input per unit area}$$

$$m = M/A = \text{land input per unit area}$$

one has

$$z = b(x_1, x_2)l^{\alpha}m^{\beta} \tag{77}$$

Consider the utility function underlying the demand function $q(p, x)$. Let $p(q, \underline{x})$ be the inverse function of (75), which exists if the demand function is strictly decreasing. The utility function is then the consumers' surplus integral

$$u(q, \underline{x}) = \int_0^q p(t, \underline{x})\,\mathrm{d}t \tag{78}$$

The cost of transportation associated with a local flow vector ϕ is given by

$$k(x_1, x_2)\,|\phi|$$

Total transportation cost in the region is then

$$\iint k\,|\phi|\,\mathrm{d}x_1\,\mathrm{d}x_2$$

Between commodity flow ϕ and the excess production $z-q$ there exists the divergence relationship

$$\operatorname{div}\phi = z - q = bl^{\alpha} m^{\beta} - q \tag{79}$$

It is well known that competitive equilibrium is attained when and only when the consumers' surplus, net of costs of production and transportation, is maximized

$$\max \iint (u(q, x_1, x_2) - wl - gm - k|\phi|)\, \mathrm{d}x_1 \mathrm{d}x_2$$

subject to the constraint (79) and the restriction on land use

$$m \leqslant 1$$

The Lagrangean of this problem is

$$\iint L \mathrm{d}x_1 \mathrm{d}x_2 \equiv \iint \{u - wl - gm - k|\phi| + \lambda[bl^{\alpha}m^{\beta} - q - \operatorname{div}\phi] + \mu[1-m]\}\mathrm{d}x_1 \mathrm{d}x_2 \tag{80}$$

Variation with respect to q, l, m, and ϕ yields

$$\left.\begin{aligned} u' - \lambda &\leqslant 0 \\ &= 0 \qquad \text{if } q > 0 \end{aligned}\right\} \tag{81}$$

$$\left.\begin{aligned} -w + \lambda\alpha\frac{z}{l} &\leqslant 0 \\ &= 0 \qquad \text{if } l > 0 \end{aligned}\right\} \tag{82}$$

$$\left.\begin{aligned} -g - \mu + \lambda\beta\frac{z}{m} &\leqslant 0 \\ &= 0 \qquad \text{if } m > 0 \end{aligned}\right\} \tag{83}$$

$$\left.\begin{aligned} -k\frac{\phi}{|\phi|} + \operatorname{grad}\lambda &= 0 \qquad \text{if } \phi \neq 0 \\ -k + |\operatorname{grad}\lambda| &\leqslant 0 \qquad \text{if } \phi = 0 \end{aligned}\right\} \tag{84}$$

$$\left.\begin{aligned} \mu &\geqslant 0 \\ &= 0 \qquad \text{when } l < 1 \end{aligned}\right\} \tag{85}$$

Notice that the Lagrangean function, the maximand, is concave so that the "efficiency conditions" are necessary and sufficient for a maximum. Their solution may also be shown to be unique.

An interpretation of these conditions now follows. It shows that we are dealing with a competitive market in which the production and consumption activities are spatially dispersed and the equilibrium is brought about through commodity arbitrage: interlocal shipments whose flow field is ϕ.

Observe first that $\lambda(x_1, x_2)$ is the commodity price and $\mu(x_1, x_2)$ the profit rate or rent per unit area of the firm(s) located at (x_1, x_2). Condition (81) states that consumption is brought to the point where its marginal utility (in money terms) equals price. No consumption takes place where price exceeds marginal utility even at the zero level of consumption. Equations (82) and (83) state that each factor is used to the point where the value of its marginal product equals the factor price. When a factor is not used – which in the Cobb–Douglas case means the product is not produced – the factor price exceeds the value of its marginal product. This can also be stated as marginal cost exceeding the product price. The factor price of land includes an extra rent μ over and above the given land rent g. According to (85) this extra rent is zero when a piece of land is not fully utilized. Finally, equation (84) states that commodity shipments are in the direction of the steepest increase in price and are just profitable in only that direction. In no case can the rate of local price increase exceed the rate of transportation cost.

To study increasing returns we must return to the production function (76). Dividing by area we obtain a production function for output density in terms of input densities and area itself

$$z = \frac{Z}{A} = b\,A^{\alpha+\beta-1}\left(\frac{L}{A}\right)^{\alpha}\left(\frac{M}{A}\right)^{\beta}$$

$$z = b\,A^{\alpha+\beta-1}\,l^{\alpha}m^{\beta} \tag{86}$$

The economic interpretation is that output intensity increases with input intensities and with the size of the contiguous area A over which production takes place. In fact, areal extent A, a measure of scale, acts as another factor of production. In the Cobb–Douglas case its output elasticity equals the degree of returns to scale, $\alpha + \beta - 1$.

This positive effect of the extension of a production activity over an area is countered by the negative effect of the increasing cost of internal communication and transportation. For simplicity we assume that all labor must move at least once during a shift from the center to the working place and back. If the area is circular and this round trip has a transportation cost of h per unit distance, then the total cost equals

$$hl\int_0^{\sqrt{(A/\pi)}} 2\pi r^2\, \mathrm{d}r = hl\tfrac{2}{3}\pi((A/\pi)^{1/2})^3$$

$$= \tfrac{2}{3}hl\,\pi^{-1/2}\,A^{3/2}$$

The cost incurred per unit area, i.e. the density of this cost, is then

$$\tfrac{2}{3}hl\,\pi^{-1/2}\,A^{1/2} \tag{87}$$

As before, the intensity of land use is restricted to unity

$$m \leqslant 1 \tag{88}$$

The density of labor may be similarly restricted. A more natural approach is to consider the cost of housing production activities directly. Increasing labor intensity can be accommodated only through the use of tall multistory buildings and this is subject to increasing cost. In fact, assuming a Cobb–Douglas function for construction and maintenance, one has cAl^{γ} for total "housing cost" of production and

$$cl^{\gamma} \qquad \gamma > 1 \tag{89}$$

for its density per unit area.

The maximand (80) is now replaced by

$$\max_{q,\,l,\,m,\,\phi,\,A} \iint \{u - wl - gm - k|\phi| - \tfrac{2}{3}\, hl\, \pi^{-1/2} A^{1/2} - cl^{\gamma} + \lambda [bA^{\alpha+\beta-1}\, l^{\alpha} m^{\beta} - \operatorname{div}\phi] + \mu[1-m]\} \mathrm{d}x_1\, \mathrm{d}x_2 \tag{90}$$

Before we carry out the optimization it is appropriate to ask what the economic interpretation of such a maximization should be. Under increasing returns the suppliers in one location or area can no longer be considered as perfect competitors. Rather, the entire contiguous area A is taken up by one enterprise. Firms in different areas are identifiable and the industry is thus characterized by oligopoly or at best by monopolistic competition. Which of the two is realized depends essentially on the size of the returns to scale relative to the size of transportation costs, as noted above.

Maximization of a welfare function is consistent with monopolistic competition and regulated entry, but constitutes only a limiting case under oligopoly. In fact, when the space is homogeneous, maximization describes the Löschian type of spatial market equilibrium with regulated entry and competitive pricing, associated with a welfare maximum. Free entry, on the other hand, would require profits to be zero, an additional condition not considered here.

The efficiency conditions are obtained through variation of the integrand with respect to the decision variables q, m, l, A, and ϕ.

$$\left.\begin{aligned} u' &\leqslant \lambda \\ &= \text{when } q > 0 \end{aligned}\right\} \tag{81}$$

$$\left.\begin{aligned} \frac{\lambda\beta z}{m} &\leqslant \mu + g \\ &= \text{when } m > 0 \end{aligned}\right\} \tag{83}$$

$$\left.\begin{aligned} \mu &\geqslant 0 \\ &= \text{when } m < 1 \end{aligned}\right\} \tag{85}$$

$$\left.\begin{aligned} \frac{\lambda\alpha z}{l} &\leqslant w + \tfrac{2}{3}\, h\pi^{-1/2} A^{1/2} + c\gamma l^{\gamma-1} \\ &= \text{when } l > 0 \end{aligned}\right\} \tag{91}$$

$$\left.\begin{aligned} \frac{\lambda(\alpha+\beta-1)z}{A} &\leqslant \tfrac{1}{3}\, h\pi^{-1/2} A^{-1/2} \\ &= \text{when } A > 0 \end{aligned}\right\} \tag{92}$$

$$\left.\begin{aligned} k\frac{\phi}{|\phi|} &= \operatorname{grad}\lambda \qquad \text{when } \phi \neq 0 \\ k &\geqslant |\operatorname{grad}\lambda| \qquad \text{when } \phi = 0 \end{aligned}\right\} \tag{84}$$

The qualifications $l > 0, m > 0, A > 0$ may be replaced by $z > 0$.

Notice that these conditions impose a consistency requirement in terms of A. Output z must be positive precisely in a contiguous area of size A. Notice also that the maximand is no longer concave since $-lA^{1/2}$ is not concave. This means that the efficiency conditions are necessary but not sufficient for a maximum. Moreover their solution need not be unique.

In the spatially homogeneous case it is possible to verify the Löschian solution (Lösch 1954). The solution consists of isolated areas of production having a regular shape at the center of regular market areas of the same shape. Whether this shape is hexagonal or circular or hexagonal with the corners rounded off by a circle depends on the relationship of "overhead cost" to transportation costs, as first observed by Mills and Lav (1964).

The versatility of the continuous modeling approach appears mainly when the homogeneous spatial setting of Löschian equilibrium is replaced by a spatially inhomogeneous environment in which locational choice and size of market areas are affected by variations in local conditions. The efficiency conditions remain the same, but all parameters must be considered as functions of location (x_1, x_2). Qualitatively we still observe contiguous production areas serving larger surrounding market areas. Their extent and location is no longer regular but is affected by a variety of local conditions. For purposes of calculation one may, of course, want to replace the continuous region by a net of discrete cells. But this in no way affects the mathematical analysis or its economic interpretation.

In conclusion we may ask what happens when returns are diminishing

$$\alpha + \beta < 1$$

rather than increasing. We must consider once more the nonspatial production function

$$Z = bL^{\alpha}M^{\beta} \qquad \alpha + \beta < 1$$

The usual justification for diminishing returns is that there is a production factor K that is being held constant but whose inclusion would then generate constant returns to scale

$$Z = bL^{\alpha}M^{\beta}K^{\gamma} \qquad \alpha + \beta + \gamma = 1 \tag{93}$$

Suppose now that K is fixed with respect to the entire area under consideration. Then the production function in terms of areal densities has the form

$$\frac{Z}{A} = b\left(\frac{L}{A}\right)^{\alpha}\left(\frac{M}{A}\right)^{\beta} K^{\gamma} A^{-\gamma} \tag{94}$$

$$z = b K^{\gamma} A^{-\gamma} l^{\alpha} m^{\beta}$$

The implication is that yields per unit area will be maximized when the area under utilization is minimized – an absurd conclusion. It brings to mind Stigler's dictum: when returns to scale are constant and returns to substitution are increasing then the world's demand for wheat could be grown in a flower pot, provided the pot is small enough! In the present case the density of labor and land inputs could be held constant (the latter at unity) and the yield per unit area increased to infinity if the pot were made small enough. We must conclude that (94) and hence (93) are misspecifications. The missing factor that is being held constant should be held at a constant level of density. The production function is accordingly

$$z = b\, l^{\alpha} m^{\beta} \qquad \alpha + \beta < 1 \tag{95}$$

Except for the constraints on $\alpha + \beta$ this has the same form as the production function with constant returns to scale. While under constant returns profit maximization would imply $m = 1$, here it need not. For solving

$$G = \max \lambda b\, l^{\alpha} m^{\beta} - wl - gm$$

without the constraint (44) yields

$$\frac{\alpha \lambda z}{l} = w$$

$$\frac{\beta \lambda}{m} = g$$

$$\frac{l}{m} = \frac{\alpha}{\beta}\frac{g}{w}$$

$$G = \max_{m} \lambda b \left(\frac{\alpha}{\beta}\frac{g}{l}\right)^{\alpha} m^{\alpha+\beta} - g\left(1 + \frac{\alpha}{\beta}\right) m$$

and finally

$$m = [\lambda g^{\alpha-1} w^{-\alpha} b \alpha^{\alpha} \beta^{1-\alpha}]^{1/(1-\alpha-\beta)} \tag{96}$$

and this is small when g is large or λ is small. Thus diminishing returns may induce a less than full utilization of locally available land (the unused part being taken up by other economic activities) and hence a larger areal spread than under constant returns. Conversely, increasing returns, while offering an inducement to large areal expansion, are checked by the increasing cost of internal transportation and communication while labor intensity per unit area is checked by the increasing costs of facilities needed to accommodate such an areal intensity of labor.

To sum up our argument: space acts as a counterforce to increasing returns in production and can thus bring about economic equilibrium in a way that is not only natural but in fact inescapable.

6.2.5 A Commuter Model

Urban structure is approached here in terms of a spatially extended labor market. We restrict ourselves to one homogeneous type of labor, e.g. white-collar employees of specified qualification. At job location (x_1, x_2) let there be $z(x_1, x_2)$ jobs of this type available. At residential location (x_1, x_2) let there be living $q(x_1, x_2)$ employees of this type. We assume free mobility between residences and between jobs. There will be a tendency for households to arrange themselves such that cross-hauling is avoided. This is true even in the presence of preferences for housing locations. We assume that the competitive market maximizes the aggregate net utility of housing and commuting. Let x_0 be the amount of housing per household, and let there be q households at location (x_1, x_2) occupying an amount $h(x_1, x_2)$ of housing. Then

$$qx_0 \leqslant h \tag{97}$$

The next constraint concerns the total number of households to be housed. Assuming (for simplicity) one jobholder per household, this number equals the number J of jobs of the type considered, and this number is assumed to be given

$$\iint q(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \iint z(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = J \tag{98}$$

The aggregate gross utility of housing is

$$\iint q(x_1, x_2)\cdot u(x_0, x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

Net utility is obtained by subtracting commuting costs.

Commuting flow is described by a flow field ϕ whose direction is that of commuting and whose strength $|\phi|$ is the number of commuters moving across a unit cross-section per working day. Actually it will be convenient to consider commuting in terms of the return trips from work to home. At any location (x_1, x_2) the number of commuting trips terminating is then $q(x_1, x_2)$ and the number of commuting trips starting is $z(x_1, x_2)$. The net origination of trips is then

$$z(x_1, x_2) - q(x_1, x_2)$$

By the "equation of continuity" for continuous flow fields, net origination or source density equals the divergence of the flow field. Thus

$$\operatorname{div}\phi = z - q \tag{1}$$

Next, consider the cost of commuting trips. In an urban context and for rush-hour traffic, it is reasonable to assume that this cost is strongly nonlinear. In fact, assume that the cost of commuting unit distance is a quadratic function of the volume of traffic $|\phi|$

$$\tfrac{1}{2}\,k(x_1, x_2)\,|\phi(x_1, x_2)|^2$$

Subtracting the aggregate commuting cost from aggregate housing utility, we obtain aggregate net utility – the welfare function that is optimized through the competitive labor and housing markets

$$\max_{\phi,\, q,\, x_0} \iint \left(qu - \frac{k}{2} \,|\phi|^2 \right) \mathrm{d}x_1 \,\mathrm{d}x_2 \tag{99}$$

subject to

$$qx_0 \leqslant h \tag{97}$$

$$\iint q \,\mathrm{d}x_1 \,\mathrm{d}x_2 = J \tag{98}$$

$$\operatorname{div} \phi + q - z = 0 \tag{1}$$

The Lagrange function of this problem may be written as

$$\iint qu - \frac{k}{2} |\phi|^2 - \lambda [\operatorname{div} \phi + q - z] + \mu [h - qx_0] \,\mathrm{d}x_1 \,\mathrm{d}x_2 + \nu \left[J - \iint q \,\mathrm{d}x_1 \,\mathrm{d}x_2 \right] \tag{100}$$

Maximization with respect to q, x_0, ϕ yields

$$\left.\begin{aligned} q &= 0 \longleftrightarrow u - \mu x_0 - \lambda < \nu \\ q &\geqslant 0 \longleftrightarrow u - \mu x_0 - \lambda = \nu \end{aligned}\right\} \tag{101}$$

$$q \frac{\partial u}{\partial x_0} - q\mu = 0$$

or

$$\frac{\partial u}{\partial x_0} = \mu \tag{102}$$

$$-k\phi + \operatorname{grad} \lambda = 0$$

or

$$\phi = \frac{1}{k} \operatorname{grad} \lambda \tag{103}$$

Efficiency condition (102) is the most straightforward: space x_0 is used by a household up to the point where its marginal utility equals rent. Efficiency condition (101) determines the net utility ν that can be obtained by a household through a wise choice of residence for a given job or a wise choice of job for a given residence. This net utility equals the gross utility of housing u, minus rent payments μx_0, minus λ. Observe that utility is in money terms, i.e. the utility function is linear in terms of "general consumption" and can thus be expressed in income units. λ represents the cost of commuting, to the nearest job. More-distant firms must make an adjustment in their wages to compensate for job location. If $\lambda = 0$ at the site of the best located firm, then $\lambda \geqslant 0$ at the location of any other firm. The value of λ at the location of a particular firm measures the extra cost to commuters of reaching this firm. Salaries must be raised by this amount above the level at the best located firm, and this will happen when the labor market is perfectly competitive. To the extent that commuting involves a significant expense as a result of energy shortages, allowances for commuting costs may be expected to be incorporated into job compensation. The rising cost of commuting can be expected similarly to encourage efficiency in the choice of location of residence in relation to a given job, or of jobs in relation to a given residence as assumed in this model.

Equation (103), perhaps the most interesting, describes how the commuter flow is oriented. It is the gradient field to a potential function λ that measures the distance from the job along the best route (except for an additive term, depending on the location of the firm). Equations (1) and (103) may be combined

$$\operatorname{div}\left(\frac{1}{k}\operatorname{grad}\lambda\right) + q - z = 0 \tag{104}$$

Thus, once the density of the household population q and the distribution of jobs z are given, the flow field is determined through (104).

6.2.6 Structure of a Dispersed City

We now consider an illustrative – essentially one-dimensional – example.

In a hypothetical city let jobs be located in the (circular) center and in an outer ring (Figure 6.4). Let the density of jobs be A in the central area and a in the suburban area. Let the density of housing be zero in the center and b households per unit area in the outer ring. Assume a single job market. The total number of jobs is

$$J = \pi A\, r_0^2 + \pi a(r_2^2 - r_1^2)$$

The total number of households is

$$H = \pi b(r_1^2 - r_0^2)$$

Feasibility requires that

$$H = J$$

or

$$Ar_0^2 + a(r_2^2 - r_1^2) = b(r_1^2 - r_0^2) \tag{105}$$

The commuter flow pattern will be as follows. There is a critical distance r_* for commuters such that those employed in the center will live at distances less than or equal to r_* away from the center, with commuter flow toward the center. Beyond r_* live those whose work is in the suburban fringe area and their commuting direction is outward with respect to the center. Now r_* is given by

$$Ar_0^2 = b(r_*^2 - r_0^2) \tag{106}$$

In view of (105) it is then also true that

$$a(r_2^2 - r_1^2) = b(r_1^2 - r_*^2) \tag{107}$$

If all housing is equally preferred, then the rent of housing will depend on distance r

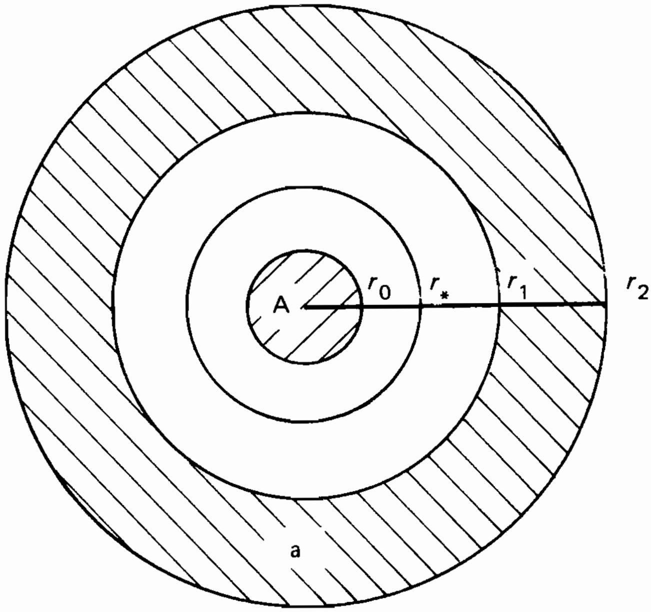

Figure 6.4. Zones of the city.

from the center only. In fact, the sum of commuting cost to the most distant job plus rent must equal a constant. In view of the *a priori* equality of supply and demand, this constant is arbitrary. If commuting costs were proportional to distance, then rents would fall from r_0 to r_* at a constant rate k and rise from r_* to r_2 at the same constant rate k. Since commuting cost depends on flow density, however, the rent pattern is slightly more complicated.

To obtain the density of traffic at any distance r

$$r_0 \leqslant r \leqslant r_*$$

Observe that an amount $\pi b(r_*^2 - r^2)$ of traffic must cross the circle of circumference $2\pi r$. Thus the density of flow is

$$|\phi(r)| = \frac{\pi b(r_*^2 - r^2)}{2\pi r} = \frac{b(r_*^2 - r^2)}{2r} \tag{108}$$

The cost of all traffic moving unit distance at a given point is

$$\frac{k}{2}|\phi(r)|^2 \tag{109}$$

and the cost to one unit of traffic is $(k/2)|\phi|$. The cost of commuting to the center is then

$$\int_0^r \frac{k}{2}|\phi(x)|\,dx = \frac{kb}{4}\int_0^r \frac{r_*^2 - x^2}{x}\,dx = \frac{kb}{4}(r_*^2 \ln r - \tfrac{1}{2}r^2) + c$$

(Actually the integral cannot be extended to $r = 0$ since traffic density would approach infinity there.) We conclude that in the inner residential ring, i.e. for

$$r_0 \leqq r \leqq r_*$$

rent is

$$\mu(r) = \mu_0 - \frac{kb}{4}(r_*^2 \ln r - \tfrac{1}{2} r^2) \tag{110}$$

where μ_0 is some constant.

In the outer residential area, total traffic passing through a circle of radius r is

$$\pi b(r^2 - r_*^2)$$

so that traffic density is

$$|\phi(r)| = \frac{b(r^2 - r_*^2)}{2r}$$

Commuting to distance r_2 costs

$$\int_r^{r_2} \frac{k}{2} |\phi(x)| \, dx = \frac{kb}{4} \int_r^{r_2} \frac{x^2 - x_*^2}{x} \, dx$$

$$= \frac{kb}{4}\left[c - \frac{r^2}{2} + r_*^2 \ln r\right]$$

Rent is then

$$\mu(r) = \mu_1 - \frac{kb}{\phi}\left(r_*^2 \ln r - \frac{r^2}{2}\right) \tag{111}$$

At $r = r_*$ the two expressions for rent must be equal. Therefore $\mu_0 = \mu_1$ and the same expressions for rent apply in both zones. In fact, minimizing $\mu(r)$ with respect to r

$$0 = \frac{r_*^2}{r} - r$$

yielding $r = r_*$

Consider also the potential function λ

$$\lambda = u - \mu x_0 - \nu \tag{112}$$

Since u, ν, and x_0 are constant this is a linear transformation of μ. It is remarkable that, although the flow field reverses its direction at distance $r = r_*$ and is thus discontinuous (when looked at in terms of unit vectors), the potential function λ remains continuous and differentiable – in fact analytic – everywhere.

7 *An Interaction Model*

7.1 INTRODUCTION

7.1.1 Basic Concepts

All of the preceding chapters have dealt with *trade* models of various kinds. Each commodity was shown to have a unique flow, representable by a well-behaved vector field. In some cases there were several commodity flows, but their number was always finite and the flow for each commodity unique.

In our simplified world of single transportation systems, where cross-hauling is ruled out, uniqueness is the result of rational behavior. Whenever this is the case it seems superfluous to record information about the origin and destination of each single commodity unit. It is immaterial whether such a unit, delivered to consumers at a certain location, has followed a flow line all the way from producers at a distant location, or has entered at an intermediate location to replace an identical unit in the original flow.

But what if the units "produced" at different locations are all unique? As long as we are dealing with commodity trade this would appear to be an unnecessary complication, since any real commodity has a sufficient degree of homogeneity to justify the mild abstraction from individual variations. We should at least be able to break the set of commodities down into a more refined, but still finite, set of brands for which the abstraction would be justified.

However, if we consider general purpose communication (or interaction) between individuals at different locations, rather than commodity trade, then the "units" produced and consumed are all different as soon as either the origins or the destinations differ. *All* locations have to communicate with *all* the other locations, and no such communication can be replaced by an equivalent communication, obtained by changing the origin or the destination. No equivalents exist!

This takes us into the world of interaction models. In order to maintain our continuous paradigm, we have to deal with a nondenumerable infinity of vector fields, each corresponding to a fixed origin or a fixed destination. These vector fields do not merge into one resultant field, as they do in any physical application. They coexist separately, so that an infinity of flows, all with different origins and directions, pass through each location. The aggregate of the *norms* of all these flows is a measure of traffic through that location.

Traffic, defined in this way, is one of the important variables in the model of regional structure that follows. Ultimately, traffic depends on the demand for communication between various locations and on the choices of optimal routes for these communications. The demand for communications is assumed to depend multiplicatively on population densities at the locations of origin and destination. This is a very simple variant of interaction theory, where cost–distance dependence is ignored completely, along with the so-called balancing factors.

Cost–distance dependence is omitted because otherwise communication costs could be minimized by the absurd method of making communication so difficult that people would abstain from it altogether. Of course, we could evaluate communication and weigh its value against its cost. But it is easier to stipulate a pattern and volume of communication as a constraint and minimize the costs of realizing it. As we have a limited urban area in mind, this is not too unreasonable.

It also seems reasonable to delete balancing factors when dealing with general-purpose communication. With regard to commuting, for example, the idea that a doubling of workers and jobs entails a quadrupling of trips is just as absurd as the idea that total communication quadruples in a doubled population is sensible. This is because population growth also increases the diversity of activities.

7.1.2 Traffic Distribution and Land Use

So far we have discussed the demand for communication. In order to determine the (infinity of) flow fields in our model, we also have to consider the choice of routes. This problem is dealt with as elsewhere in the book, by choosing routes so that the path integrals of local transit costs are minimized along them. The transit cost is again a location-dependent, but direction-independent (isotropic), scalar field. However, we do not take it as a given datum, but assume that it depends on congestion measured as the ratio of traffic to road capacity at the location concerned.

These two theoretical aspects, the simple interaction model for communication demand and the optimal routing paradigm, can be used to derive all the flow fields and hence the traffic distribution. It should be noted that a complicated feedback mechanism is involved, as traffic depends on optimal routing, which depends on traffic! So, the resulting traffic distribution is an equilibrium and may be difficult to compute in practice. The given data, which result in an equilibrium traffic distribution, are the distributions of population and of road capacity.

Before considering these data and the rest of our model, let us clarify the fact that we are attempting to compute the communication costs for each location within the flow field. We know the numbers of communications terminating at each of the other locations, we know the best communication routes, and we know the local transit costs along them. A location separated from the main part of the population by some highly congested area will obviously suffer from high communication costs.

Suppose people are free to move from one location to another. What could make them accept such high communication costs? Low housing costs is one self-evident answer. So, let us stipulate a spatial invariance of the sum of communication and housing costs as a condition for equilibrium in the spatial population distribution.

In line with our assumption that each individual needs a certain number of communications with each of the other individuals, we also assume that each individual requires a certain living space. If few people live in an area, they can be housed in one-story buildings of relatively weak construction; but with a growing population density we have to build higher and higher, at an increasing capital cost per unit of artificial housing space created. The assumption is that the higher the ratio of population to the space available for housing the more expensive it is to provide an individual with the required living space.

We are now in a position to close the model. The natural space available for housing is obviously the part of it not used for the transportation network. We have the option of using each piece of land to facilitate either housing or communication. If we use it for the first purpose, the result will be a decrease in population crowding and residential construction costs and an increase in traffic congestion and local transit costs. A balance obviously has to be struck for land use at each location so that the sum of housing and communication costs is as small as possible. On the other hand, we also stipulated that this sum should be a spatial invariant. This was the condition for a population distribution in spatial equilibrium. This equilibrium condition eliminates our last degree of freedom.

7.2 OPTIMAL FLOWS

7.2.1 Flow Fields and Population

Let us now formalize the model. As usual, we denote the region studied A and its boundary ∂A. In this model we deal with pairs of origin and destination locations. Let them be denoted $\xi = (\xi_1, \xi_2)$ and $x = (x_1, x_2)$, respectively. Next, define the population density function

$$p = p(x_1, x_2) \tag{1}$$

For convenience we abbreviate population $\bar{p} = p(\xi_1, \xi_2)$ at the origin, whereas we let $p = p(x_1, x_2)$ denote population at the destination. This convention is useful because the origins are kept fixed as long as we deal with individual flow fields. (We could have chosen the destination as fixed instead, which would have worked equally well.) Total population is

$$P = \iint_A p \, dx_1 \, dx_2 \tag{2}$$

As already indicated, an individual flow field can be defined uniquely when the origin is fixed. Denote it

$$\phi = (\phi_1(x_1, x_2), \phi_2(x_1, x_2)) \tag{3}$$

Of course, the vector field also depends on the location ξ of origin, but by keeping it fixed we can delete it as an explicit argument in (3). It should be noted that with a fixed origin, all the vector operations (such as taking the divergence) are carried out with respect to the variable x-coordinates, not the fixed ξ-coordinates.

According to our assumption on communication demand, the number of "communications" originating in ξ and leaving the flow ϕ in the destination x equals the product $\bar{p}p$ of population densities. This is the sink density, so that we obtain

$$\operatorname{div} \phi = -\bar{p}p \tag{4}$$

as the relevant divergence law. In order to avoid confusion, it should be kept in mind that the divergence, $\partial\phi_1/\partial x_1 + \partial\phi_2/\partial x_2$, is taken in terms of the x-coordinates.

In order to formulate the gradient law, the transit cost function has to be defined. It is not a given function of location, but depends on the ratio of traffic to capacity as a measure of congestion. Denote traffic by i and capacity by m. Then transit cost is

$$k = k(i/m) \tag{5}$$

Using expression (5), the gradient law, as always, reads

$$k\frac{\phi}{|\phi|} = \operatorname{grad} \lambda \tag{6}$$

7.2.2 The Potentials

Two observations concerning λ are in order. First, it is similar to ϕ in its dependence on the location of origin ξ, and we cannot delete its coordinates as arguments of the scalar field unless the origin is kept fixed. When ξ is regarded as variable, we get a double continuum of vector fields ϕ, and likewise a double continuum of scalar fields λ. Second, λ is now an undetermined Lagrangean multiplier function, associated with the constraint (4). An interpretation is given below. But note that (6) says nothing more than that the unit flow field is gradient to *some*, as yet undetermined, function whose gradient norm equals local transit costs.

Let us multiply both sides of (6) by the unit vector $\phi/|\phi|$. On the left-hand side the unit vectors multiply to scalar unity, so that $k(\phi/|\phi|)^2 = k$. On the right-hand side we get $\operatorname{grad} \lambda \cdot \phi/|\phi| = \mathrm{d}\lambda/\mathrm{d}\sigma$, where σ is an arc length parameter. This is so because $\phi/|\phi|$ is the unit vector in the direction of the optimal route. Thus

$$k = \mathrm{d}\lambda/\mathrm{d}\sigma \tag{7}$$

and, integrating along any optimal route having its origin at ξ, we obtain

$$\lambda = \int_0^s k\,\mathrm{d}\sigma \tag{8}$$

because $(\mathrm{d}\lambda/\mathrm{d}\sigma)\mathrm{d}\sigma = \mathrm{d}\lambda$ is an exact differential. The arbitrary integration constant in equation (8) was chosen to be zero. This makes λ the path integral of local transit costs along the most efficient routes of communication. Thus, the potential λ will have zero value at the origin and increase in all directions at the rate of local transit costs. Any positive λ defines a closed curve surrounding the origin ξ, and consists of points as far as possible from ξ when the total amount λ is spent on transportation.

7.2.3 A Digression

Let us go somewhat deeper into the matter of determining the unit flow field $\phi/|\phi|$ from equation (6) to ensure that it does not matter if λ is an unknown Lagrangean. In doing so we abuse the terminology from vector analysis somewhat and regard curls and cross products as scalar quantities. A cross product of two vectors (in three-space) is actually a vector, perpendicular to the plane spanned by these two vectors, and pointing in the direction that forms a right-handed set of axes with them. The norm of the cross product is the area of the parallelogram spanned by the two original vectors.

Likewise, the curl is actually a vector along the axis of rotation in a flow, pointing in the direction that makes the rotation counterclockwise, and having a norm equal to the velocity of revolution.

When considering vectors in the plane, both the cross products and the curls always point in directions perpendicular to this plane. Thus, they have only one nonzero component. Our "abuse" entails disregarding the vectorial character of these two concepts and treating them as if they were identical with the (scalar) values of the single nonzero components. This simplification cannot cause any confusion. The only remnant of the vectorial character is the sign (or sense), which depends on whether the resultant vectors point outwards or inwards from the plane.

Using this terminology, the formal definition of the cross product of two arbitrary vectors ϕ and ψ is

$$\phi \times \psi = \phi_1 \psi_2 - \phi_2 \psi_1$$

Similarly, for an arbitrary vector field ϕ, we define

$$\operatorname{curl} \phi = \partial \phi_2 / \partial x_1 - \partial \phi_1 / \partial x_2$$

With respect to the cross product, we should note the trigonometric formula $\phi \times \psi = |\phi|\,|\psi| \sin \alpha$, where α is the angle between the directions of the vectors. As we similarly have $\phi \cdot \psi = |\phi|\,|\psi| \cos \alpha$, we derive the useful relation $(\phi \times \psi)/(\phi \cdot \psi) = \tan \alpha$.

7.2.4 Equation for Route Directions

After these preliminaries we may begin with equation (6) by taking the curls of both sides. Now a gradient field is always irrotational and the curl is hence identically zero. So

$$\operatorname{curl}(k\,\phi/|\phi|) = 0 \tag{9}$$

Expanding this expression we get

$$\operatorname{grad} k \times \phi/|\phi| + k \operatorname{curl}(\phi/|\phi|) = 0 \tag{10}$$

(Note the similarity of this expression to the corresponding one for the divergence.)

Next, denote the direction of grad k by ω and the direction of $\phi/|\phi|$ by θ. Using our trigonometric relation between cross and dot products, and noting that grad $k \cdot \phi/|\phi| = \mathrm{d}k/\mathrm{d}\sigma$, we get

$$\mathrm{d}k/\mathrm{d}\sigma \sin(\theta - \omega) + k \operatorname{curl}(\phi/|\phi|) \cos(\theta - \omega) = 0 \tag{11}$$

But, $\phi/|\phi| = (\cos\theta, \sin\theta)$ and so, by definition of the curl and using the chain rule

$$\operatorname{curl}(\phi/|\phi|) = \cos\theta\, \partial\theta/\partial x_1 + \sin\theta\, \partial\theta/\partial x_2 \tag{12}$$

However, as $(\cos\theta, \sin\theta) = (\mathrm{d}x_1/\mathrm{d}\sigma, \mathrm{d}x_2/\mathrm{d}\sigma)$, we immediately transform (12) into

$$\operatorname{curl}(\phi/|\phi|) = \mathrm{d}\theta/\mathrm{d}\sigma \tag{13}$$

Substituting into (11) we get

$$\sin(\theta - \omega)\, \mathrm{d}k/\mathrm{d}\sigma + k \cos(\theta - \omega)\, \mathrm{d}\theta/\mathrm{d}\sigma = 0 \tag{14}$$

Let us briefly consider (14). The angle ω is defined by the gradient direction to k, the local transit cost, which is known. The variation of this cost in the direction of the route, $\mathrm{d}k/\mathrm{d}\sigma$, depends only on the direction. Accordingly, as we follow the route, (14) involves only the direction of the route θ, and its rate of change $\mathrm{d}\theta/\mathrm{d}\sigma$, as unknowns. We thus have a differential equation for the route direction with arc length as argument.

This differential equation in fact justifies our assertion that (6) would allow us to derive the flow lines, despite the fact that λ is unknown. Its character is most easily understood by considering some special cases.

7.2.5 Refraction of Traffic

First, we assume that ω is invariant in space, so that $\mathrm{d}k/\mathrm{d}\sigma = 0$ and we can drop the first term in (14). What remains can be written

$$\mathrm{d}/\mathrm{d}\sigma\, (k \sin(\theta - \omega)) = 0 \tag{15}$$

which has the first integral

$$k \sin(\theta - \omega) = \text{constant} \tag{16}$$

Obviously, the sine of angular difference between the direction of maximum transit cost increase and the direction of the route is related reciprocally to transit costs. If the route takes us to locations where transit costs increase, we decrease the angular difference in order to pass through the high-cost region as rapidly as possible. If transit costs decrease along the route, we increase the difference in order to profit from the low costs during transit, which is as long as possible.

Equation (16) again reminds us of geometrical optics. At a separation point between two media with refractive indices k_1 and k_2 (and $\omega = 0$ arbitrarily, as it is not defined when k_1 and k_2 are sectionally constant) we have

$$k_1/k_2 = \sin\theta_1/\sin\theta_2 \tag{17}$$

This is Snell's law, which states that the sines of angles of incidence have the same ratio as the refractive indices. It is noteworthy that the corresponding refraction law for transportation, on land and sea combined, with different transit costs (k_1, k_2), was formulated by two economists, Palander (1935) and von Stackelberg (1938).

7.2.6 Circular Symmetry

Second, relax the constraint of a constant k, but assume that it displays circular symmetry. Thus, we can write $k(\rho)$, where $\rho = (x_1^2 + x_2^2)^{1/2}$. In view of our complete model this, of course, means that the congestion ratio i/m itself depends on location x via $\rho = |x|$ only. From this circular symmetry of k, we get

$$\operatorname{grad} k = dk/d\rho(x_1/\rho, x_2/\rho) \tag{18}$$

But, as ω was *defined* to be the angle of the gradient of k, we can identify the vectors $(x_1/\rho, x_2/\rho)$ and $(\cos\omega, \sin\omega)$. So

$$x_1 = \rho\cos\omega \tag{19}$$

$$x_2 = \rho\sin\omega \tag{20}$$

and we note that we have introduced polar coordinates in place of the Cartesian ones. Let us now differentiate (19)–(20) with respect to the arc length parameter σ, and for convenience denote derivatives with respect to arc length by a dot. So

$$\dot{x}_1 = \dot{\rho}\cos\omega - \rho\dot{\omega}\sin\omega \tag{21}$$

$$\dot{x}_2 = \dot{\rho}\sin\omega + \rho\dot{\omega}\cos\omega \tag{22}$$

However, as θ is the direction of the route, $\dot{x}_1$ and $\dot{x}_2$ denote the direction cosines, since differentiation is with respect to arc length. Accordingly

$$\dot{x}_1 = \cos\theta \tag{23}$$

$$\dot{x}_2 = \sin\theta \tag{24}$$

We substitute from (23)–(24) into (21)–(22) and use Cramer's rule to solve for $\dot{\rho}$ and $\rho\dot{\omega}$, which are treated as the two unknowns in the resulting system. In the explicit solutions we use the formulas for the cosine and the sine of a difference to obtain

$$\dot{\rho} = \cos(\theta - \omega) \tag{25}$$

$$\rho\dot{\omega} = \sin(\theta - \omega) \tag{26}$$

These trigonometric expressions are now substituted back into our original differential equation (14), which now reads

$$\rho\dot{\omega}\dot{k} + k\dot{\rho}\dot{\theta} = 0 \tag{27}$$

If we now differentiate (26) with respect to arc length once more and use (25) for cos $(\theta - \omega)$, we get $k\dot{\rho}\dot{\theta} = 2k\dot{\rho}\dot{\omega} + k\rho\ddot{\omega}$. Substituting this into (27) and collecting terms, (27) becomes

$$\mathrm{d}/\mathrm{d}\sigma(k\rho\dot{\omega}) + k\dot{\rho}\ddot{\omega} = 0 \tag{28}$$

But this is the same as

$$\mathrm{d}/\mathrm{d}\sigma(k\rho^2\dot{\omega}) = 0 \tag{29}$$

which has the first integral

$$k\rho^2\dot{\omega} = \text{constant} \tag{30}$$

Expression (30) is well known from the mechanics of central fields (such as planetary motion). In order to understand the character of this new differential equation, let us denote the constant by c, write out $\dot{\omega} = \mathrm{d}\omega/\mathrm{d}\sigma$, and note that the arc length element $\mathrm{d}\sigma$ equals $(\rho^2 + \rho'^2)^{1/2}\,\mathrm{d}\omega$ (where $\rho' = \mathrm{d}\rho/\mathrm{d}\omega$). Thus $\dot{\omega}$ is the reciprocal of $(\rho^2 + \rho'^2)^{1/2}$ and (30) reads

$$k\rho^2 = c(\rho^2 + \rho'^2)^{1/2} \tag{31}$$

which is an ordinary differential equation expressed in polar coordinates. It has been studied widely in theoretical mechanics, and in fact its solution can always be obtained by integration (if the independent and dependent variables are interchanged).

Explicit solutions are either easy or difficult to obtain, depending on the character of k. Before giving some illustrations, let us note that if we substitute arc length $\mathrm{d}\sigma = (\rho^2 + \rho'^2)^{1/2}\,\mathrm{d}\omega$, (8) reads

$$\lambda = \int_0^s k(\rho^2 + \rho'^2)^{1/2}\,\mathrm{d}\omega \tag{32}$$

If we regard $\rho(\omega)$ as an unknown function that we have to choose so as to minimize λ, then we obtain (31) as the appropriate Euler equation for this variational problem. This corroborates the gradient law, as we obviously get the same condition by seeking the optimal routes one by one (as parameterized curves $\rho(\omega)$) so that they minimize transportation costs. Once we have solved for the flow lines, so that we know $\rho(\omega)$, we can calculate λ. All of this, of course, applies to the given origin ξ. For another origin, we would have to work through the whole process again.

7.2.7 An Illustration

Let us conclude this section by giving a very simple illustration using power functions

$$k = \rho^{a-1} \tag{33}$$

Unless a is zero, the solution is

$$\rho^a = \alpha \sec(a\omega + \beta) \tag{34}$$

This is a two-parameter family of routes, but one can be removed by fixing a point of origin, so that we obtain a set of radiating curves. From (32) we can also calculate λ, using (33)–(34), and obtain

$$a\lambda = [\bar{\rho}^{2a} + \rho^{2a} - 2\bar{\rho}^a \rho^a \cos(a\omega - a\bar{\omega})]^{1/2} \tag{35}$$

where $\bar{\rho}$ and $\bar{\omega}$ are the polar coordinates for the fixed point of origin ξ, and ρ and ω are the polar coordinates for the variable point of destination. The calculation of (35) is somewhat awkward and is therefore not reproduced here. The logic, however, is simply evaluation of (32) with substitutions from (31)–(34).

For constant λ, (35) describes a set of concentric transportation cost contours to which the routes defined by (34) are orthogonal. It should be emphasized that the sufficiency conditions for extremality are fulfilled for (34)–(35) only in the neighborhood of the origin, more specifically in a wedge with vertex in the origin (of the coordinate system, not the central flow field).

The geometrical characters of these solution curves are easily recognizable for low integral values of a. The value zero is a special case for which (34) does not hold (see below). The simplest of the remaining cases is that where $a = 1$. Then, transforming (34) back into Cartesian coordinates, we get

$$\cos\beta\, x_1 - \sin\beta\, x_2 = \alpha \qquad a = 1$$

This is obviously a two-parameter family of *straight lines*. This nature of the routes is intuitively obvious, as $k = \rho^\circ = 1$ makes transportation costs equal to route length according to (32).

In the case where we let $a = 2$, so that $k = \rho$, transportation is cheap in the center and becomes increasingly expensive towards the periphery. We expect optimal routes to be deflected from the straight line and to become convex to the origin. This is verified by the formal solution. Turning again to Cartesian coordinates, we have

$$\cos\beta(x_1^2 - x_2^2) - \sin\beta(2x_1x_2) = \alpha \qquad a = 2$$

This formula represents the family of all *hyperbolas* that can be arranged symmetrically around the center (of the coordinate system). By varying α we fill the four sectors, formed by a pair of orthogonal axes through the center, by rectangular hyperbolas. By varying β, we simply rotate any such set of axes and its corresponding family of hyperbolas.

Next, letting $a = 3$, so that $k = \rho^2$, we note that the transportation advantages in the central areas become even greater. We suspect that the convexity of the routes will be even more pronounced. This is confirmed by the formal solution

$$\cos\beta(x_1^3 - 3x_1x_2^2) - \sin\beta(3x_1^2x_2 - x_2^3) = \alpha \qquad a = 3$$

Now, for any fixed β, the space is split into six equal sectors (with vertices in the center). This is similar to the previous case where space was split into four sectors. Again, the sectors are filled by hyperbolic curves, now more sharply convex as they are compressed within angles of 60° (instead of 90°). This is a so-called *monkey saddle* flow, whereas the saddles were ordinary in the previous case. A change of the value of β again rotates the whole system of solution curves.

Since space is split into sectors (of 60° or 90°) in the last two cases, we can understand that there is no solution curve according to (34) that joins an origin and a destination separated by an acute angle larger than 60° or 90°. Thus it becomes clear that (34) only provides a local solution – as previously hinted. Fortunately, there exists another solution to the optimal routing problem for the remaining cases, i.e. radially from the origin towards the center and out towards the destination again.

As we have covered positive values of a in sufficient detail, let us now move in the reverse direction. Let $a = -1$. Then (34) in Cartesian coordinates reads

$$\cos\beta\, x_1 - \sin\beta\, x_2 = \alpha(x_1^2 + x_2^2) \qquad a = -1$$

This equation represents the set of all *circular arcs* through the center of coordinate space. As expected, the shape of the routes is now concave with respect to the origin. As $k = \rho^{-2}$ it is least expensive to travel in the periphery and to avoid the center as much as possible.

Our final illustration of (34) is to set $a = -2$. We expect the pressure to avoid the central areas to be even more pronounced in this case. The formal solution

$$\cos\beta(x_1^2 - x_2^2) - \sin\beta(2x_1x_2) = \alpha(x_1^2 + x_2^2)^2 \qquad a = -2$$

represents the family of *lemniscates* through the center. Again, for each β, space is split into four sectors. These sectors are now elliptic (not hyperbolic). These cases should suffice for an intuitive understanding of solution (34) in general.

Let us also record the special case of (33) where $a = 0$. Then

$$\ln\rho = \alpha + \beta\omega \tag{36}$$

is the solution that replaces (34). The value of (32) is obtained according to

$$\lambda = [(\ln\rho - \ln\bar{\rho})^2 + (\omega - \bar{\omega})^2]^{1/2} \tag{37}$$

which replaces (35).

7.3 TRAFFIC

7.3.1 Differential Equation for Flow Volume

As indicated in the Introduction, we define traffic at location x by

$$i = \iint_A |\phi| \, d\xi_1 \, d\xi_2 \tag{38}$$

The norms of all the vector fields passing through x are integrated with respect to all possible points of origin ξ. But we cannot integrate according to (38) until we have calculated the norms $|\phi|$, which are not yet known. The discussion thus far has concerned the routes of communication, not the volumes.

In the Introduction we also indicated how the demand for communication determines sink density and thus flow volumes by means of a gravity type of model. As a matter of fact, the exact mathematical condition for how flow volume changes with sink density has already been presented in equation (4). Note that when we know the unit flow field $\phi/|\phi|$ equation (4) becomes a partial differential equation in flow volume alone. As $\phi = |\phi|(\phi/|\phi|)$, we get

$$\operatorname{div} \phi = \operatorname{grad} |\phi| \cdot \phi/|\phi| + |\phi| \operatorname{div} (\phi/|\phi|) \tag{39}$$

where $\phi/|\phi|$ and $\operatorname{div}(\phi/|\phi|)$ are known as soon as we know the flow lines. The only unknowns in (39) are $|\phi|$ and grad $|\phi|$, i.e. the flow volume and its partial derivatives. So, (4) does indeed provide a differential equation for $|\phi|$.

Once we know $|\phi|$ for all ξ we can proceed to the integration (38) and calculate traffic.

7.3.2 An Illustration: Constant Transit Cost

The procedure may be illustrated by a few examples. For our *first* example, suppose that transit cost is constant, i.e. $k = 1$ over the whole region. Moreover, suppose population density is uniform, $p = 1$, everywhere and that the region is the unit disk $A = \{(x_1, x_2) | x_1^2 + x_2^2 \leqslant 1\}$. This is the simplest imaginable case.

Insert $a = 1$ into (33). We then know that (34) is a solution. For the present case it reads

$$\rho = \alpha \sec (\omega + \beta) \tag{40}$$

which is the familiar equation of a straight line written in polar coordinates. It is not surprising that the optimal routes with constant transit costs are straight lines, as in all classical location models.

It is more convenient to put the equation of these straight lines into parametric form. Using our familiar notation, where ξ_1, ξ_2 are the coordinates of the point of origin and θ is the constant angle of the flow line, we write

$$x_1 = \xi_1 + \sigma \cos \theta \tag{41}$$

$$x_2 = \xi_2 + \sigma \sin \theta \tag{42}$$

As always, σ denotes the arc length parameter. Obviously (41)–(42) are equivalent to (40) when we have a set of lines with a common point of intersection. Using (41)–(42), we can easily calculate arc length

$$\sigma = [(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2]^{1/2} \tag{43}$$

Let us now check the gradient of this arc length measure. Obviously

$$\begin{aligned} \operatorname{grad} \sigma &= [(x_1 - \xi_1)/\sigma, (x_2 - \xi_2)/\sigma] \\ &= (\cos \theta, \sin \theta) \\ &= \phi/|\phi| \end{aligned} \tag{44}$$

We can thus identify the unit flow field with the gradient of the arc length. As arc length is measured along straight lines, we obviously have a Euclidean metric. The loci of equal distance are then concentric circles as defined by (43) for any given σ, and the pencil of radials through their common center is obviously the gradient field to this, the simplest of all metrics. Now, using (44)

$$\text{grad}\,|\phi| \cdot \phi/|\phi| \;=\; \text{grad}\,|\phi| \cdot \text{grad}\,\sigma \tag{45}$$

and

$$\text{div}\,(\phi/|\phi|) \;=\; \text{div grad}\,\sigma \tag{46}$$

But the Laplacian div grad $\sigma = \partial^2\sigma/\partial x_1^2 + \partial^2\sigma/\partial x_2^2$ can easily be calculated, from (43), to equal $1/\sigma$. On the other hand, grad $|\phi| \cdot$ grad σ is obviously the derivative $\partial|\phi|/\partial\sigma$. So, according to (39) and (45)–(46)

$$\text{div}\,\phi \;=\; \partial|\phi|/\partial\sigma + |\phi|/\sigma \tag{47}$$

Using the information that $\bar{p} \equiv p \equiv 1$, equation (4) becomes

$$\partial|\phi|/\partial\sigma + |\phi|/\sigma + 1 \;=\; 0 \tag{48}$$

which is quite easy to solve. As neither θ nor any derivative with respect to it appears in (48), we can treat it as an ordinary differential equation with σ as the only independent variable. Dependence on the angle θ is confined to a variation of the arbitrary integration constant. Denoting this constant by S^2, we obtain the solution

$$|\phi| \;=\; \tfrac{1}{2}(S^2 - \sigma^2)/\sigma \tag{49}$$

As we are concerned with communication solely *within* the closed disk, there is no flow crossing the boundary. Moreover, as the routes are straight lines radiating from interior points of the circular region, no route can be tangential to the boundary curve. So, the condition that no flows cross the boundary translates into a condition that all flow volumes are zero on the boundary, i.e. $|\phi| = 0$. From (49) we see that $S = \sigma$ on the boundary, which means that S can be interpreted as the straight-line distance to the boundary ∂A from the point ξ. In other words, S is the distance from ξ to the boundary in the direction θ. This last formulation indicates *how* S depends on θ.

7.3.3 Solution by Elliptic Integrals

Our next task is to evaluate the double integral (38) from (49). But in order to perform the integration efficiently we start by changing integration variables from ξ_1, ξ_2 to σ, θ. Now, (41)–(42) tell us that $\xi_1 = x_1 - \sigma\cos\theta$ and $\xi_2 = x_2 - \sigma\cos\theta$. Note that when we integrate according to (38), we treat the point x_1, x_2 as fixed, thus reversing the roles of ξ and x. It is easy to evaluate the Jacobian of the coordinate transformation as

$$\frac{\partial(\xi_1, \xi_2)}{\partial(\sigma, \theta)} \;=\; \sigma \tag{50}$$

Accordingly

$$d\xi_1 d\xi_2 \;=\; \sigma\, d\sigma d\theta \tag{51}$$

and, from (38) and (49)

$$i \;=\; \tfrac{1}{2}\iint_A (S^2 - \sigma^2)\, d\sigma d\theta \tag{52}$$

The evaluation of the innermost integral is messy, but straightforward, and yields

$$i = \tfrac{1}{6} \int_0^{2\pi} ((S' + S'')^3 - S'^3 - S''^3)\mathrm{d}\theta \tag{53}$$

where

$$S' = (1 - \rho^2 \sin^2\theta)^{1/2} - \rho \cos\theta \tag{54}$$

$$S'' = (1 - \rho^2 \sin^2\theta)^{1/2} + \rho \cos\theta \tag{55}$$

Note that S' and S'' are the lengths of the two segments into which ξ divides a chord of the unit circle in direction θ. Now

$$(S' + S'')^3 - S'^3 - S''^3 = 3S'S''(S' + S'') \tag{56}$$

and from (54)–(55)

$$S'S'' = (1 - \rho^2) \tag{57}$$

$$(S' + S'') = 2(1 - \rho^2 \sin^2\theta)^{1/2} \tag{58}$$

Substituting from (56)–(58) into (53) yields

$$i = (1 - \rho^2) \int_0^{2\pi} (1 - \rho^2 \sin^2\theta)^{1/2}\, \mathrm{d}\theta \tag{59}$$

where we note that $S'S'' = (1 - \rho^2)$ could be moved outside the integration sign, as it does not depend on θ. The rest of our expression is also handy. The integral of $(1 - \rho^2 \sin^2\theta)^{1/2}$ taken over an angle $\pi/2$ defines the complete elliptic integral of the second kind. As $\sin^2\theta$ has a perfect periodicity over $\pi/2$, our integral is simply four times the elliptic integral, denoted as usual by $E(\rho)$. Thus

$$i(\rho) = 4(1 - \rho^2)E(\rho) \tag{60}$$

For convenience, we record the Taylor series for $E(\rho)$, which is the most accessible way of computing it. Thus

$$E(\rho) = \frac{\pi}{2}\left(1 - \left(\frac{1}{2}\right)^2 \frac{\rho^2}{1} - \left(\frac{3}{8}\right)^2 \frac{\rho^4}{3} - \left(\frac{15}{48}\right)^2 \frac{\rho^6}{5} - \ldots\right) \tag{61}$$

The resulting traffic distribution is illustrated in Figure 7.1. Several comments are in order. First, even though the volumes of each flow, according to (49), do not possess circular symmetry, the traffic distribution has such symmetry. This is reasonable as the model as a whole is symmetric. The region is a circular disk, population is uniformly distributed, and transit costs are spatially invariant. According to intuition, traffic i should have the symmetric property. On the other hand the origin ξ, associated with the flow volume $|\phi|$, is in general asymmetrically located in the disk, so that we should not expect any symmetry.

Second, the traffic distribution was relatively difficult to derive despite the fact that we were dealing with an extremely simple case. In general, considerable computational

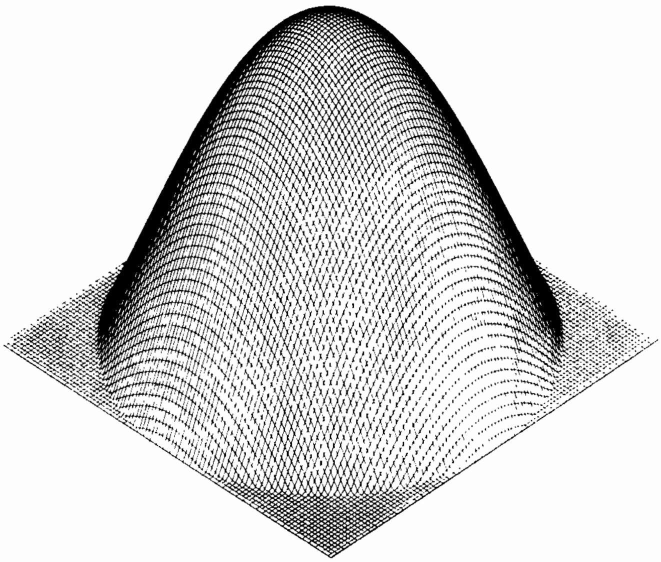

Figure 7.1. Traffic: linear paths.

difficulties can be expected, regardless of the type of example chosen. For a detailed discussion of traffic distributions and simulation techniques, the reader is referred to Puu (1979b).

7.3.4 A Second Example

Let us now turn to our *second* example, which is much easier to treat – although this, of course, is an exception. What happens to the solution (34) if the exponent in (33) increases? If we draw the curves (34) for increasing a, we see that they become more and more sharply convex with respect to the origin. In the limit, as a goes to infinity, the routes become as convex as they can, i.e. they degenerate into pairs of radials joining the points of origin and destination to the center of the region, which is still the unit disk. Thus we arrive at the case of radial transportation, familiar from von Thünen and the new urban economics with its central business district.

Along with the disk-shaped region we retain the assumption of a uniformly dispersed population. From each point of origin, all communications now go to the center first and then radiate out from it in all directions. Thus we conclude that the present flows are all in the direction of grad ρ. So, $\phi/|\phi| = \text{grad}\,\rho$ and formulas (43)–(48) continue to hold true, but with σ replaced by ρ and $\xi_1 = \xi_2 = 0$. The differential equation equivalent to (48) is now

$$\partial|\phi|/\partial\rho + |\phi|/\rho + 1 = 0 \tag{62}$$

Its solution resembles (49), but is simpler

$$|\phi| = \tfrac{1}{2}(1-\rho^2)/\rho \tag{63}$$

Most of the simplicity is due to the fact that the distance to the boundary is a unitary constant, independent of θ. Accordingly, integration with respect to all the origins amounts to multiplication of (63) by the area π of our region. This results from the invariance of (63) with regard to ξ. It should be borne in mind that we have only accounted for communication radiating *out* from the center. There is just as much communication radiating *in* towards the center, so that our measure has to be doubled. Thus, we get

$$i(\rho) = \pi(1-\rho^2)/\rho \tag{64}$$

This traffic distribution is illustrated in Figure 7.2. As traffic becomes infinite at the center, we have removed the infinite peak at a certain level. It is not surprising that in comparison with linear transportation, radial transportation leads to a higher degree of traffic concentration at the center.

It should be noted that even though traffic is infinite at the center, this does not mean that total traffic, the volume under the surface shown (including the infinite peak), is infinite. In fact total traffic $\iint_A i \mathrm{d}x_1 \mathrm{d}x_2$ is an improper integral that converges. Thus

$$\iint_A \pi(1-\rho^2)/\rho \, \mathrm{d}x_1 \mathrm{d}x_2 = 4\pi^2/3 \tag{65}$$

which can be compared to

$$\iint_A 4(1-\rho^2)E(\rho)\mathrm{d}x_1 \mathrm{d}x_2 = 128\pi/45 \tag{66}$$

As $4\pi^2/3 \approx 13.2$ and $128\pi/45 \approx 8.9$, about 50 per cent more traffic is created by radial than by linear transportation. Since linear transportation should lead to a minimum of total traffic, because it corresponds to the choice of the shortest route for each communication, the excess created by radial transportation is surprisingly small.

Our two examples, expressly chosen to admit analytical treatment, should not convey the impression that it is easy to derive explicitly all traffic distributions for any case we would like to treat. On the contrary, the computation is in general very difficult. This is particularly unfortunate in the sense that we should deal with the formidable task of deriving an *equilibrium* traffic distribution when traffic is fed back, via congestion, into local transit costs, which determine the choice of routes and ultimately the traffic distribution itself. We have to conclude that the final equilibrium traffic distribution cannot actually be computed using analytical methods. Computer simulation could be helpful, but would entail a formidable task with respect to the model as a whole.

7.4 COMMUNICATION COST

7.4.1 Alternative Expressions for Communication Costs

Once we are far enough along to be able to calculate traffic, the communication costs for each point of origin can – in principle, of course – also be calculated.

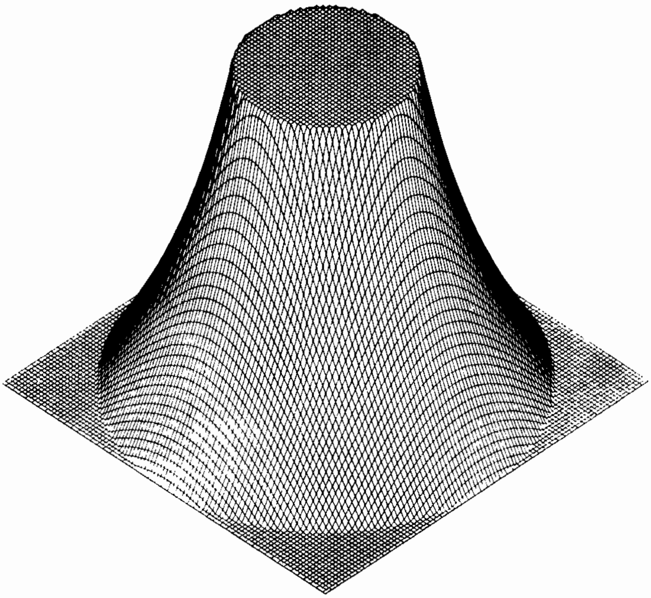

Figure 7.2. Traffic: radial paths.

Let us begin by deriving a general relation between various expressions for communication costs. According to our assumption, if $\bar{p}$ people live at ξ and p live at x, then they need a number $\bar{p}p$ of communications. Each of these has a cost λ, as defined by (8) when optimal routes are chosen with respect to the given transit cost function k. The most obvious expression for transportation costs is

$$T = \iint_A \bar{p}p \, \lambda \, dx_1 \, dx_2 \tag{67}$$

To be exact, T depends on the location of origin ξ. This point is fixed and the integration runs over all destination points. Note that this is the reverse of the case in which traffic distributions were derived.

Now, equation (2) makes it possible to substitute $-\operatorname{div}\phi$ for the product $\bar{p}p$. So

$$T = -\iint_A \lambda \operatorname{div}\phi \, dx_1 \, dx_2 \tag{68}$$

This expression can be transformed conveniently by using Gauss's theorem, but there is one snag. The theorem is not applicable to region A, because the vector field is not regular within it. The troublesome point is the single location ξ of origin. If there were no net outflow from this singularity there would be no difficulty, but we know there is!

So, we use the artifice of defining a new sort of region with a small hole in it. The hole must contain ξ, but can be as small as we wish. For convenience, since we know that the constant λ contours are concentric closed curves surrounding ξ, we let the boundary of the hole be defined by some $\lambda = $ constant. Denote this boundary $\partial' A$ and the region with the hole A'. Obviously, we can make the hole as small as we like by letting $\lambda \to 0$. In other words, we can make A' approach A as closely as we wish by this limiting procedure. The important feature of A' is that ϕ is regular within it, which makes Gauss's theorem applicable.

Consider the formula

$$\iint_{A'} \operatorname{div}(\lambda\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\partial' A} \lambda(\phi)_n\,\mathrm{d}\sigma \tag{69}$$

which states that the surface integral of the divergence of value flow $\lambda\phi$ equals the curve integral of the normal component of this flow along the boundary. This boundary $\partial' A$ is the inner boundary of the hole. Of course, there is an outer boundary ∂A of the whole region. However, we are studying only internal communication within the region, so that this boundary integral can be deleted from the outset, as $(\phi)_n$ is zero on all of ∂A.

In equation (69), λ can be moved outside the sign of integration, since the curve $\partial' A$ was conveniently defined by a constant λ. Next we use Gauss's theorem once again to transform the remaining curve integral of $(\phi)_n$ to a surface integral of div ϕ. Thus

$$\int_{\partial' A} \lambda(\phi)_n\,\mathrm{d}\sigma = \lambda \iint_{A'} \operatorname{div}\phi\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{70}$$

But, div $\phi = -\bar{p}p$, where $\bar{p}$ as a constant can be moved outside the integration signs. Population density in (70) now remains to be integrated. Let us denote the total population of A' as P', by analogy with (2). Accordingly

$$\int_{\partial' A} \lambda(\phi)_n\,\mathrm{d}\sigma = -\lambda\bar{p}P' \tag{71}$$

By letting λ approach zero $\bar{p}$ remains constant, whereas P' goes to P, the population of the whole region A. Formally,

$$\lim_{\lambda \to 0} \int_{\partial' A} \lambda(\phi)_n\,\mathrm{d}\sigma = 0 \tag{72}$$

because $\bar{p}$ and P are finite, whereas λ goes to zero. In this limiting process A' goes to A, so that the following relation is obtained from (69) for our (improper) integral over A

$$\iint_{A} \operatorname{div}(\lambda\phi)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0 \tag{73}$$

Next, use $\operatorname{div}(\lambda\phi) = \operatorname{grad}\lambda \cdot \phi + \lambda \operatorname{div}\phi$ to get

$$\iint_{A} \operatorname{grad}\lambda \cdot \phi\,\mathrm{d}x_1\,\mathrm{d}x_2 = -\iint_{A} \lambda \operatorname{div}\phi\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{74}$$

We are now prepared for the last step. From (6)

$$\operatorname{grad}\lambda \cdot \phi = k|\phi| \tag{75}$$

and substituted into (74), this yields

$$-\iint_A \lambda \operatorname{div} \phi \, dx_1 dx_2 = \iint_A k|\phi| dx_1 dx_2 \tag{76}$$

But according to (68), this equals transportation costs and so

$$T = \iint_A k|\phi| dx_1 dx_2 \tag{77}$$

It is interesting to compare the initial equation (67) with the final derived equation (77). Transportation costs, originally expressed as the aggregate of the number of trips from the origin to other locations multiplied by the cost of each trip, can obviously also be obtained by taking the aggregate of the flow volume at each of the other locations multiplied by the local transit cost. In passing we should note that the equivalence of (67) and (77) applies to all flow fields, not only the optimal (cost-minimizing) field, provided λ is defined as accumulated transit cost along the arbitrary flow lines. This is because we do not need the optimality condition (6) itself, but only its weaker consequence (75).

Equation (77) is much more useful than (67), both in actual computation and in the general discussion that follows.

7.4.2 Derivation of Transportation Costs

We will now give a simple example of how transportation costs can be derived for the case of $k \equiv 1$, $p \equiv 1$, and $A = \{(x_1, x_2)|x_1^2 + x_2^2 \leqslant 1\}$. This is the familiar case of homogeneous space, and hence linear transportation and uniformly distributed population on the unit disk. Traffic distribution has already been derived for this case.

As indicated, it is useful to start from (77). Since $k = 1$ we get

$$T = \iint_A |\phi| dx_1 dx_2 \tag{78}$$

Note the difference between this and expression (38), which defined traffic. Integration with respect to destinations x, not origins ξ, makes a big difference, and the outcome will be different from (60). However, part of the derivation leading to (60) is still relevant, so we can use (49) directly. In order to facilitate integration, we again use the coordinate transformation (41)–(42). The Jacobian is

$$\frac{\partial(x_1, x_2)}{\partial(\sigma, \theta)} = \sigma \tag{79}$$

which happens to be the same as that in (50), due to the symmetry of x and ξ in (41)–(42). Accordingly,

$$dx_1 dx_2 = \sigma \, d\sigma d\theta \tag{80}$$

By substituting (49) and (80) into (78), we thus get

$$T = \frac{1}{2}\iint_A (S^2 - \sigma^2) \mathrm{d}\sigma \mathrm{d}\theta \tag{81}$$

Now, recall that S denotes the distance from ξ in direction θ to the boundary circle. The chord segments (distances in direction θ and $\theta + \pi$) have been recorded in (54) and (55). However, we only need one of them. So, we can set

$$S = (1 - \bar{\rho}^2 \sin^2 \theta)^{1/2} - \bar{\rho} \cos \theta \tag{82}$$

Note that we take $\bar{\rho}$, not ρ, which again has to do with the fact that the origin rather than the destination is fixed.

We still have to fix the limits of integration in (81). Obviously, θ has to make a full round of 2π, but as the second half-round merely repeats the first, we can let θ range from 0 to π and take twice the integral (81) with the limits for θ thus fixed. As for σ, it obviously ranges from 0 to S.

Evaluation of the innermost integral is trivial. We simply obtain

$$\int_0^S (S^2 - \sigma^2) \mathrm{d}\sigma = \frac{2}{3} S^3 \tag{83}$$

Thus (81) becomes

$$T = \frac{2}{3} \int_0^\pi S^3 \mathrm{d}\theta \tag{84}$$

where S is defined in (82). The process of evaluating this last integral is somewhat complicated. By expanding the third power of (82) we get four terms, two of which involve $\cos\theta$ and $\cos\theta \sin^2\theta$. Now, the integrals of these from 0 to π are zero. The remaining terms are $(1 - \bar{\rho}^2)(1 - \bar{\rho}^2 \sin^2\theta)^{1/2}$ and $4\bar{\rho}^2 \cos^2\theta (1 - \bar{\rho}^2 \sin^2\theta)^{1/2}$, respectively. Both have a perfect periodicity over $\pi/2$ so that

$$T = \frac{4}{3}(1 - \bar{\rho}^2) \int_0^{\pi/2} (1 - \bar{\rho}^2 \sin^2\theta)^{1/2} \, \mathrm{d}\theta + \frac{16}{3} \int_0^{\pi/2} \bar{\rho}^2 \cos^2\theta (1 - \bar{\rho}^2 \sin^2\theta)^{1/2} \, \mathrm{d}\theta \tag{85}$$

We again recognize the definition of the complete elliptic integral of the second kind in the first integral. The second can also be evaluated in terms of complete elliptic integrals, but of both the first and second kinds. The series expansion of the elliptic integral of the second kind has already been recorded in (61). For convenience we write the corresponding expression for the elliptic integral of the first kind

$$F(\rho) = \frac{\pi}{2}\left(1 + \left(\frac{1}{2}\right)^2 \rho^2 + \left(\frac{3}{8}\right)^2 \rho^4 + \left(\frac{15}{48}\right)^2 \rho^6 + \dots\right) \tag{86}$$

which is similar to (61). In fact, the minus signs have been reversed and the denominators of the powers of ρ deleted, but otherwise they are the same.

Using these elliptic integrals, we have

$$T = \frac{4}{3} (1 - \bar{\rho}^2) E(\bar{\rho}) + \frac{16}{9}((1 + \bar{\rho}^2) E(\bar{\rho}) - (1 - \bar{\rho}^2) F(\bar{\rho})) \tag{87}$$

This distribution of transportation costs is illustrated in Figure 7.3. Obviously, transportation costs are lowest for those living in the center and increase monotonically, the farther the origin of communications is from the center. This is intuitively appealing.

We might also compare Figures 7.1 and 7.3. Both resulted from integration of the same $|\phi|$, the first with respect to ξ, the second with respect to x. The difference between the two figures illustrates the importance of the coordinates taken for integration. However, the volume under the two surfaces is equal. Regardless of whether we take the integral of (60) with respect to x or the integral of (87) with respect to ξ, we arrive at

$$\iint_A \iint_A |\phi| \, dx_1 dx_2 \, d\xi_1 d\xi_2 = 128\pi/45 \tag{88}$$

since the order of integration is immaterial. One interpretation of this integral is total traffic, as we have seen. Another is total communication costs. As local transit cost is unitary, we can equate total communication costs to total communication distance or, in more familiar terminology, total transport work. Hence, total traffic equals total transportation work. This conclusion is not limited to the case illustrated, but holds in general.

7.5 LAND USE AND EQUILIBRIUM SETTLEMENT

7.5.1 The Problem

Thus far we have considered how the choice of optimum routes, in connection with the demand for transportation, determines the distribution of traffic in the region in question and the distribution of communication costs for various points of origin. The computational aspects were covered in some detail in order to show how complicated an analytical solution can become in even mildly complex cases. In this process the local transit cost was taken as given. We noted only that it depended on the congestion ratio of traffic to space available for transportation. Moreover, due to the feedback mechanism via traffic, we could not regard transit cost as a datum, even if the fraction of space allocated to transportation was assumed given everywhere.

The quantity of land available for transportation results from a decision concerning the use of land. The quantity allocated to transportation is determined by the value of the best alternative use of land, which, in the framework of our model, is housing. The value of land use for housing, on the other hand, depends on population. If, as indicated in the Introduction, we seek a spatial equilibrium where locations are indifferent, due to an exact balance between housing and communication costs, then the distribution of population must be regarded as variable in the model. But let us deal with the problems one at a time. We first consider land use, and then proceed to equilibrium settlement.

Local transit cost k was defined in equation (5) as an increasing function of the (traffic to carrying capacity) congestion ratio i/m. We already have a lengthy derivation of i. Total land available at a location is divided into two fractions, m, used for the transportation network and n, used for housing. If our simplified model is to make any

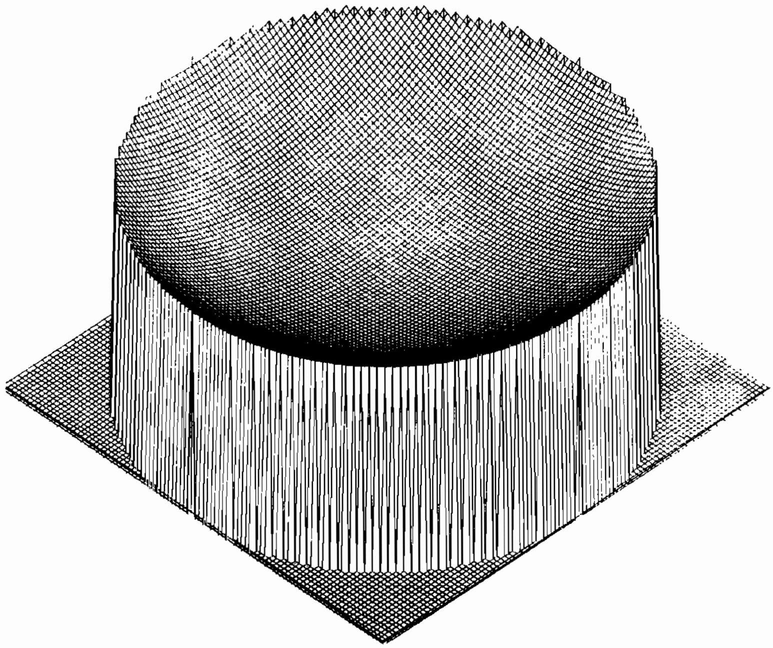

Figure 7.3. Cost and location.

sense, "housing" should be interpreted broadly to include the construction of buildings for productive purposes along with residential construction.

Let us briefly discuss the dependence of k on i/m. Obviously, numerous empirical and theoretical studies (of, for example, the "follow-the-leader" type) suggest a monotonically increasing relation. The increase is very drastic since there is usually a critical congestion level at which the velocity of traffic flow is reduced to zero, and hence its reciprocal, transit time (a proxy for transit cost), goes to infinity. The general picture is not altered, even if we let k include capital costs for maintenance, since repair requirements due to wear obviously increase with congestion, as do the locomotion costs proper. The transit cost function could also take care of the fact that it is possible to push the critical congestion ratio to a higher value by creating *artificial* space, i.e. by setting up several storys of elaborate networks. However, capital costs for such constructions obviously increase with the ratio of traffic to *natural* space available, so that we can retain our specification.

7.5.2 Allocation of Land

As indicated earlier, a decision on land use has to be reached. Now, the use for transportation has been accounted for, but we still have to formalize the use for housing (in the broad sense). Let us suppose that there is a cost function

$$h = h(p/n) \tag{89}$$

for providing each individual with his or her required living space. This cost increases with the crowding ratio, measured by the quotient of population to natural space available for housing. As in the case of land use for transportation, we have in mind a process whereby artificial space is created at an ever-increasing capital cost, the greater the amount of artificial space, in relation to natural space, that has to be constructed. As the need for space was proportional to population, p/n is the correct argument. For convenience, we can restate equation (5)

$$k = k(i/m) \tag{90}$$

and the fact that the proportions of land used for the two purposes included in the model add up to unity

$$m + n = 1 \tag{91}$$

We now have to select the proper expression to be optimized by the choice of m and n. In the Introduction we argued that it would be reasonable to let the sum of housing and communication costs be minimized. However, it would be somewhat absurd to do this for each location separately. From an empirical point of view, public agencies are usually responsible for the planning of land use with respect to entire regions. In theory, all the communications, and not only those from a certain point of origin, will be affected by changing transit cost at this point. Therefore, we can expect to encounter analytical difficulties if we set up a local optimum condition for something that has global effects.

So, dealing with the region as a whole, the total transportation costs are obtained by integrating (77) with respect to ξ and using definition (38), as

$$\iint_A k(i/m)\, i \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{92}$$

On the other hand, total housing costs are

$$\iint_A h(p/n)\, p \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{93}$$

Accordingly, we can minimize the sum of housing and transportation costs (92)–(93) with respect to m and n, subject to (91). This yields

$$k'(i/m)\left(\frac{i}{m}\right)^2 = h'(p/n)\left(\frac{p}{n}\right)^2 = \mu(x_1, x_2) \tag{94}$$

where μ is a (location-dependent) Lagrangean multiplier associated with the constraint. This optimum condition has the attractive property of stipulating a universal relation that must hold everywhere between the local traffic congestion and population crowding ratios. We conclude, assuming second derivatives to be positive (as is reasonable in terms of our discussion), that a high transit cost due to congestion is linked to a high housing cost due to crowding. This seems reasonable as a condition for optimal land use.

It can also be seen that (94) and (91) combined determine both m and n, once i and p are given. The same is then true for k and h. Supposing that we have somehow managed to solve the complicated feedback process of traffic as a determinant for route choice and obtained the equilibrium traffic distribution, we find that the single remaining degree of freedom is the spatial distribution of population.

7.5.3 Equilibrium Settlement

It is assumed that there are no migration incentives if the sum of transportation costs and housing costs is a spatial invariant, i.e. if

$$\iint_A k(i/m)|\phi| \,\mathrm{d}x_1 \mathrm{d}x_2 + h(\bar{p}/\bar{n})\bar{p} = \text{constant} \tag{95}$$

Observe that the barred symbols again refer to conditions at the fixed point of origin ξ.

The model is now complete. If we limit our discussion to the case of a region with circular symmetry, we conclude that something like the case illustrated in Figures 7.1 and 7.3 has come part of the way towards an equilibrium solution. Of course, this is true only in a very general sense.

However, the traffic displayed in Figure 7.1 arose from unit population density and linear communication routes. The latter occurred if transit cost was a spatial constant. Now, transit cost depended on the traffic congestion ratio. As seen from Figure 7.1, there is traffic concentration towards the center of the disk. Accordingly, we have to allocate more land to transportation in the central parts in order to arrive at the constant transit cost (i.e. constant congestion ratio).

On the other hand, little land is available for housing in the central parts as it is used for transportation instead. So, housing should be expensive in the central parts. As for communication costs, we see from Figure 7.3 that they are low in the center and high in the outskirts. It is thus possible that housing and communication costs could balance everywhere.

There is only one qualitative feature in this case that violates our conditions. It was shown that, for optimal land use, the high population crowding in the center should be balanced by a high congestion ratio. The latter, however, is a spatial constant.

Thus it seems that we should either have lower crowding or higher congestion in the center. In equilibrium, a higher congestion ratio would lead to avoidance of the center, and, via feedback, to a reduction in the concentration of traffic there. This could be brought about by allocating less land in the center to transportation and more to housing. This change would result in a better balance between crowding and congestion. However, the cost of communication would be increased for all those who have to communicate via the central parts, not only for those who live there, whereas housing costs would be decreased only locally. Therefore, it is likely that such a reallocation of land would make the center more attractive and encourage people to migrate there.

Finally, we recognize the features of reality: congestion and crowding in the center; a tendency to avoid the central parts for trips not originating or destined there, but nevertheless a considerable concentration of traffic; more land used for transportation than for housing in the center; high costs of housing, offset by centrality of location; etc. In other words, the general case is far too complicated to allow explicit solution.

8 *Spatial Business-Cycle and Growth Models*

8.1 INTRODUCTION

Most of the foregoing analysis has been static in the sense that time was not explicitly included along with the space coordinates. In a very formal sense all our models have been "dynamic," as space plays the same role for us as does time in the dynamic analysis of traditional economics. Spatial analysis is even more complicated than temporal analysis due to the two dimensions of the space we are dealing with. This, however, does not remove the need to include time along with space in the models.

In spaceless economics, dynamic analysis usually means either of two things: the study of price adjustments in a dynamized multimarket equilibrium, or the study of business cycles in macroeconomic aggregates. We have occasionally touched upon the former type of dynamics in the study of price–flow adjustments and in the stability analyses. The present chapter will be devoted to the study of spatial business cycles.

As trade is the natural conveyor of economic change a coupling of business-cycle theory, as developed by Hicks (1950) and Samuelson (1939),* with interregional trade theory is an obvious step towards understanding economic fluctuations. However, this coupling really only gives new insights once we have conceived of trade among geographical locations related to each other by some well-defined distance metric.

If we choose the continuous format we have the advantage of using all that is known about the wave equation for physical oscillating systems. If we keep this analogy in mind then the usual business cycle of the multiplier–accelerator type can be compared to a simple harmonic oscillator. Typically, it has only one frequency of oscillation. In order to admit superposed cycles of different periods, as observed in reality, it has been necessary to introduce complicated lag structures for economic responses so that the beautiful simplicity of the basic model is destroyed.

In comparison to the simple harmonic oscillator a vibrating string is capable of producing any compound of the infinity of natural harmonics. As Fourier's analysis demonstrates, such a system is capable of producing any periodic motion, however irregular it is. This is already much better than the single cosine or sine waves of the simple harmonic oscillator. We attain realism without having to complicate the basic physical laws with something like a distributed lag structure. The single spatial extension of the string makes all the difference!

Obviously, a one-dimensional region in business-cycle theory would give us the same

---

*J.R. Hicks, *A Contribution to the Theory of Trade*, Oxford University Press, London, 1950. P.A. Samuelson, Interactions Between the Multiplier Analysis and the Principle of Acceleration, *Review of Economic Statistics*, 21 (1939): 75–78.

possibility. But, we can do better than this. The oscillations produced by a string are still strictly periodic, though possibly irregular. This is so because the vibration is compounded by natural harmonics. In everyday life we note this fact by ascribing a fundamental pitch to each string.

With a two-dimensional membrane things are completely different. A string can only produce resonance with its harmonics. A membrane, as we know from violin plates or the soundboards of pianos, can be resonant with any of the tones produced from the strings mounted upon it. Accordingly, the membrane can vibrate at the same time in frequencies that are no longer natural multiples of each other, and the compound motion is no longer strictly periodic. Again, it is the two dimensions that make the difference, since the result comes from the same fundamental laws that govern the simple harmonic oscillator.

We can therefore expect a business-cycle model for a two-dimensional region to be capable of producing irregular and nonperiodic change, even though it is produced by the simple multiplier–accelerator mechanism. This is because the physical and economic models are equivalent in introducing second-order differential equations with constant coefficients. We note that the mere introduction of space yields completely new conclusions, even for the single point.

As is the case in acoustics, we have to deal with idealized systems. In particular, we will deal with rectangular and circular plane regions, and with the sphere (as an approximation to global trade-cycle diffusion in the world economy). The regions dealt with are assumed to be uniform in the sense that the economic reactions are the same in all parts of them. This is obviously unrealistic, since the impulses of change are communicated much more rapidly through locations provided with good transportation facilities than through deserts and forests. Likewise, the economic inertia of various parts is different in the sense that highly industrialized areas are not easily switched into modes of change from modes of rest, but once they are excited exert an enormous influence on the surrounding areas.

But the situation is quite similar in physics. To learn something about the vibrations of a real physical system, like a violin plate (an irregular curved shell of nonuniform thickness), uniform plane shells of regular shape or systems of similar simplicity are studied. Like physicists we can hope to learn something about reality from the study of idealized (unrealistic, but analyzable) systems.

8.2 THE BASIC MODEL

8.2.1 The Spaceless Original

As indicated above we will start out from the well known multiplier–accelerator model. However, we wish to deal with differential (not difference) equations, and therefore the appropriate formulation is not the original Samuelson–Hicks model, but a continuous equivalent due to Phillips. A complete discussion of it may be found in Allen (1956).

The model has two parts. First, aggregate demand is given by

$$Z = A + I + (1 - s)Y$$

where s denotes the rate of saving out of income Y (so that $(1 - s)Y$ is consumption), I denotes induced investments (i.e. those induced by the acceleration principle), and A denotes autonomous expenditures (autonomous investments, government expenditures, and the like). In equilibrium, aggregate demand Z would be equal to aggregate supply or income Y. Thus $Y = A + I + (1 - s)Y$, or in another form $sY = A + I$. The equivalent of this equilibrium condition in adaptive form with lagged response of savings would read

$$\dot{Y} = \lambda(A + I - sY) \tag{1}$$

This is the multiplier part of the model.

Second, the principle of acceleration states that capital stock be in a certain proportion to income, or, which amounts to the same thing, the rate of change of capital stock (investments) be in that proportion to the rate of change of income. Denoting the proportionality factor by v, the formal condition is $I = v\dot{Y}$. Again, we assume a lagged response in the change of capital stock from the actual to the optimal, and write in adaptive form

$$\dot{I} = \kappa(v\dot{Y} - I) \tag{2}$$

One way of handling these relations is to ignore the lagged response structure and start out from the equilibrium conditions. Then, deleting autonomous expenditures, $sY = v\dot{Y}$ is obtained. This is simply the Harrod–Domar model of balanced growth generating exponential growth at the rate s/v. Another way to use the relations is to take (1) and (2) together, differentiate (1) once more and then substitute for $\dot{I}$ from (2). Finally, we substitute from the original form of (1) for I and get

$$\ddot{Y} + (\lambda s + \kappa - \kappa\lambda v)\dot{Y} + \kappa\lambda sY = \kappa A + \lambda\dot{A}$$

The autonomous expenditures can, in principle, lead to forced vibrations of the system. In the original model, however, A is treated as a constant so that the time derivative is zero and a constant equilibrium $\bar{Y}$ can be obtained as a particular solution. Oscillatory deviations from this can be obtained from the solution of the homogeneous form, where the right-hand side is zero.

The homogeneous form, in fact, corresponds to a damped (or explosive) simple harmonic oscillator. The solution, as stated in the Introduction, is a simple sinusoidal wave with one definite period. And this is the construct we are going to study in a spatial format. Since the basic linearity of the model is not destroyed by the spatial generalization, the particular solution can always be added to the one obtained from the homogeneous form. Therefore we can, without loss of information, disregard autonomous expenditures in the discussion that follows.

8.2.2 Space and Trade Added

Our next step in making the model a spatial one is to introduce exports and imports. The simplest way is to treat them in the same way as consumption, i.e. assuming that imports are in a certain proportion to income. In this way exports too are determined by the incomes at other locations. Net exports should therefore be proportional to the difference between income abroad and at home. In traditional international economics "abroad" is a simple concept; it is just another abstract region without spatial dimensions. Once we introduce an economic space, however, things become more difficult.

To simplify matters we assume local action between spatially contiguous locations only. This is equivalent to the assumptions regarding actions in time. Interaction spatially with remote locations would be equivalent to a structure of distributed lags in the temporal interactions. But even with local actions only we have to be careful in defining a local income difference between a location and its immediate surroundings. If our region is one-dimensional and we are dealing with an interior point, we have one income difference to the right and another one to the left. Since in continuous space the income differences are spatial derivatives, the net difference between these, in the limit, equals the second spatial derivative. Denoting net export surplus by X and the space coordinate by x we get $X = m\,\mathrm{d}^2 Y/\mathrm{d}x^2$, where m is the propensity to import.

Our main interest, however, is two-dimensional space, and there we are faced with an infinity of directions in which we can move from an interior point to its immediate surroundings. To define a net change of income between a location and its immediate surroundings we can again use Gauss's theorem in the plane. Let us consider the divergence of the gradient of income

$$\operatorname{div} \operatorname{grad} Y = \partial^2 Y/\partial x_1^2 + \partial^2 Y/\partial x_2^2$$

This is also called the Laplacian and denoted $\nabla^2 Y$. Let us take the double integral of this over any bounded portion of the plane. Then, according to Gauss's theorem, we have

$$\iint \nabla^2 Y \mathrm{d}x_1 \mathrm{d}x_2 = \int (\operatorname{grad} Y)_n \mathrm{d}s$$

The right-hand side is a curve integral along the closed boundary of the region of integration on the left-hand side. The integrand, $(\operatorname{grad} Y)_n$, is the component normal to the boundary of the gradient of income. Therefore the integrand at each point of the boundary is the rate of change of income as we leave the enclosed region in a direction normal to the boundary, and the curve integral is the net change of income as we leave the enclosed area, all possible points of departure being taken into account. Let us now shrink the enclosed region to a point. In the limit we conclude that the Laplacian actually defines the net change of income as we pass from a point to its immediate surroundings, all possible directions leaving the point being considered. So, we define local export surplus as $X = m\nabla^2 Y$. Defining net exports in this way we can, by Gauss's theorem, see that trade is conserved in the model. What is exported from one location is always imported to another one. Hence, the model is consistent. We also note that in the one-dimensional case the Laplacian degenerates into the second derivative of the single spatial coordinate as conjectured above.

But so far we have only stated an equilibrium condition for trade. As was done for the multiplier and accelerator relations, it is logical to assume a delay in the adjustment of trade and introduce an appropriate adaptive process. We do this as Phillips and Allen would have done, by putting

$$\dot{X} = \mu(m\nabla^2 Y - X) \tag{3}$$

Together with (1) and (2) this defines the whole model. We will make a slight simplification before discussing the model further, by assuming that the delay coefficients κ, μ, and λ are equal. This corresponds to the way Samuelson and Hicks treat delays in their discrete models and has the advantage of keeping the order of the resulting differential equation relatively low. By doing this we can also define a unit of time so that these coefficients are unitary. This saves us a lot of symbols. As already mentioned, we also ignore autonomous expenditures.

Before proceeding, we must make a slight change in equation (1) and in its equilibrium counterpart. The latter becomes $A + I + X = sY$ in the "open" economy, and the corresponding adaptive relation becomes

$$\dot{Y} = \lambda(A + I + X - sY) \tag{1a}$$

which we will use from now in place of the original equation (1).

Then, differentiating (1a) with respect to time, using (2) and (3) to eliminate $\dot{I}$ and $\dot{X}$ in the resulting equation, and (1a) as it stands to eliminate I and X, we obtain

$$\ddot{Y} + (1 + s - v)\dot{Y} + sY = m\nabla^2 Y \tag{4}$$

which is the differential equation we are going to study. Before proceeding, note that we could utilize the trade condition in equilibrium form together with the other conditions in the same form to yield

$$\dot{Y} - (s/v)Y = -(m/v)\nabla^2 Y \tag{5}$$

which is the spatial counterpart of the Harrod–Domar balanced-growth model. The counterpart in physics to (5) is the heat diffusion equation, whereas the counterpart to (4) is the wave equation.

8.2.3 Solution by Separating Time and Space

In setting out to solve (4), we first introduce separation between the spatial and the temporal coordinates, i.e. we suppose that the solution can be written in the form $Y = T(t)S(x_1, x_2)$. Equation (4) can then be split in two parts, namely

$$T'' + (1 + s - v)T' + k^2 T = 0 \tag{6}$$

and

$$m\nabla^2 S + (k^2 - s)S = 0 \tag{7}$$

This splitting arises because, on introducing the attempted solution, the differential equation (4) breaks down into two different parts, which exclusively depend on the temporal and the spatial coordinates, respectively. The only way two expressions that depend on different independent coordinates can stay in a certain algebraic relation to one another is if they both equal some constant. This constant is arbitrary and has been introduced here as k^2.

The ordinary differential equation (6) is familiar from business-cycle theory. It has the solution

$$T = e^{-\alpha t}(A \cos \beta t + B \sin \beta t) \tag{8}$$

where

$$\alpha = (1 + s - v)/2 \tag{9}$$

and

$$\beta = (k^2 - \alpha^2)^{1/2} \tag{10}$$

We have a damped solution when $(1 + s - v)$ is positive, and an explosive solution when it is negative. The stability condition is exactly the same as in Hicks (1950).

The great difference, however, is that the period β of the simple harmonic motion represented by the trigonometric factor in (8) now depends on the arbitrary constant k^2. The latter is not determined by the structural coefficients of the model. It can take any value that suits us in a solution to the spatial equation (7), and when several k^2 values are possible, we need to take a sum of the harmonic functions of all the corresponding frequencies β according to (10) instead of the single harmonic. We remember at this point our introductory remarks that a one-dimensional region admits all natural multiples of the basic frequency, while a two-dimensional region admits both natural multiples and other frequency combinations. These facts now remain to be formally demonstrated.

8.3 EXAMPLES

We will now examine in more detail the behavior of our model for rectangular and circular plane regions and then for the curved surface of a two-dimensional sphere. In treating the rectangular case we will also cover the one-dimensional case in the limit where the rectangle becomes very long and narrow. Finally we will make some remarks on the general case of more irregular shapes.

8.3.1 The Rectangular Region

Our discussion will now concentrate on equation (7), which is a partial differential equation, not an ordinary one like (6). In the science of acoustics, (7) is usually called the Helmholtz equation. The degree of difficulty encountered in solving this equation depends on the shape of the boundary and the boundary conditions. For simple boundaries a suitable choice of coordinates, together with the separation method (already used to separate the temporal and spatial factors), will yield separate ordinary differential equations for the separate coordinates, and from these the solution can be simply compounded.

The natural starting point is the rectangular case, since our equations have been expressed in Cartesian coordinates from the outset. The region dealt with will be a rectangle with sides a and b. Suppose the boundary condition states that there is total income stability on the boundary of this rectangular region, which means that it is zero at all times. (Observe that zero here means zero deviation from the equilibrium value determined by the autonomous expenditures.) We will try the solution $S = X_1(x_1)X_2(x_2)$. Then equation (7) splits in two equations

$$X_1'' + (\pi i/a)^2 X_1 = 0 \tag{11}$$

and

$$X_2'' + (\pi j/b)^2 X_2 = 0 \tag{12}$$

where the new constants can be introduced, provided that they fulfil the condition

$$(\pi i/a)^2 + (\pi j/b)^2 = (k^2 - s)/m \tag{13}$$

We can thus choose i and j in any way that fulfils equation (13). As k^2 was linked to the temporal period of the variation we thus know that, by (9) and (10), i and j are related to β.

The solutions to (11) and (12) are

$$X_1 = \sin\frac{i\pi x_1}{a} \tag{14}$$

and

$$X_2 = \sin\frac{j\pi x_2}{b} \tag{15}$$

for any natural numbers i and j. The solutions are restricted to the natural numbers, and only the sines (not the cosines) appear, due to the boundary conditions that X_1 and X_2 be zero for $x_1 = 0$ and a and for $x_2 = 0$ and b, respectively.

The complete solution to (4) is then

$$Y = e^{-\alpha t} \sum_i \sum_j \sin\frac{i\pi x_1}{a} \sin\frac{j\pi x_2}{b} (A_{ij} \cos\beta t + B_{ij} \sin\beta t) \tag{16}$$

where the summations range over all natural numbers i and j from unity to infinity. In each of the terms β is determined by the combination of natural numbers. For convenience, we condense here the conditions (10) and (13)

$$\pi^2\left(\frac{i^2}{a^2} + \frac{j^2}{b^2}\right) = \frac{\alpha^2 + \beta^2}{m} - \frac{s}{m} \tag{17}$$

Remember that the α coefficient (the damping factor) is determined by the structure of the model according to (9), whereas the β coefficient (the temporal period) is determined by the spatial wavelengths according to (17). The damping factor is the same for all vibration frequencies.

We will illustrate the four lowest periods and the corresponding spatial configurations in terms of the nodal lines (lines of constant income). For illustrative purposes we assume that the region is square with $a = b = \pi$. Suppose the structural constants are $s = m = v = 0.25$. The result is a slightly damped motion with $\alpha = 0.5$, as we can see from (9).

For this special case (17) is $\beta = (i^2 + j^2)^{1/2}/2$. We can see that, even for this simple case, the sequence of successively shorter periods contains numbers that are not simple fractions of each other. For values of i, j equal to 1,1, 1,2 or 2,1, 2,2, and 1,3 or 3,1 the corresponding β values are $\sqrt{2}/2$, $\sqrt{5}/2$, $\sqrt{8}/2$, and $\sqrt{10}/2$, respectively. The implication is that whenever several of these modes coexist, the motion of income at a certain location is no longer strictly periodic.

We can also see from this example that cyclic variation of a given period is not always associated with any given spatial variation pattern in terms of the nodal lines. Obviously, the same period is obtained when we interchange the values of i and j. Moreover, any weighted sum of vibrations due to i, j and j, i also leads to the same period for the compound vibration. These considerations lead to the study of expressions of the type

$$C \sin ix_1 \sin jx_2 + D \sin jx_1 \sin ix_2 \tag{18}$$

We can get some ideas about the possible spatial variation modes associated with a given frequency by studying the four basic cases: $C = 0$, $D = 0$, $C + D = 0$, and $C = D$. This will be done for the four lowest frequencies enumerated above. First we note that if $i = j = 1$, then the two terms in (18) are equal and there are no nodal lines (except the edges of the square). The oscillations are greatest in the center and diminish as we get closer to the edges. All points of the square oscillate in phase. The frequency $\beta = \sqrt{2}/2 = 0.71$ is associated with this mode.

Next, for $i, j = 1,2$ and 2,1

$$C \sin x_1 \sin 2x_2 + D \sin 2x_1 \sin x_2 \; = \; \sin x_1 \sin x_2 (C \cos x_1 + D \cos x_2)$$

The factor outside the parentheses only gives the edges as nodal lines. Therefore we need only study the parenthetic term. If $C = 0$ there is an additional nodal line when $\cos x_2 = 0$, i.e. when $x_2 = \pi/2$. As the side of the square was given as π there is a horizontal nodal line halving the square. In the same way $D = 0$ gives a vertical nodal line bisecting the square. As $\cos(\pi - x_1) = -\cos x_1$ we see that $x_1 + x_2 = \pi$ results for the case where $C + D = 0$. This nodal line is a downward-sloping diagonal. Finally, $C = D$ obviously results in $x_1 = x_2$, which gives the other diagonal. The four modes are illustrated in Figure 8.1. For all of them we have $\beta = \sqrt{5}/2 = 1.12$. We see that the ratio of this frequency to the lowest frequency (corresponding to $i = j = 1$) is $1.12/0.71 = 1.58$. So, compounding only the two lowest frequencies destroys perfect periodicity.

The case of $i = j = 2$ is also easy to deal with. Then the trigonometric expressions in (18) are once again equal, and the relation of C to D does not matter. As $\sin 2x_1 \sin 2x_2$ takes zero value for $x_1 = \pi/2$ and for $x_2 = \pi/2$ (and on the edges, of course), there are both horizontal and vertical nodal lines through the center of the square. The square is divided by these lines into four smaller squares of equal size. The oscillation in two adjacent squares is always in opposite phase, so that when there is prosperity in the NE and SW squares there is depression in the SE and NW squares. This demonstrates the fact

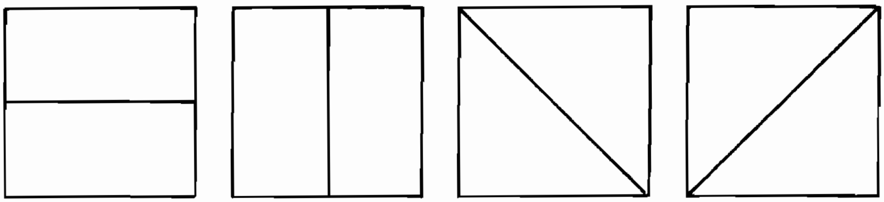

Figure 8.1. Basic oscillatory modes of a square (ratio of frequency to lowest one = 1.58).

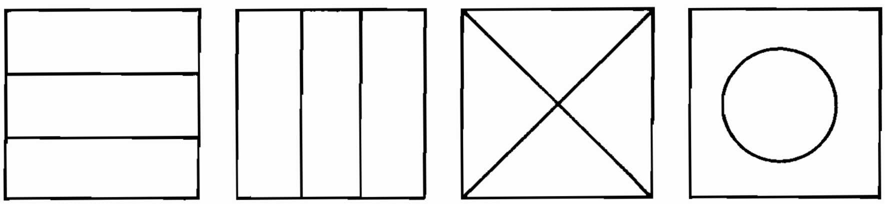

Figure 8.2. Basic oscillatory modes of a square (ratio of frequency to lowest one = 2.24).

that cyclic change in various parts of the region under study may be out of phase due to the friction in impulse transmission through space. The frequency associated with this mode is $\beta = \sqrt{8}/2$. The ratio of this to the lowest one is 2, so that we are actually dealing with a natural harmonic. This case is not illustrated since the reader may easily visualize it.

From this case we can make a general observation. When the square oscillates in parts separated by nodal lines then the frequency is higher than when it oscillates as a whole. In other words, cycles of long period may be expected to extend over large areas (all moving in phase), whereas fast cycles are confined to small areas (with the immediate surroundings moving in opposite phase).

Let us finish with the case of $i, j = 1{,}3$ and $3{,}1$. Then (18) becomes

$$C \sin x_1 \sin 3x_2 + D \sin 3x_1 \sin x_2 = \sin x_1 \sin x_2 (4C \cos^2 x_1 + 4D \cos^2 x_2 - C - D)$$

Again the multiplicative factor gives the edges as nodal lines, and hence only the parenthetic term is of interest. If $C = 0$, we get $\cos^2 x_2 = 1/4$, or $x_2 = \pi/3$ and $2\pi/3$ as the solution; i.e. there are two horizontal nodal lines dividing the square into equal strips. By the same reasoning, $D = 0$ gives three vertical strips of equal breadth. When $C + D = 0$, we get $\cos^2 x_1 = \cos^2 x_2$, which is satisfied when $x_1 = x_2$ and when $x_1 + x_2 = \pi$. Accordingly, both diagonals are nodal lines. Finally, if $C = D$, we get $\cos^2 x_1 + \cos^2 x_2 = 1/2$. This defines a closed curve that is almost but not quite circular. The four basic modes are illustrated in Figure 8.2.

The temporal frequency associated with all these modes is $\sqrt{10}/2 = 1.58$. Its ratio to the lowest frequency is $\sqrt{5} = 2.24$. Again we are dealing with a frequency ratio that is not rational. There is one interesting observation to make in this case. The frequency of temporal oscillation is not even associated with a given number of separately oscillating parts of the square. In Figure 8.2 we can see cases of two, three, and four areas into which the nodal lines divide the square.

The square is particularly simple to deal with. Even if not all the modes have frequencies in rational ratios, there exists the whole sequence of natural harmonics with twice, three times, . . . the basic frequency. These arise when the square is subdivided in 4, 9, . . . equal small squares.

For an oblong rectangle with sides whose lengths are not in rational proportions things are entirely different. The sequence of rational harmonics does not exist and it is generally impossible to compound a strictly periodic motion, even by selecting particular modes of vibration. An exception is when the rectangle becomes a very long and narrow strip. Then the system degenerates into the one-dimensional case and, as in (17), there remains only one ratio in the left-hand parenthetic term so that we are left with a series of natural harmonics.

We conclude this discussion by returning to the general case of the solution (16) and state a way in which the undetermined coefficients A_{ij} and B_{ij} may be calculated from the initial conditions. As any initial state is compatible with the solution, we are convinced of its full generality. By the orthogonality of the various sine and cosine functions we have

$$A_{ij} = \frac{4}{ab} \int\int Y_0 \sin \frac{i\pi x_1}{a} \sin \frac{j\pi x_2}{b} \, dx_1 \, dx_2 \tag{19}$$

and

$$B_{ij} = \frac{4}{ab\beta} \int\int Y_0' \sin \frac{i\pi x_1}{a} \sin \frac{j\pi x_2}{b} \, dx_1 \, dx_2 \tag{20}$$

where Y_0 is the income distribution at $t = 0$ and Y_0' is the rate of change of income at $t = 0$. Integration is over the whole rectangular region considered. When the system is initially at rest the B_{ij} coefficients vanish.

8.3.2 The Circular Region

The circular case can be dealt with more briefly, since it so much resembles the rectangular case just discussed. As a matter of fact, it is included merely in order to demonstrate that many of the conclusions have a greater generality than might otherwise be supposed. To deal with the circular disk efficiently we introduce polar coordinates, by the transformation

$$x_1 = \rho \cos \omega \tag{21}$$

and

$$x_2 = \rho \sin \omega \tag{22}$$

A little calculation then shows that

$$\nabla^2 S = \frac{\partial^2 S}{\partial x_1^2} + \frac{\partial^2 S}{\partial x_2^2} = \frac{\partial^2 S}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial S}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 S}{\partial \omega^2} \tag{23}$$

is the Laplacian in polar coordinates. The choice of coordinates already suggests the method of separation to be attempted. Assume a solution of the form

$$S = R(\rho)\Omega(\omega) \tag{24}$$

Equation (7) then splits again into two ordinary differential equations

$$\Omega'' + i^2\Omega = 0 \tag{25}$$

and

$$R'' + \frac{1}{\rho}R' + \left(\frac{k^2 - s}{m} - \frac{i^2}{\rho^2}\right)R = 0 \tag{26}$$

Equation (25) very much resembles equations (11)–(12) above. Its solution is any compound of sine and cosine functions with frequency i. Since we can thus represent the solution by a simple cosine function with a suitable phase lead, and as the phase shifts yield nothing of general interest, we represent the solution by

$$\Omega = \cos i\omega \tag{27}$$

In the present case i must be a natural number because of the angular character of the ω coordinate. Only then will (27) end up at the same value after a full round at $\omega = 2\pi$.

Equation (26) is Bessel's differential equation. The solutions are the two kinds of Bessel functions, much studied in the context of physical applications and tabulated in great detail. Only the Bessel functions of the first kind stay finite at the center of the disk. Since only this makes economic sense we will only use Bessel functions of the first kind as solutions to (26). The traditional notation for these functions is uppercase J subscripted by the "order" i. So

$$R = J_i(h\rho) \tag{28}$$

where

$$h^2 = (k^2 - s)/m \tag{29}$$

solve (26). As a boundary condition we again assume constant stability. So, (28) must vanish when ρ equals the radius of the disk that presently represents our region. Without loss of generality we can assume this radius to be unity. This assumption amounts to the choice of a unit of distance (in the same way that above we adopted a time unit that made the adjustment speeds unitary). With this convention, h must be so chosen that

$$J_i(h) = 0 \tag{30}$$

As in the rectangular case there is a uniform exponential damping (or antidamping) of all frequencies at a rate α, which is determined by the structural constants of the model according to (9). The frequencies for each mode i, j are determined by (29) and (10). For convenience, we condense these equations to

$$h^2 = \frac{\alpha^2 + \beta^2}{m} - \frac{s}{m} \tag{31}$$

Since α is fixed by the structural constants, (31) determines the temporal frequency β for

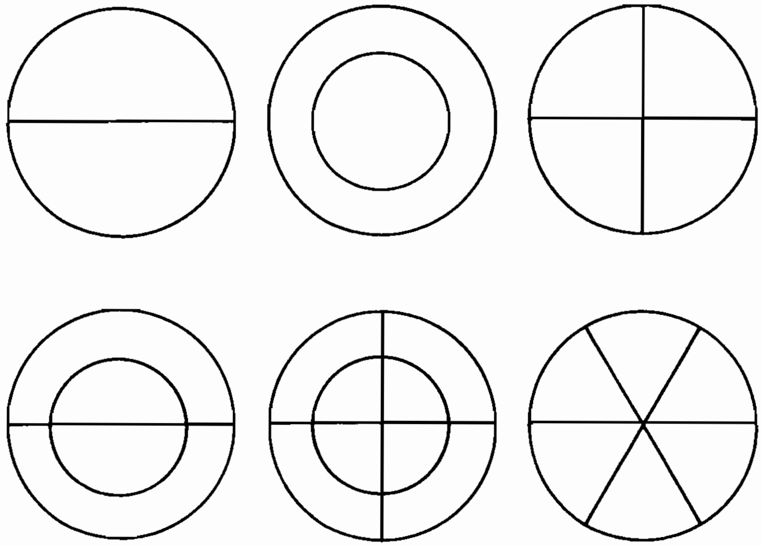

Figure 8.3. The lowest oscillating modes for a circular disk.

each h that is a solution to equation (30). And since the solutions (28) and (27) depend solely on the radial and angular coordinates, respectively, we can see that the nodal lines must be either concentric circles or equally spaced radials. At present there is no ambiguity concerning the spatial mode of oscillation that corresponds to a given temporal frequency, as was the case with a square region. Interchange of the subscripts i and j never results in the same β, because the polar coordinates do not have the symmetry of Cartesian coordinates.

In Figure 8.3 we illustrate the spatial modes of vibration associated with the six lowest frequencies (deleting the first one that arises when the whole disk oscillates in phase). The frequencies of the illustrated modes, expressed as ratios to the lowest one, are 1.59, 2.18, 2.30, 2.65, 2.92, and 3.16, in this order. These numbers result from again assuming $s = m = v = 0.25$ for the structural coefficients.

In the general case we can again determine the A_{ij} and the B_{ij} coefficients from initial conditions, since the Bessel functions, like the elementary trigonometric ones, are orthonormal. Multiplying the initial income distribution Y_0 by $J_i(h_{ij}\rho)$ and by $\cos i\omega$ and integrating over the whole disk we get A_{ij}. If we substitute Y_0' for Y_0 we get B_{ij}. (In order to obtain the actual values we have to divide by certain constants as in (19)–(20).)

8.3.3 The Spherical Region

The idealized simple regions examined so far have been approximations to, at most, single countries or continents. As a first approximation to world-wide trade and cycle propagation we will now study the spherical shell. For simplicity we can assume the radius to be unity. Again, this only amounts to a convenient choice of distance unit. For the circular region we found it *convenient* to introduce curvilinear coordinates (polar ones). For the present case, as we are dealing with a curved surface, we *must* use curvilinear coordinates. No Cartesian ones will do.

The simplest coordinate system for the present case is the spherical, given by the angles of longitude and colatitude. Suppose the sphere is embedded in three-dimensional Euclidean space with Cartesian coordinates ξ, η, ζ. Denoting the spherical coordinates θ, ϕ we have the transformation

$$\xi = \sin\theta\cos\phi \tag{32}$$

$$\eta = \sin\theta\sin\phi \tag{33}$$

$$\zeta = \cos\theta \tag{34}$$

Computation shows that the Laplacian in these coordinates equals

$$\nabla^2 S = \cos\theta\frac{\partial S}{\partial\theta} + \frac{\partial^2 S}{\partial\theta^2} + \frac{1}{\sin^2\theta}\frac{\partial^2 S}{\partial\phi^2} \tag{35}$$

This expression is the one to be inserted into the Helmholtz equation (7). The appropriate separation of variables in the solution is obvious. Let us try $S = \Theta(\theta)\Phi(\phi)$. The result, as before, is that our partial differential equation breaks down into two ordinary ones

$$\Theta'' + \cos\theta\,\Theta' + \left(j(j+1) - \frac{i^2}{\sin^2\theta}\right)\Theta = 0 \tag{36}$$

and

$$\Phi'' + i^2\Phi = 0 \tag{37}$$

For convenience we define a new constant

$$j(j+1) = \frac{k^2 - s}{m} \tag{38}$$

From (9) we see that hence

$$j(j+1) = \frac{\alpha^2 + \beta^2}{m} - \frac{s}{m} \tag{39}$$

which relates j to the temporal frequency β, everything else being determined by the structural constants of the model.

The solution to (37) is obvious. Any compound of sine and cosine functions of periodicity i will do. As in the case of the circular plane disk, the solution must end up at the starting value after a full round, and therefore the argument of the trigonometric functions must be $i\phi$, where i is a natural number. Since compounding sines and cosines, or, equivalently, introducing phase leads, does nothing but rotate the whole set of nodal lines, we can without loss of generality put

$$\Phi = \cos i\phi \tag{40}$$

This solves (37) for any integral value of i. Equation (36) may seem more complicated, but as a matter of fact it is Legendre's associated differential equation. Solutions to this are the Legendre functions of the first and second kinds. Since the latter take on infinite values only the first ones make economic sense. They are traditionally denoted as $P_j^i(\cos\theta)$, and there is a simple analytical expression

$$P_j^i(\cos\theta) = \frac{\sin^i\theta}{2^j j!}\frac{\mathrm{d}^{i+j}(\sin^{2j}\theta)}{\mathrm{d}(\cos\theta)^{i+j}} \tag{41}$$

These functions have been thoroughly studied and tabulated in most mathematical handbooks. One thing to note from definition (41) is that the functions obviously become zero whenever i exceeds j.

Accordingly we have

$$\Theta = P_j^i(\cos\theta) \tag{42}$$

as the solution to (36) and can compound the general solution to (4) from the solutions (40), (42), and (8). Thus

$$Y = \mathrm{e}^{-\alpha t}\sum_{i\leqslant j}\sum P_j^i(\cos\theta)\cos i\phi(A_{ij}\cos\beta t + B_{ij}\sin\beta t) \tag{43}$$

where the temporal frequency is determined from (39) for each spatial mode of vibration. As before the whole system of oscillations is uniformly damped or antidamped (due to the Hicksian condition for the structural coefficients).

We will illustrate some basic modes of oscillation by studying the nodal line systems for cases with low i and j. (For these, the expressions for P_j^i are tabulated below. The formulas can be derived quite easily from the definition (41).) We note that expressions are only tabulated for i at most equal to j, since otherwise the P_j^i become identically zero.

i	j	$P_j^i(\cos\theta)\cos i\phi$
0	0	1
0	1	$\cos\theta$
1	1	$\sin\theta\cos\phi$
0	2	$3\cos^2\theta - 1$
1	2	$\cos\theta\sin\theta\cos\phi$
2	2	$\sin^2\theta\cos 2\phi$
0	3	$5\cos^3\theta - 3\cos\theta$
1	3	$\sin\theta(5\cos^2\theta - 1)\cos\phi$
2	3	$\cos\theta\sin^2\theta\cos 2\phi$
3	3	$\sin^3\theta\cos 3\phi$

Two facts from this table are noteworthy. First, if $i = 0$, the tabulated expression only involves the coordinate θ. Second, if $i = j$, then, except for factors that are powers of $\sin\theta$, the expressions only involve the coordinate ϕ. As $\sin\theta$ only produces zeros at the poles of the sphere, the corresponding nodal lines are degenerate points (at the

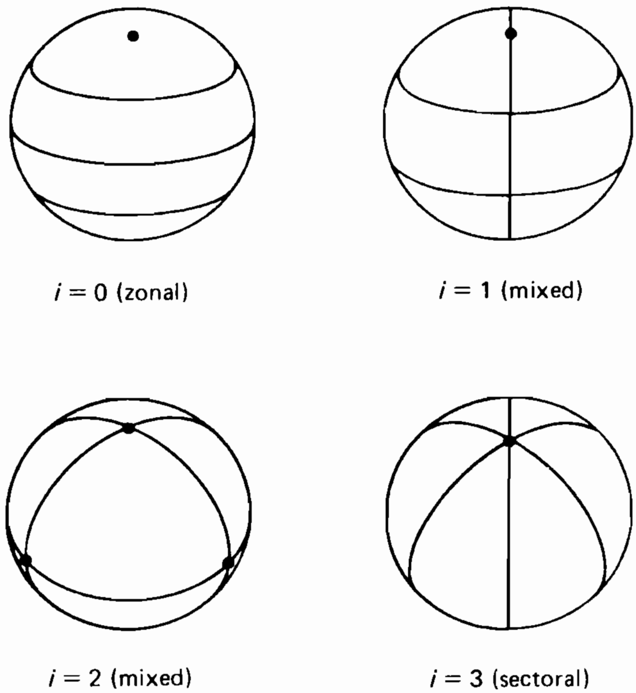

Figure 8.4. Oscillating modes of a sphere ($j = 3$, $\beta = 1.73$).

intersections of other nodal lines). Accordingly $i = 0$ produces nodal lines parallel to the equator, whereas $i = j$ produces nodal lines that are great circles through the poles. The modes associated with these cases are called "zonal" and "sectoral", respectively. The intermediate cases where $0 < i < j$ represent the "mixed" modes.

In Figure 8.4 we illustrate the different modes. We choose the case of $j = 3$, which out of the cases tabulated yields the richest set of possibilities. If $i = 0$, we get nodal lines for $\cos\theta = 0$ and for $\cos\theta = \pm\sqrt{(3/5)}$, i.e. for $\theta = 39°$, $90°$, and $141°$. So, the nodal lines are the equator and two parallel circles. This zonal mode is the first case illustrated.

Next, let $i = 1$ (and $j = 3$ still). Once again, as we can see from the table, the nodal lines are produced by $\sin\theta = 0$, $\cos\theta = \pm\sqrt{(1/5)}$, and $\cos\phi = 0$. The middle possibility gives $\theta = 63°$ and $117°$, i.e. two parallel circles, whereas the third one gives $\phi = 0°$, i.e. a great circle through the poles. The first possibility is degenerate, giving $\theta = 0°$ and $180°$. These are the poles, which we already have through the great circle. The corresponding mixed mode has two parallel circles and one polar great circle.

Letting $i = 2$, we read from our table that $\cos\theta = 0$ and $\cos 2\phi = 0$ define the nodal lines. The first gives $\theta = 90°$, which is the equator, whereas the second gives $\phi = 45°$ and $135°$, i.e. two polar great circles. The remaining possibility, with $\sin^2\theta = 0$, can be discarded from the outset, since it yields only the poles, which we already have twice by the great circles. Thus, the mixed mode has the equator and two polar great circles as nodal lines.

There remains the case with $i = j = 3$. Again discarding $\sin^3\theta = 0$, we get the nodal lines from $\cos 3\phi = 0$. Equivalently $\phi = 30°$, $90°$, and $150°$. These define three polar great circles, dividing the sphere into six sectors. This, finally, illustrates the purely sectoral mode.

We obviously have a great variety of possibilities as to the spatial modes when only j is given. In addition to the four cases illustrated we can of course combine them as we did in the section on the quadratic plane region. As there are four modes to combine (in contrast to the two resulting from interchanging the subscripts i and j in the quadratic case), even more combinations are possible.

It should be borne in mind that, due to (39), only the value of j (and not that of i) is related to the temporal frequency β. So, all the four cases illustrated (and also all linear combinations thereof) have the same period of oscillation. If, for instance, we again have the structural coefficients $s = m = v = 0.25$, as in the above illustrations, then from (39) we get $3 \cdot 4 = 4\beta^2$. Hence $\beta = \sqrt{3} = 1.73$.

We should also remember that the oscillations of different periods can be combined linearly in any way, which results in a superposition of the various corresponding spatial oscillation modes, and in the composite temporal motion having no perfect periodicity. This is due to the fact that the sequence of $\beta = (j(j+1))^{1/2}$ for ascending integral j is not one of integers.

Nevertheless, it is possible to calculate the A_{ij} and B_{ij} from initial conditions as in the previous cases. For the general solution we have

$$A_{ij} = \frac{(j-i)!}{(j+i)!}\,\frac{2n+1}{\pi} \iint Y_0 P_j^i(\cos\theta) \cos i\phi \,\mathrm{d}\theta\,\mathrm{d}\phi \tag{44}$$

$$B_{ij} = \frac{(j-i)!}{(j+i)!}\,\frac{2n+1}{\pi} \iint Y_0' P_j^i(\cos\theta) \cos i\phi \,\mathrm{d}\theta\,\mathrm{d}\phi \tag{45}$$

due to the orthogonality of the Legendre functions. In fact, (44)–(45), as in the previous cases, allow the representation of any initial distribution of income and rate of change of income over the region. So, the solution is again perfectly general.

With this example we conclude our set of illustrations. For more complicated shapes of regions there exist approximate methods of computing the eigenvalues. On this topic, and on much of the other mathematics used in the present chapter, Sneddon (1957) or Courant and Hilbert (1953) can be usefully consulted. Other valuable references are Duff and Naylor (1966) and Rayleigh (1945).*

8.4 THE REGIONAL GROWTH MODEL

Before summing up the discussion, let us say something about the spatial variant of the Harrod–Domar growth theory. Thus we are dealing with (5) rather than (4). Let us try separation of variables again. Put $Y = T(t)S(x_1,x_2)$. The result is two equations

$$T' - \left(k + \frac{s}{v}\right) T = 0 \tag{46}$$

*Lord Rayleigh, *Theory of Sound*, Vols. 1 and 2, MacMillan, London, 1894; reprinted as a Dover paperback, New York, 1945.

and

$$\nabla^2 S + k \frac{v}{m} S = 0 \tag{47}$$

The solution to (46) is

$$T = e^{(s/v+k)t} \tag{48}$$

multiplied by any arbitrary constant, which we ignore here. Equation (47) will be discussed for the case of a rectangular region only. Separation of variables by $S = X_1(x_1) X_2(x_2)$ again splits the equation. Denoting the sides of the rectangle by a and b, we get

$$X_1'' + \left(\frac{i\pi}{a}\right)^2 X_1 = 0 \tag{49}$$

and

$$X_2'' + \left(\frac{j\pi}{b}\right)^2 X_2 = 0 \tag{50}$$

where

$$\pi^2 \left(\frac{i^2}{a^2} + \frac{j^2}{b^2}\right) = k \frac{v}{m} \tag{51}$$

Solutions to (49)–(50) are any sine or cosine functions of frequency i and j, respectively. If we assume again that income does not change on the boundary, we can forget all about phase leads and use the pure sine functions. So

$$X_1 = \sin \frac{i\pi x_1}{a} \tag{52}$$

and

$$X_2 = \sin \frac{j\pi x_2}{b} \tag{53}$$

are solutions. We can compound (52) and (53) with (48) to obtain the solution for any pair of i and j that are integers and a constant k that satisfies (51) for these. The general solution is a weighted sum of all these solutions.

The rectangular area is thus subdivided into ij small rectangles in which S alternates between positive and negative values. This makes Y alternate in the same manner, which might seem absurd, but remember that Y is the deviation from the level of income induced by autonomous expenditures. So, the alternation means that income alternates between values above and below the values it would locally have in stationary equilibrium. The discrepancy that initially exists between the actual and the equilibrium values of income grows everywhere at the rate $s/v + k$. So to the Harrod–Domar growth rate there is added a growth rate k, which depends on interregional trade.

This additional growth rate is higher, the higher the values of i and j, as we can see from (51). The process hence goes faster, the finer the mesh of spatial subdivision. Observe that even the Harrod–Domar model with autonomous expenditures generates an exponential "growth" downwards when the initial income is lower than the equilibrium value obtained by applying the multiplier to autonomous expenditures alone. Actually, it is the discrepancy between the initial and the stationary equilibrium values that grows exponentially.

In our spatial variant of the growth model with trade there is an additional growth rate due to the dynamics from interregional trade. If, spatially, locations with initial income above the equilibrium value alternate with locations where the reverse holds true, the additional growth rate is higher the more inhomogeneous space is in this respect. Actually it is proportional to the square root of the sum of squares of the horizontal and vertical subdivisions.

In addition, the additional growth rate is proportional to the ratio m/v, as we see from (51). This is similar to the case of the original Harrod–Domar model, where the natural growth rate is s/v, i.e. the ratio between the propensity to save and the accelerator. This ratio appears in our model as well. The additional growth rate is proportional to the ratio between the propensity to import and the accelerator. This seems logical since imports, like savings, constitute a "leakage" of expenditures.

Let us finally state the general solution, where different growth rates are compounded together

$$Y = \sum \sum A_{ij} \mathrm{e}^{(g+k)t} \sin ix_1 \sin jx_2 \tag{54}$$

where we have put $g = s/v$ as a convenient abbreviation. Observe that the spatial model, by combining as many growth rates as we wish, can generate any pattern of irregular growth, not just one single exponential growth rate. This is an advantage of the spatial version. The coefficients can be determined from the initial income distribution by again using the orthogonality of the trigonometric functions. So

$$A_{ij} = \int\int Y_0 \sin ix_1 \sin jx_2 \, \mathrm{d}x_1 \mathrm{d}x_2 \tag{55}$$

By Fourier's analysis any initial income distribution can be expressed in the form (54) by an appropriate choice of the A_{ij} coefficients.

The reader who feels uneasy about the very precise conclusion that interregional trade in the growth process tends to speed up the process of ever-increasing regional inequality should keep one thing in mind. All linear growth models, whether spatial or not, sooner or later become unacceptably unrealistic. With business-cycle models, where the oscillatory movement is the primary result, the variables tend to stay within certain bounds for a sufficiently long period to justify the assumption of constant coefficients, at least as a first approximation. But a model generating exponential growth, by its very nature, sooner or later violates any linear approximation, however generous a tolerance we admit. This is why we regard the business-cycle dynamics as the primary contribution of this chapter. It should, however, be borne in mind that the reservations above also apply to business-cycle models in cases where the exponential part of the solution creates explosive cycles.

8.5 A MORE GENERAL CASE

The cases of rectangular, circular, and spherical regions examined above were not meant to be anything more than examples. The general conclusions drawn, however, remain true for any region with a sectionally differentiable boundary curve and a homogeneous boundary condition. Such a homogeneous boundary condition may, for instance, state that the deviation of income from the equilibrium is zero on the boundary, or, more generally, that it displays a certain relation to its rate of change.

We can even dispense with homogeneity, as a nonhomogeneous boundary condition can always be translated into a nonhomogeneity of the differential equation. We already have one nonhomogeneity, i.e. the autonomous expenditures, and can thus absorb the one translated from a nonhomogeneous boundary condition into the one that already exists.

At the same time we will make another generalization. Trade surplus was supposed to result from income differences to which a constant import propensity was applied. We had X adapted to $m\nabla^2 Y$. A more general version is obtained by replacing $m\nabla^2 Y$ by $\operatorname{div}(m \operatorname{grad} Y)$, with $m(x_1, x_2)$ being a function of location. This is interpreted as meaning that there is a net trade flow proportional to the gradient of income. The divergence of the flow equals local export surplus.

As long as m is a spatial constant the resulting equations are not changed, but once it is location dependent we get

$$\operatorname{div}(m \operatorname{grad} Y) = \operatorname{grad} m \cdot \operatorname{grad} Y + m\nabla^2 Y \tag{56}$$

This differs from our previous $m\nabla^2 Y$ in two respects. First, m is now a function of the location coordinates. Second, there is an additional term arising from projecting the gradient of income on the direction of the gradient of m. As the main reason for spatial variation of the propensity to import is the variation of transportation facilities, we see that the present generalization introduces inhomogeneous space with local differences in the facility of transportation.

Mathematically, we arrive at the two-dimensional version of the Sturm–Liouville problem (which originally dealt with vibrations of a nonhomogeneous string). The general theory for this case is well developed. Adapting to the business-cycle case we obtain

$$\operatorname{div}(m \operatorname{grad} S) + (k^2 - s)S = 0 \tag{57}$$

to replace (7), whereas (6) remains the same as before. (Observe that s, unlike m, must still be a spatial constant as the separation of variables would otherwise not work.)

For clarity, let us define the linear, self-adjoint, differential operator

$$L(S) = \operatorname{div}(m \operatorname{grad} S) - sS \tag{58}$$

so that we have an eigenvalue problem defined by

$$L(S) + k^2 S = 0 \tag{59}$$

for an arbitrary region with a sectionally smooth boundary. As the boundary condition we have $S = 0$, which means that any inhomogeneity in the original boundary condition has been absorbed into the corresponding inhomogeneous equation obtained by adding the autonomous expenditures.

It is well known (Courant and Hilbert 1953) that (59) has a denumerable infinity of solutions for a sequence of eigenvalues $k_1^2, k_2^2, \ldots$, and that there correspond solutions $S_1(x_1, x_2), S_2(x_1, x_2), \ldots$, to each of these eigenvalues. Moreover, the solutions can be normalized so that they are orthogonal and fulfill the conditions

$$\iint S_i S_j \, \mathrm{d}x_1 \, \mathrm{d}x_2 = \delta_{ij} \tag{60}$$

Accordingly, it is possible to represent the general solution by

$$Y = \mathrm{e}^{-\alpha t} \sum (A_i \cos \beta_i t + B_i \sin \beta_i t) S_i(x_1, x_2) \tag{61}$$

where α is defined by the structural coefficients as in (9) and the β_i are defined from the eigenvalues k_i by (10).

Owing to the orthogonality of the eigenfunctions (60), the solution is perfectly general and the constants of the series expansion can be computed from

$$A_i = \iint Y_0 S_i \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{62}$$

and

$$B_i = \iint \dot{Y}_0 S_i \, \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{63}$$

This general case in fact covers all the particular cases dealt with, since the sines, cosines, Bessel functions, and Legendre polynomials are all sets of orthogonal eigenfunctions presently denoted by $S_i(x_1, x_2)$. The reader may wonder why we now use a single sequence of solutions, whereas in all the examples the solutions were obtained as double sums. We could, of course, have defined the set of solutions as $S_{ij}(x_1, x_2)$, but this would only be meaningful if we could work out some further separation of coordinates by putting $S_{ij} = S_i^1(\xi_1) S_j^2(\xi_2)$ for some suitable change of coordinates $x \rightarrow \xi$. It must, however, be remembered that the separation of variables is not always possible, and the reasoning above is applicable to cases much more general than those for which the possibility of separation exists.

In this context we should also record a useful method of approximating the eigenvalues. By starting from a set of functions that need not be the real eigenfunctions for the problem but that have the useful property of vanishing on the boundary, it is possible to get fairly good approximations of the eigenvalues; see Weinstock (1974).

We have been specifically discussing business cycles, but an analogous line of reasoning is applicable in the growth model as well. Thus, our general conclusions hold for regional shapes that are as general as we wish to deal with, as long as the boundary is sectionally smooth. (From this we only rule out boundaries of the Mandelbrot type that tend to be space filling.)

8.6 CONCLUSION AND PERSPECTIVES

The analysis presented in this chapter has been based on a spatial generalization of the multiplier–accelerator models of economic growth and business cycles to a spatial variant with interregional trade. The models chosen as appropriate starting points were the continuous variants due to Harrod and to Phillips, respectively, though all of those that emerged from the "golden age" of the growth and cycle models were based on the same principles and had very similar results.

The principal difference betweeen the results of the spatial and spaceless models is the irregular and nonperiodic temporal patterns of economic change that the former can generate. For this we need not complicate the delay structure, only introduce space. (In a very formal sense we do, of course, introduce a distributed lag structure, since in bounded space the delays in propagation of impulses reflected back later on are in fact equivalent to distributed lags. Through spatial friction the history of economic development at a given location does continue to exert an influence over a longer period.)

Obvious generalizations concern the removal of the linearity assumptions, so that such phenomena as unlimited growth disappear. The same is true of the spatial uniformity that was assumed even in cases where world trade over the whole globe was implied. However, any discussion must necessarily start with models as simple as those that we have introduced.

One interesting exercise would be to use statistical spectral analysis to implement the model for economic activity over a vast region. There already exist time series of indicators of economic activity for sufficiently many locations to make such an analysis possible.

Finally, it is clear that a generalization to a more complicated metric, where impulses spread very fast to centers all over the world, and diffuse more slowly from there, would be most desirable and timely. Some impulses of economic change, like a stock exchange collapse at a major economic center, obviously spread very fast, whereas changing orders through chains of enterprises responsible for processing at various stages of the refinement of a product propagate in a more gradual way for which the model of local action is more realistic.

9 *Conclusion*

The continuous flow model has emphasized theoretical analysis rather than computation. In the tradition of economic theory the discussion has been on general structural properties rather than numerical results. This is in line with classical land-use and location theory, but not with recent trends in regional science where discrete models based on graph theory have superseded the traditional geometrical ones. These recent developments in regional science seem to be due to the evolution of computers and algorithms, as discretization is a precondition for access to these facilities. However, there is the drawback that the intuitively appealing picture of shapes in geometrical two-space disappears with discretization of space.

We do not suggest the continuous model as a competitor to the discrete one, but rather as a companion. One need not choose between the two approaches on the basis of which one provides the most "realistic" description of real-world observations. Reality can always be described at different microscopic or macroscopic levels, and it is only a matter of convenience which description we choose for a given purpose.

This point is beautifully illustrated by Benoit Mandelbrot (1977) in his discussion of what a ball of thread is. *"Indeed, at the resolution possible to an observer placed 10 m away, it appears as a point, that is, as a zero-dimensional figure. At 10 cm it is a ball, that is, a three-dimensional figure. At 10 mm it is a mess of threads . . . At 0.1 mm each thread becomes a sort of column . . . At 0.01 mm resolution, each column is dissolved into filiform fibers . . .* " So, each system in reality can be conceived in different ways; on certain levels, reality seems to be more or less chaotic, but on others, it may lend itself to systematic description and modeling involving different kinds of abstraction and idealization.

Mandelbrot's argument could easily be transferred to a system of physical matter, or to a regional economy, among many other systems. A fluid can be described as a collection of particles whose configuration is governed by a huge set of ordinary interrelated differential equations, or as a continuum that evolves according to a few partial differential equations.

Likewise, in the regional economy, we can conceive of the system at the level of a detailed map, where each piece of land is used for some particular purpose and the roads represent some given network, keeping track of each vehicle as it traverses the arcs of the network. Taking a more macroscopic view, we can disregard the anisotropic character of roads at each street corner, smoothing out the kinks of flows, and regard different land-uses in terms of densities that, in principle, can be combined. It is this last viewpoint that we have adopted throughout most of this book.

Though discrete and continuous modeling are complementary rather than competitive approaches, our view is that the former has dominated the market to the point of excess. The aim of this book has been to stimulate renewed interest in continuous-space modeling.

One of the reasons for the sparsity of recent literature in this vein may be that the classical location theorists have exploited almost everything that can be done with Euclidean geometry (using the assumption that space is homogeneous with respect to transportation cost). Once constant cost is abandoned the possibilities become very rich, in fact, so rich that specific structures tend to have illustrative character, rather than the logical necessity characteristic of the location classics.

One way out of this dilemma has been hinted at in the book, viz. the use of topology instead of Euclidean geometry and the principle of structural stability as an additional modeling principle. The discussion by no means exhausts all that can be done using these instruments. We only intend to provide starting points for exploration of the field.

This is also true for the book as a whole. We cannot claim to have exhausted the possible uses of continuous flow models in spatial economics. Rather, our purpose has been to stimulate a wider audience to adopt the continuous flow approach to spatial economic phenomena.

Appendix

DERIVATION OF THE DIVERGENCE LAW

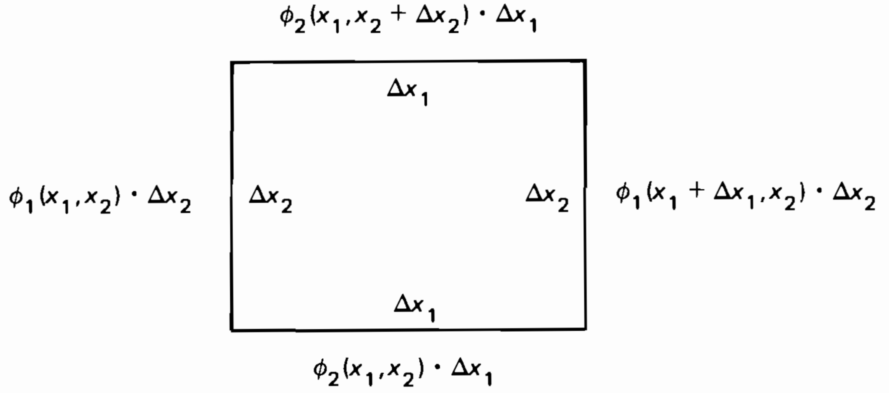

Figure A.1. Net outflow of a cell $(\Delta x_1, \Delta x_2)$.

In Figure A.1, the inflow into and outflow from a rectangular cell of side lengths Δx_1, Δx_2 are broken down into horizontal and vertical components:

Horizontal inflow	$\phi_1(x_1, x_2)\Delta x_2$
Horizontal outflow	$\phi_1(x_1 + \Delta x_1, x_2)\Delta x_2$
Vertical inflow	$\phi_2(x_1, x_2)\Delta x_1$
Vertical outflow	$\phi_2(x_1, x_2 + \Delta x_2)\Delta x_1$

The difference of outflow over inflow is therefore

$$[\phi_1(x_1 + \Delta x_1, x_2) - \phi_1(x_1, x_2)]\,\Delta x_2 + [\phi_2(x_1, x_2 + \Delta x_2) - \phi_2(x_1, x_2)]\,\Delta x_1 \quad \text{(A1)}$$

We assume the flow field $\phi(\underset{\sim}{x})$ is continuously differentiable. For the flow variables a Taylor expansion can then be used

$$\phi_1(x_1 + \Delta x_1, x_2) = \phi_1(x_1, x_2) + \frac{\partial \phi_1}{\partial x_1} \cdot \Delta x_1 + 0(\Delta x_1^2)$$

$$\phi_2(x_1, x_2 + \Delta x_2) = \phi_2(x_1, x_2) + \frac{\partial \phi_2}{\partial x_2} \cdot \Delta x_2 + 0(\Delta x_2^2)$$

Substituting this in (A1) and dropping all terms in $0(\Delta x_1^2)$ and $0(\Delta x_2^2)$, the net outflow becomes

$$\left[\frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2}\right] \Delta x_1 \cdot \Delta x_2$$

Then net outflow equals the net supply of fluid in the area $\Delta x_1 \cdot \Delta x_2$, which is

$$-q(x_1, x_2)\Delta x_1 \Delta x_2$$

Therefore

$$\frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2} + q = 0 \tag{A2}$$

or

$$\operatorname{div} \phi + q = 0$$

Bibliography

A. LOCATION CLASSICS

Christaller, W. (1935) *Die zentralen Orte in Süddeutschland*. Fischer, Jena; translated into English as Christaller, W. (1966) *Central Places in Southern Germany*. Prentice-Hall, New York.

Hotelling, H. (1927, 1978) A Mathematical Theory of Migration. M.A. thesis reprinted in *Environment and Planning A* 10, 1223–1239.

Hotelling, H. (1929) Stability in Competition, *Economic Journal*, 39: 41–57.

Launhardt, W. (1869) *Rentabilität und Richtungsfestellung der Strassen*. Hannover.

Launhardt, W. (1882) Der beste Standort einer gewerblichen Anlage. *Zeitschrift des Vereins deutscher Ingenieure*, 26: 106–115.

Launhardt, W. (1885) *Mathematische Begründung der Volkswirtschaftslehre*. B.G. Teubner, Leipzig. Reprinted (1945) by Scientia Verlag, Aalen.

Lösch, A. (1954) *The Economics of Location*; translation from the German *Die räumliche Ordnung der Wirtschaft*. Yale University Press, New Haven, Connecticut. (German editions 1940, 1944.)

Palander, T.F. (1935) *Beiträge zur Standortstheorie*. Almqvist and Wiksell, Uppsala.

Ponsard, C. (1983) *History of Spatial Economic Theory*. Texts and Monographs in Economics and Mathematical Systems. Springer-Verlag, Berlin.

von Stackelberg, H. (1938) Das Brechungsgesetz des Verkehrs. *Jahrbücher für Nationalökonomie und Statistik*, 148: 680–696.

von Thünen, H. (1921) *Der isolierte Staat*. Waentig, Jena.

Weber, A. (1909) *Über den Standort der Industrien*. J.C.B. Mohr, Tübingen; translated into English as Weber, A. (1929) *Alfred Weber's Theory of Location of Industries*. University of Chicago Press, Chicago, Illinois.

B. CURRENT LITERATURE ON SPATIAL ECONOMICS AND THEORETICAL GEOGRAPHY

Alonso, W. (1965) *Location and Land Use*. Harvard University Press, Cambridge, Massachusetts.

Alonso, W. (1967) A Reformulation of Classical Location Theory and its Relation to Rent Theory. *Papers and Proceedings of the Regional Science Association*, 26: 85–99.

Angel, S. and Hyman, G.M. (1970) Urban Velocity Fields. *Environment and Planning*, 2: 211–224.

Angel, S. and Hyman, G.M. (1971) Urban Travel Time. *Papers and Proceedings of the Regional Science Association*, 26: 85–99.

Angel, S. and Hyman, G.M. (1972a) Urban Transportation Expenditures. *Papers and Proceedings of the Regional Science Association*, 29: 105–123.

Angel, S. and Hyman, G.M. (1972b) Urban Spatial Interaction. *Environment and Planning*, 4: 99–118.

Angel, S. and Hyman, G.M. (1976) *Urban Fields – A Geometry of Movement for Regional Science*. Pion, London.

Beckmann, M.J. and Ingene, C.A. (1976) The Profit Equivalence of Mill and Uniform Pricing Policies. *Regional Science and Urban Economics*, 6: 327–329.

Beckmann, M.J., McGuire, C.B., and Winsten, C.B. (1956) *Studies in the Economics of Transportation*. Yale University Press, New Haven, Connecticut.

Bos, H.C. (1965) *Spatial Dispersion of Economic Activity*. University Press of Rotterdam. Rotterdam.

Brakke, K.A. and Zitron, N.R. (1977) Further Optimal Cost Routes in the Circular City Model. *Journal of Optimization Theory and Applications*, 21: 531–534.

Bunge, W. (1962) *Theoretical Geography*. C.W.K. Gleerup, Lund.

Bunge, W. (1964) *Patterns of Location*. Discussion Paper 3, Michigan Inter-University Community of Mathematical Geographers, Ann Arbor, Michigan.

Dafermos, S.C. (1976) Integrated Equilibrium Flow Models for Transportation Planning. *Lecture Notes in Economics and Mathematical Systems*, 118. Springer-Verlag, New York.

Dafermos, S.C. (1980) Continuum Modelling of Transportation Networks. *Transportation Research*, 14B: 295–301.

Dendrinos, D.S. (1980) *Catastrophe Theory in Urban and Transport Analysis*. U.S. Dept. of Transportation, Washington D.C.

Dunn, E.S., Jr. (1954) *The Location of Agricultural Production*. University of Florida Press, Gainsville, Florida.

Erlenkotter, D. (1978) A Dual Based Procedure for Uncapacitated Facility Location. *Operations Research*, 26: 992–1009.

Fales, R.L. and Moses, L.N. (1972) Land Use Theory and the Spatial Structure of the Nineteenth-Century City. *Papers of the Regional Science Association*, 28: 49–80.

Ford, L.R. and Fulkerson, D.R. (1962) *Flows in Networks*. Princeton University Press, Princeton, New Jersey.

Fox, K.A. and Taeuber, R.C. (1955) Spatial Equilibrium Models of the Livestock-Feed Economy. *American Economic Review*, 45: 584–608.

Greenhut, M.L. and Ohta, H. (1975) *Theory of Spatial Pricing and Market Areas*. Duke University Press, Durham, North Carolina.

Griffith, D.A. and MacKinnon, R.D. (1981) *Dynamic Spatial Models*. Sijthoff and Noordhoff, Alphen, The Netherlands.

Griffith, D.A. and Lea, A.C. (1983) *Evolving Geographical Structures*. Nijhoff, The Hague.

Haggett, P. (1965) *Locational Analysis in Human Geography*. Edward Arnold, London.

Henderson, J.M. (1955) A Short-Run Model for the Coal Industry. *Review of Economics and Statistics*, 37: 336–346.

Hicks, J.R. (1946) *Value and Capital*, 2nd edn. Oxford University Press, London.

Hoken, J., Kojina, M., and Nakagawa, N. (1981) *Experimental Study of Continuous Models Applied for Migration.* Paper presented at Tsukuba Urbanization Conference, University of Tsukuba, Japan.

Holroyd, E.M. (1966) *Theoretical Average Journey Lengths in Circular Towns with Various Routing Systems.* Road Research Laboratory, Report No. 43, Crowthorne, Berks., England.

Hoover, E.M. (1948) *The Location of Economic Activity*. McGraw-Hill, New York.

Iri, M. (1969) *Network Flow, Transportation and Scheduling.* Academic Press, New York.

Isard, W. (1956) *Location and Space-Economy*. MIT Press, Cambridge, Massachusetts.

Isard, W. and Liossatos, P. (1979) *Spatial Dynamics and Optimal Space-Time Development.* North-Holland, Amsterdam.

Jutila, S.T. (1971) A Linear Model for Agglomeration, Diffusion, and Growth of Regional Economic Activity. *Regional Science Perspectives*, 1: 83–108.

Jutila, S.T. (1972) Dynamics of Regional Economic Development. *Regional Science Perspectives*, 2: 95–110.

Jutila, S. (1973) Generation of Regional Business Cycles through Interregional Feedback Mechanism. *Regional Science Perspectives*, 3: 52ff.

Kansky, K.J. (1963) *Structure of Transportation Networks.* Research Report No. 84, Department of Geography, University of Chicago, Chicago, Illinois.

Kantorovich, L. (1942) On the Translocation of Masses. *Doklady Akademii Nauk SSSR,* 37: translated into English in *Management Science,* 5: 1–4.

Koopmans, T.C. (1949) Optimum Utilization of the Transportation Systems. *Econometrica*, 17: 136–146.

Koopmans, T.C. and Reiter, S. (1951) *A Model of Transportation*. Cowles Commission Monograph No. 13, Chap. XIV, pp. 222–259. Wiley, New York.

Lam, T.N. and Newell, G.F. (1967) Flow Dependent Traffic Assignment on a Circular City. *Transportation Science*, 1: 318–361.

Lefeber, L. (1958) *Allocation in Space. Production, Transport and Industrial Location*. North-Holland, Amsterdam.

Leonardi, G. (1978) Optimal Facility Location by Accessibility Maximization. *Environment and Planning A*, 10: 1287–1305.

Lloyd, P.E. and Dicken, P. (1972) *Location in Space – A Theoretical Approach to Economic Geography*. Harper and Row, New York.

Melut, P. and O'Sullivan, P. (1974) A Comparison of Simple Lattice Transport Networks for a Uniform Plain. *Geographical Analysis*, 6: 163–173.

Mills, E.S. (1972) *Studies in the Structure of the Urban Economy.* Resources for the Future, Baltimore, Maryland.

Mills, E.S. and Lav, M.R. (1964) A Model of Market Areas with Free Entry. *Journal of Political Economy*, 72: 278–288.

Moses, L. (1960) A General Equilibrium Model of Production, Interregional Trade, and Location of Industry. *Review of Economics and Statistics*, 42: 373–397.

Mosler, K.C. (1976) *Optimale Transportnetze – Zur Bestimmung ihres Kostengünstigsten Standorts bei gegebener Nachfrage*. Lecture Notes in Economics and Mathematical Systems, Vol. 125. Springer-Verlag, Berlin.

Muth, R. (1969) *Cities and Housing.* University of Chicago Press, Chicago, Illinois.

Newell, G.F. (1980) *Traffic Flow on Transportation Networks.* MIT Press, Cambridge, Massachusetts.

Philips, L. and Thisse, J.F. (1982) Spatial Competition and Theory of Differentiated Markets: An Introduction. *Journal of Industrial Economics*, 31: 1–9.

Ponsard, C. (1955) *Economie et Espace.* Sedes, Paris.

Ponsard, C. (1958) *Histoire des Théories Economiques Spatiales.* Colin, Rennes.

Potts, R.B. and Oliver, R.M. (1972) *Flows in Transportation Networks.* Academic Press, New York.

Samuelson, P.A. (1947) *Foundations of Economic Analysis.* Harvard University Press, Cambridge, Massachusetts.

Samuelson, P.A. (1952) Spatial Price Equilibrium and Linear Programming. *American Economic Review*, 42: 283–303.

Schweizer, U. and Varaiya, P. (1976) The Spatial Structure of Production with a Leontief Technology. *Regional Science and Urban Economics*, 6: 231–251.

Sheppard, E.S. (1979a) Geographic Potentials. *Annals of the Association of American Geographers*, 69: 438–447.

Sheppard, E.S. (1979b) Interaction and Potential in Spatial Systems. *Ontario Geography*, 13: 47–60.

Singer, H.W. (1937) A Note on Spatial Price Discrimination. *Review of Economic Studies*, 5: 75–77.

Skellam, J.G. (1951) Random Dispersal in Theoretical Populations. *Biometrika*, 38: 196–218.

Solow, R. (1972) Congestion, Density and the Use of Land in Transportation. *Swedish Journal of Economics*, 74: 161–173.

Solow, R. and Vickrey, W. (1971) Land Use in a Long Narrow City. *Journal of Economic Theory*, 4: 430–447.

Stewart, C. (1958) The Size and Spacing of Cities. *Geographical Review*, 48: 222–245.

Stewart, J.Q. (1947) Empirical Mathematical Rules Governing the Distribution and Equilibrium of Population. *Geographical Review,* 37: 461–485.

Taguchi, A. and Iri, M. (1982) Continuum Approximation to Transport Networks and its Application to the Analysis of Urban Road Networks. *Mathematical Programming Study*, 20: 178–217.

Tobler, W.R. (1961) *Map Transformations of Geographic Space*. Ph.D. Dissertation, University of Washington, Seattle, Washington.

Tobler, W.R. (1973) *Visual Evidence for Urban Potential Fields*. Mimeo. University of Michigan, Ann Arbor, Michigan.

Tobler, W.R. (1975) *Spatial Interaction Patterns.* Research Report RR-75-19, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Tobler, W.R. (1978a) Comparison of Plane Forms. *Geographical Analysis*, 10: 155–162.

Tobler, W.R. (1978b) *A Geographical Migration Probability Density Function.* Mimeo. University of California, Santa Barbara, California.

Tobler, W.R. (1979a) Lattice Tuning. *Geographical Analysis*, 11: 36–44.

Tobler, W.R. (1979b) Cellular Geography. In S. Gale and G. Olsson (Eds.), *Philosophy in Geography*, pp. 379–386. Reidel, Dordrecht.

Vernon, R. (1972) *Structure and Growth of a von Thünen Economy.* Ph.D. Dissertation, Brown University, Providence, Rhode Island.

Wardrop, J.G. (1952) Some Theoretical Aspects of Road Traffic Research. *Proceedings of the Institute of Civil Engineers*, 1: 325–378.

Wardrop, J.G. (1966) Journey Speed and Flow in Central Urban Areas. *Traffic Engineering and Control*, 9: 528–532.

Wardrop, J.G. (1969) Minimum Cost Paths in Urban Areas. *Strassenbau- und Strassenverkehrstechnik*, 86: 184–190.

Wardrop, J.G. (1971) *Minimum Cost Paths when Cost per Unit Length Depends on Location and Direction.* Paper presented at the 5th Int. Symposium on the Theory of Traffic Flow and Transportation, University of California, Berkeley, California.

Warntz, W. (1959) *Toward a Geography of Price: A Study in Geo-econometrics*. University of Pennsylvania Press, Philadelphia, Pennsylvania.

Warntz, W. (1965) *A Note on Surfaces and Paths and Applications to Geographic Problems.* Discussion Paper 6, Michigan Inter-University Community of Mathematical Geography, Ann Arbor, Michigan.

Warntz, W. (1967) Global Science and the Tyranny of Space. *Papers and Proceedings of the Regional Science Association*, 19: 7–19.

Werner, C. (1966) *Zur Geometrie von Verkehrsnetzen.* Abhandlungen des 1. Geografischen Instituts der FU, Berlin.

Werner, C. (1968) Research Seminar in Theoretical Transportation Geography. In F. Horton (Ed.), *Geographic Studies of Urban Transportation and Network Analysis.* Northwestern University, Evanston, Illinois.

Williams, H.C.W.L. (1974a) *Some Generalizations and Applications of the Velocity Field Concept: 2. Problems Relating to Traffic Restraint, Optimal Investment and Velocity Field Dynamics.* WP 31, Institute of Transport Studies, University of Leeds, Leeds, England.

Williams, H.C.W.L. (1974b) *Some Generalizations and Applications of the Velocity Field Concept: 3. The Location and Pricing of Park-and-ride in Urban Areas.* WP 32, Institute of Transport Studies, University of Leeds, Leeds, England.

Williams, H.C.W.L. and Orthuzar Salas, J.D. (1976) Some Generalizations and Applications of the Velocity Field Concept. *Transportation Research,* 10: 65–73.

Wilson, A.G. (1970) *Entropy in Urban and Regional Modelling.* Pion, London.

Wingo, L. (1961) *Transportation and Urban Land.* Resources for the Future, Washington, D.C.

Zipf, G.K. (1941) *National Unity and Disunity.* Principal Press, Bloomington, Indiana.

Zitron, N.R. (1974) A Continuous Model of Optimal-Cost Routes in a Circular City. *Journal of Optimization Theory and Applications*, 14: 291–303.

Zitron, N.R. (1978) A Critical Condition for the Cost Density in the Circular City Model. *Journal of Optimization Theory and Applications*, 24: 507–512.

C. RELATED PUBLICATIONS BY THE AUTHORS

Beckmann, M.J. (1952) A Continuous Model of Transportation. *Econometrica*, 20: 643–660.

Beckmann, M.J. (1953) The Partial Equilibrium of a Continuous Space Market. *Weltwirtschaftliches Archiv*, 71: 73–89.

Beckmann, M.J. (1957) On the Equilibrium Distribution of Population in Space. *Bulletin of Mathematical Biophysics*, 19: 81–90.

Beckmann, M.J. (1968) *Location Theory*. Random House, New York.

Beckmann, M.J. (1969) Market Shares and Distance. *The Swedish Journal of Economics*, Special Issue in Honor of Tord Palander: 53–63.

Beckmann, M.J. (1970) The Analysis of Spatial Diffusion Processes. *Papers of the Regional Science Association*, 25: 109–117.

Beckmann, M.J. (1971) Equilibrium versus Optimum: Spacing of Firms and Patterns of Market Areas. *Northeast Regional Science Review*, 1: 1–20.

Beckmann, M.J. (1972) Spatial Cournot Oligopoly. *Papers of the Regional Science Association*, 28: 37–47.

Beckmann, M.J. (1976a) Equilibrium and Stability in a Continuous Space Market. *Operations Research Verfahren*, 14: 48–63.

Beckmann, M.J. (1976b) Spatial Price Policies Revisited. *The Bell Journal of Economics*, 7(2): 619–630.

Beckmann, M.J. (1980) Continuous Models of Transportation and Location Revisited. *Papers of the Regional Science Association*, 45: 45–53.

Beckmann, M.J. (1981) Continuous Models of Transportation and Location. *Sistemi Urbani*, 3: 403–413.

Beckmann, M.J. (1982a) *Perspectives of Location Theory*. Public lecture delivered on assuming the Tinbergen Chair at the Erasmus University, Rotterdam, May 27.

Beckmann, M.J. (1982b) Locational Aspects of Resource Depletion. In W. Eichhorn and R. Henn (Eds.), *Economic Theory of Natural Resources*. Birkhauser, Boston, Massachusetts, pp. 89–95.

Puu, T. (1977) A Proposed Definition of Traffic Flow in Continuous Transportation Models. *Environment and Planning*, 9: 559–567.

Puu, T. (1978a) Towards a Theory of Optimal Roads. *Regional Science and Urban Economics*, 8: 225–248.

Puu, T. (1978b) On the Existence of Optimal Paths and Cost Surfaces in Isotropic Continuous Transportation Models. *Environment and Planning*, 10: 1121–1130.

Puu, T. (1978c) On Traffic Equilibrium, Congestion Tolls and the Allocation of Transportation Capacity in a Congested Urban Area. *Environment and Planning*, 10: 29–36.

Puu, T. (1979a) Regional Modelling and Structural Stability. *Environment and Planning A*, 11: 1431–1438.

Puu, T. (1979b) *The Allocation of Road Capital in Two-Dimensional Space*. North-Holland, Amsterdam.

Puu, T. (1981a) Stability and Change in Two-Dimensional Flows. In D.A. Griffith and R.D. MacKinnon (Eds.), *Dynamic Spatial Models*. Sijthoff and Noordhoof, Alphen, The Netherlands.

Puu, T. (1981b) Structural Stability and Change in Geographical Space. *Environment and Planning A*, 13: 979–989.

Puu, T. (1981c) Catastrophic Structural Change in a Continuous Regional Model. *Regional Science and Urban Economics*, 11: 317–333.

Puu, T. (1982a) Outline of a Trade Cycle Model in Continuous Space and Time. *Geographical Analysis*, 14: 1–9.

Puu, T. (1982b) The General Equilibrium of a Spatially Extended Market Economy. *Geographical Analysis*, 14: 145–154.

Puu, T. (1982c) *Structurally Stable Transport Flows and Patterns of Location.* Research Report RR-82-42, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Puu, T. (1982d) Continuous Flow Modelling in Regional Science. In M. Albegov, Å.E. Andersson, and F. Snickars (Eds.), *Regional Development Modelling: Theory and Practice.* North-Holland, Amsterdam.

Puu, T. (1984) A Model of Spatial Flows and Growth of Capital and Labor Stocks. *Applied Mathematics and Computation*, 14: 3–9.

D. MATHEMATICAL REFERENCE WORKS

Allen, R.G.D. (1956) *Mathematical Economics.* Macmillan, New York.

Andronov, A. and Pontryagin, L. (1937) Systèmes grossiers. *Comptes Rendus de l'Académie des Sciences de l'URSS,* 14: 247–250.

Arnold, V.I. (1978) *Mathematical Methods of Classical Mechanics.* Springer-Verlag, New York.

Arnold, V.I. (1983) *Geometrical Methods in the Theory of Ordinary Differential Equations.* Springer-Verlag, New York.

Arrow, K.J. and Hahn, F.H. (1971) *General Competitive Analysis*, Holden-Day, San Francisco, California.

Beck, A. (1974) *Continuous Flows in the Plane.* Springer-Verlag, Berlin.

Bolza, O. (1904) *Lectures on the Calculus of Variations*. The University of Chicago Press, Chicago, Illinois.

Buck, R.C. (1965) *Advanced Calculus.* McGraw-Hill, New York.

Cartan, H. (1963) *Elementary Theory of Analytic Functions of One or Several Complex Variables.* Addison-Wesley, Reading, Massachusetts.

Coral, M. (1931/32) The Euler–Lagrange Multiplier Rule for Double Integrals. *Contributions to the Calculus of Variations*, 1. Johnson Reprint Corporation, Subsidiary of H.B. Johnson, New York.

Courant, R. and Hilbert, D. (1953) *Methods of Mathematical Physics.* Interscience, New York.

Courant, R. and John, F. (1965) *Introduction to Calculus and Analysis*, Vol. 1. Interscience, New York.

Courant, R. and John, F. (1974) *Introduction to Calculus and Analysis*, Vol. 2. Interscience, New York.

Coxeter, H.S.M. (1961) *Introduction to Geometry*. Wiley, New York.

DoCarmo, M.D. (1976) *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Englewood Cliffs, New Jersey.

Dodson, C.T.J. and Poston, T. (1977) *Tensor Geometry.* Pitman, London.

Duff, G.F.D. and Naylor, D. (1966) *Differential Equations of Applied Mathematics*. Wiley, New York.

Elsgolc, L.E. (1961) *Calculus of Variations.* Pergamon, Oxford.

Fejes Toth, L. (1969) *Regular Figures.* Pergamon, Oxford.

Forsyth, A.R. (1927) *Calculus of Variations.* Cambridge University Press, Cambridge.

Fox, C. (1950) *Introduction to the Calculus of Variations.* Oxford University Press, London.

Gilmore, R. (1981) *Catastrophe Theory for Scientists and Engineers.* Wiley, New York.

Goldstein, H. (1950) *Classical Mechanics.* Addison-Wesley, Reading, Massachusetts.

Guckenheimer, J. (1973) One-Parameter Families of Vector Fields on Two-Manifolds: Another Nondensity Theorem. In: M.M. Peixoto (Ed.), *Dynamical Systems*, pp. 111–127. Academic Press, New York.

Guggenheimer, H.W.(1963) *Differential Geometry.* McGraw-Hill, New York.

Haken, H. (1978) *Synergetics.* Springer-Verlag, Berlin.

Henle, M. (1979) *A Combinatorial Introduction to Topology.* Freeman, San Francisco, California.

Hilbert, D. and Cohn-Vossen, S. (1932) *Anschauliche Geometrie.* Springer-Verlag, Berlin.

Hirsch, M.W. and Smale, S. (1974) *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, New York.

Kreyszig, E. (1979) *Advanced Engineering Mathematics.* Wiley, New York.

Lefschetz, S. (1977) *Differential Equations: Geometric Theory.* Dover, New York.

Mandelbrot, B.B. (1977) *Fractals – Form, Chance, and Dimension.* Freeman, San Francisco, California.

Marsden, J.E. (1973) *Basic Complex Analysis.* Freeman, San Francisco, California.

Marsden, J.E. and Tromba, A.J. (1976) *Vector Calculus.* Freeman, San Francisco, California.

Morse, M. (1934) The Calculus of Variations in the Large. *American Mathematical Society Colloquia*, Publication 18.

Nemytskii, V.V. and Stepanov, V.V. (1960) *Qualitative Theory of Differential Equations.* Princeton University Press, Princeton, New Jersey.

Nicolis, G. and Prigogine, I. (1977) *Self-Organization in Nonequilibrium Systems.* Wiley, New York.

Palis, J. (1969) On Morse–Smale Dynamical Systems. *Topology*, 8: 385–405.

Peixoto, M.M. (1962) Structural Stability on Two-Dimensional Manifolds. *Topology*, 1: 101–120.

Peixoto, M.M. (1967) Qualitative Theory of Differential Equations and Structural Stability. In J.K. Hale and J.P. LaSalle (Eds.), *Proceedings of a Symposium on Differential Equations and Dynamical Systems, Mayaguez, Puerto Rico, 1965.* Academic Press, New York.

Peixoto, M.M. (1973) On the Classification of Flows on 2-Manifolds. In M.M. Peixoto (Ed.), *Dynamical Systems.* Academic Press, New York.

Peixoto, M.M. (1977) Generic Properties of Ordinary Differential Equations. *Studies in Ordinary Differential Equations*, 14: 52–92.

Poston, T. and Stewart, I. (1978) *Catastrophe Theory and its Applications.* Pitman, London.

Rutherford, O.E. (1948) *Vector Methods.* Oliver and Boyd, Edinburgh.

Santalo, L.A. (1976) Integral Geometry and Geometric Probability. *Encyclopedia of Mathematics and its Applications.* Addison-Wesley, Reading, Massachusetts.

Simmons, G.F. (1972) *Differential Equations.* McGraw-Hill, New York.

Smale, S. (1967) Differentiable Dynamical Systems. *Bulletin of the American Mathematical Society*, 73: 747–817.

Sneddon, I. (1957) *Elements of Partial Differential Equations.* McGraw-Hill, New York.

Sotomayor, J. (1973) Generic Bifurcation of Dynamical Systems. In M.M. Pexioto (Ed.), *Dynamical Systems*, pp. 561–582. Academic Press, New York.

Struik, D.J. (1950) *Differential Geometry.* Addison-Wesley, Reading, Massachusetts.

Synge, J.L. and Schild, A. (1949) *Tensor Calculus.* University of Toronto Press, Toronto, Ontario.

Thom, R. (1969) Topological Models in Biology. *Topology*, 8: 313–335.

Thom, R. (1975) *Structural Stability and Morphogenesis.* Benjamin, Reading, Massachusetts.

Weatherburn, C.E. (1943a) *Elementary Vector Analysis*. Bell and Sons, London.

Weatherburn, C.E. (1943b) *Advanced Vector Analysis*. Bell and Sons, London.

Weinstock, R. (1974) *Calculus of Variations.* Dover, New York.

Whittle, P. (1954) On Stationary Processes in the Plane. *Biometrika*, 41: 434–449.

Name Index

Subject Index

ISBN : 0 444 87771 1