



Det här verket har digitaliserats vid Göteborgs universitetsbibliotek och är fritt att använda. Alla tryckta texter är OCR-tolkade till maskinläsbar text. Det betyder att du kan söka och kopiera texten från dokumentet. Vissa äldre dokument med dåligt tryck kan vara svåra att OCR-tolka korrekt vilket medför att den OCR-tolkade texten kan innehålla fel och därför bör man visuellt jämföra med verkets bilder för att avgöra vad som är riktigt.

This work has been digitized at Gothenburg University Library and is free to use. All printed texts have been OCR-processed and converted to machine readable text. This means that you can search and copy text from the document. Some early printed books are hard to OCR-process correctly and the text may contain errors, so one should always visually compare it with the images to determine what is correct.



**Rapport**

**R15:1972**

TEKNISKA HOGSKOLAN I LUND  
SEKTIONEN FOR VAG. OCH VASTEN  
BIBLIOTEKET

**Dynamiska  
lokaliseringsmodeller**

**Anders Karlqvist et al.**

**Byggforskningen**



# Dynamiska lokaliseringsmodeller

Sammanställning av föredrag vid forskarseminarium  
15–16 december 1970

Den rapport som sammanfattas nedan utgör en sammanställning av ett antal föredrag som presenterades vid ett forskarseminarium i Stockholm 1970. Syftet med seminariet var att diskutera olika problem kring tids- och rumsdimensionen i samhällsplaneringsmodeller med tyngdpunkten på matematiska lokaliseringsmodeller och därmed förknippade metodfrågor. De åtta uppsatser som ingår i rapporten kan i princip ses som helt fristående från varandra. Några av dem är av översiktsskärakt, andra är i huvudsak referat av tidigare forskningsarbeten medan det i övriga fall är fråga om diskussioner kring inte fullt genomförda forskningsidéer.

Bland de ämnen som tas upp till behandling märks analys av samhällsplaneringsmodeller från systemteoretiska utgångspunkter, prediktionsproblem vid rumsliga beskrivningsmodeller, optimeringsmodeller för bebyggelselokalisering över tiden och industriell kapacitetsutbyggnad på nationell nivå, långsiktig inomregional lokaliseringsanalys samt markanvändningskonkurrens mellan privata och kollektiva intressen. De olika uppsatserna förutsätter i varierande utsträckning förkunskaper inom matematik och ekonomi. För att i någon mån underlätta förståelsen av vissa fackuttryck har ett antal ordförklaringar införts i en bilaga.

Under de senaste decennierna har analyser av lokaliseringsproblem med hjälp av formella lokaliseringsmodeller ägnats allt större uppmärksamhet. Detta kan ses dels som resultat av ett växande intresse för långsiktig fysisk planering i allmänhet, dels som en följd av utvecklingen av den moderna datatekniken, vilken gjort det möjligt att omsätta teoretiska modellansatser i praktiskt planeringsarbete.

Ett utmärkande drag hos de flesta existerande modeller inom lokaliseringsteorin är deras statiska karaktär. Medan rumsdimensionen ges en relativt detaljerad beskrivning inskränks behandlingen av tiden till att omfatta en enda tidsperiod. Inom den ekonomiska teorin är förhållandet snarast det motsatta: expansionsförlopp i tiden ägnas stor uppmärksamhet, medan de rumsliga förändringarna behandlas tämligen summariskt. En utveckling av lokaliseringsmodeller, som samtidigt behandlar förlopp i rummet och tiden, måste därför anses vara av stor betydelse för mera

realistiska modellstudier av problem inom långsiktig fysisk planering.

De uppsatser som ingår i denna rapport utgör en relativt heterogen samling, inom ramen för ämnesområdet dynamiska lokaliseringsmodeller. I den följande sammanfattningen har målet varit att ge en separat presentation av varje uppsats för att därigenom ytterligare framhäva att rapporten består av ett antal fristående delar.

## Inledning om dynamiska system Anders Karlqvist

Lokaliseringsmodeller systematiseras traditionellt efter matematisk komplexitet, varvid de dynamiska modellerna vanligen hänförs till de mest komplexa.

En jämförande analys av de egenskaper som karakteriserar dynamiska system inom fysiken, reglertekniken och samhällsplaneringen görs i termer av systemteoretiska begrepp. Härvid ges först en generell systemteoretisk definition av dynamiska system, varefter en precis innebörd i dessa termer ges sådana dynamiska begrepp som invariants, tidsfördröjning, kontroll och feedback.

På praktiskt taget alla punkter visar sig samhällsplaneringsmodellerna ställa större krav på komplexitet än de övriga; tidsinvariants råder inte, stora tidsfördröjningar finns inbyggda i systemen, möjligheterna till kontroll och styrning spelar en central roll etc. Dessutom är systemen inom samhällsplaneringen ofta av adaptiv natur, vilket i allmänhet gör dem svåra att behandla med matematiska metoder.

## Beskrivande modeller och prediktion — en översikt Folke Snickars

Det är vanligt att vid modellarbete göra en uppdelning i beskrivande, förklarande och planerande modeller. I anslutning till en översikt över några beskrivande modeller ges exempel på prediktionsproblem som uppstår när data utgöres av rums- och tidsserier. Några exempel ges på andra matematiska tekniker än de klassiska regressionsanalytiska, som kan användas i beskrivande och förklarande modeller. Speciellt diskuteras modeller som utnyttjar Markovteknik och exempel ges på tillämpningar i form av studier av befolkningstäthetsvariation i ett stadsområde. Det blir alltmera vanligt att utnyttja simulering för analys av urbana lokaliseringsförlopp. Några

# Bygghforskningen Sammanfattningar

## R15:1972

Nyckelord:

dynamiska lokaliseringsmodeller, beskrivande modeller, optimeringsmodeller, transportmodeller

systemteori, nyttoteori, kontrollteori

matematisk programmering, klustringsmetoder

markanvändningskonkurrens, kapacitetsutbyggnad

Rapport R15:1972 avser anslag Bs 471 från Statens råd för byggnadsforskning till tekn.lic. Anders Karlqvist, institutet för matematik, KTH, Stockholm.

UDK 711.1.001.57  
333.013.4

SfB A  
ISBN 91-540-2020-4

Sammanfattning av:

Karlqvist, A, Snickars, F, Marksjö, B, Lundqvist, L, Holvid, A, Andersson, Å, E & Lönnroth, J, 1972, *Dynamiska lokaliseringsmodeller*. (Statens institut för byggnadsforskning) Stockholm. Rapport R15:1972, 174 s., ill. 35 kr.

Rapporten är skriven på svenska med svensk och engelsk sammanfattning.

Distribution:

Svensk Byggtjänst  
Box 1403, 111 84 Stockholm  
Telefon 08-24 28 60  
Grupp: samhällsplanering

exempel på användning av denna teknik diskuteras och som avslutning analyseras några grunddrag i en större urban simuleringsmodell.

## **Teori för och bestämning av nyttofunktionaler**

**Bertil Marksjö**

Begreppet nyttofunktional definieras och några alternativa egenskaper hos nyttor diskuteras. En skala definieras som en ekvivalensrelation mellan nyttofunktionaler och exempel på olika skalor ges.

Vid numerisk bestämning av en nyttofunktional med ändligt många argument utnyttjas lotterier i enlighet med von Neumann-Morgensterns teoretiska arbeten. Begreppen riskaversion och nyttooberoende diskuteras med hjälp av liknande lotteriförfaranden. Avslutningsvis behandlas problem som rör bestämning av nyttofunktionaler vid tidsutvecklingar och det visas, att en beslutsfattare, som inte ångrar tidigare fattade beslut, bör arbeta med en konstant diskontineringsfaktor.

## **Optimerings- och styrproblem**

**Lars Lundqvist**

Problemet att lokalisera över tiden växande mängder av verksamheter till givna ytor betraktas som ett kontrollproblem och formuleras generellt i en optimeringsmodell. Genom att införa viktiga begränsningar kan denna modell ges strukturen av ett deterministiskt flerstegigt beslutsproblem. Olika lösningsmetoder i matematisk programmering diskuteras, speciellt med avseende på metodmässiga och beräkningsmässiga restriktioner. Som illustration visas enkla resultat för problemet att över tiden lokalisera verksamheter till områden längs en rät linje.

## **En planeringsmodell för kapacitetsexpansion**

**Anders Karlqvist & Lars Lundqvist**

Problemet att bestämma optimal utbyggnadsstrategi för en enproduktindustri har behandlats i en avhandling av D. Erlenkotter, Stanford, USA. Givet är ett antal platser med känd, över tiden växande, efterfrågan av en viss produkt. Uppgiften är att söka tidpunkt, storlek och plats för de kapacitetsexpansioner som svarar mot minsta kostnad, sammansatt av diskonterade investerings-, produktions- och transportkostnader.

Problemet kan formuleras och lösas med hjälp av dynamisk programmering. Vanligen sker detta genom att diskret tid införes. Erlenkotter visar dock att man kan reducera beräkningsmängden väsentligt genom att behålla kontinuerlig tid och istället använda kapacitetsnivåer som stegvariabel. Först behandlas det icke-stationära planeringsproblemet med ändlig tidshorisont. Lösningsmetoder utvecklas sedan även för motsvarande stationära problem i oändlig tid. Tillståndsrummets expansion gör att den med normalstora datorer lösbara problemstorleken begränsas till tre platser. Det stationära problemet illustreras med några exempel: potentiell vinst av regional samordning av investeringsbeslut, utseendet hos optimal utbyggnadsstrategi samt möjlig förlust vid icke-optimala initiala beslut.

## **Heltalsprogrammeringsmodeller för dynamisk lokalisering**

**Anders Holvid**

I nuläget är en viss fördelning av produktionsenheter given. Produktionsfördelningen är avsedd att möta den rådande konsumtionsstrukturen. Det problem som behandlas är hur man skall förändra produktionsstrukturen för att få bästa resultat i förhållande till prognostiserad framtida konsumtion. Modellförslaget innebär en utvidgning av det så kallade simple plant location-problemet till flera tidsperioder genom att formulera ett antal möjliga strategier för varje produktionsenhet och utvärdera dessa. Problemet blir ett blandat heltalsproblem som dekomponeras i två delar, en strategivald och en transportoptimeringsdel.

## **Lokaliseringsanalys för Stockholmsregionen**

**Åke E. Andersson**

Som underbyggnad till ett förslag till dynamisk lokaliseringsanalys för Stockholmsregionen behandlas ett antal problem rörande regional dynamisk lokalisering av ekonomiska aktiviteter. Med utgångspunkt från en jämviktsmodell av Lefebvre diskuteras lönsamheten av ett vidgat pendlingsarbete av arbetskraften i en region. Härvid arbetas bl a med produktionsfunktioner som är strikt konkava. Därefter diskuteras realismen i detta antagande med hänsyn till förekomsten av stordriftsfördelar och odelbarheter. En transportoptimeringsmodell av hel-

talstyp visas möjliggöra en lokaliseringsanalys av odelbara enheter efter lämpligt val av objektfunktion. För att komma till rätta med ömsesidiga beroenden i det ekonomiska systemet presenteras en klustringsanalys, som i princip avser att identifiera de ekonomiska aktiviteter som med fördel kan samlokaliseras. En väg som föreslås för detta ändamål är användning av faktoranalytiska metoder. När klustringsanalysen väl genomförd kan problemet att placera klustren optimalt i stadsregionen studeras. För genomförande av en dynamisk lokaliseringsanalys av klustren föreslås en till rekursiv programmering omarbetad transportoptimeringsmodell av den typ som tidigare behandlats, utvidgad till att endogent behandla även bostads- och transportpolitiken. Avslutningsvis diskuteras några sammanhörande problem rörande ekonomisk tillväxttakt, ränta och optimal planhorisont.

## **Om kriterier för markanvändningsbeslut**

**Johan Lönnroth**

För att belysa marknadsmekanismens roll vid markanvändningsbeslut diskuteras en modell med markkonkurrens mellan två verksamheter, den ena varuproducerande med internationell marknad (t ex oljeraffinering), den andra icke-produktiv med svårsmärkt avkastning (t ex fritidsboende). Den första verksamheten antas beredd att betala en lägesränta av upp till skillnaden mellan marknadspriset på oljeprodukter och deras produktionskostnader för att få ta en eftertraktad lokalisering i besittning. Motsvarande lägesräntor för fritidsboendet visas kunna uppfattas som de ränteintäkter från konkurrerande markanvändare som samhället är berett att avstå ifrån för att bereda plats för en politiskt beslutad fritidsbebyggelse. Samtidigt kan de belopp härledas, med vilka samhället i egenskap av fritidsboende-producent är berett att subventionera fritidsägare bosatta i olika områden.

Slutsatsen är att det endast i speciella fall går att hitta ett system av lägesräntor och subventioner, som leder till en ur samhällets synpunkt optimal markanvändning om verksamheterna är varandra ömsesidigt uteslutande.

Avslutningsvis presenteras ett forskningsprogram för studier av markanvändningsplanering, där en utvidgning av den tidigare presenterade analysen skulle kunna ingå.

# Dynamic location models

Collection of lectures given at a research seminar held on  
15–16 December 1970

The report summarized below is a compendium of a number of lectures held at a research seminar in Stockholm in 1970. The object of this seminar was to discuss different problems concerning the time and space dimension in urban planning models, with the emphasis on mathematical location models and methodological problems associated with these. In principle, the eight lectures covered by the report may be considered to be independent of one another. Some of them are reviewing in character, others are mainly accounts of previous research work, while the remainder are discussions concerning research projects which have not yet been completely realised.

Among the subjects dealt with may be mentioned an analysis of urban planning models on the basis of systems theory, prediction problems associated with spatial descriptive models, optimization models for the location of intra-regional building development over time and for industrial capacity expansion on the national level, long-term intra-regional location analysis and land use competition between private and public interests.

To a varying extent, the different lectures require prior knowledge of mathematics and economics. In order to facilitate the understanding of some of the terms, a limited vocabulary has been included in an appendix.

The analysis of location problems with the aid of formal location models has received increasing attention over the past few decades. This may be seen as a result of a growing interest in long-term physical planning in general, and also as a consequence of the development of modern computer techniques which has made possible the use of theoretical models in practical planning.

A distinctive characteristic of most existing models in the field of location theory is their static character. While the spatial dimension receives a comparatively detailed description, the treatment of time is confined to a single period. In the field of economic theory, the situation is exactly the opposite: expansion processes over time are given considerable attention while spatial changes are dealt with in a fairly summary manner. A development of location models which simultaneously treat processes in time and space must therefore be considered of great significance for more realistic model studies of problems in long-term physical planning.

The lectures covered in this report are a relatively heterogeneous collection within the framework of dynamic location

models. The intention in the summary below has been to present each lecture on its own, in order to emphasise yet again that the report consists of a number of independent parts.

## An introduction to dynamic systems Anders Karlqvist

Location models are traditionally systematized on the basis of their mathematical complexity, the dynamic models being usually classified among the most complex ones.

A comparative analysis of the properties which characterise dynamic systems in physics, automatic control and urban planning is performed in terms of systems theory concepts. There is first of all a general systems-theoretical definition of dynamic systems, after which a precise meaning of dynamic concepts such as invariance, time-lag, control and feedback is stated in these terms.

It is found that, in practically all respects, urban planning models demand a more complex treatment than other models; there is no time invariance, there are large time-lags inherent in the systems, the possibilities of control play a central part, etc. In addition to this, systems in urban planning are often adaptive which, as a rule, makes them intractable by mathematical methods.

## Descriptive models and prediction — a survey Folke Snickars

It is common in working with models to divide these into descriptive, explanatory and planning models. In conjunction with a survey of some descriptive models, examples are given of prediction problems which arise when the data consist of space and time series. Some examples are given of mathematical techniques other than the classical regression analysis methods which can be used in descriptive and explanatory models. There is a special discussion of models which utilise Markovian techniques, and examples are given of applications in the form of studies of the variation in population density in an urban area. The use of simulation for the analysis of urban location processes is becoming increasingly common. Some examples of the use of this technique are discussed and, in conclusion, some basic characteristics of a major urban simulation model are analysed.

## Theory for and evaluation of utility functionals Bertil Marksjö

The concept utility functional is defined

# National Swedish Building Research Summaries

## R15:1972

Key words:

*dynamic location models*, descriptive models, optimization models, transportation models

*systems theory*, utility theory, control theory

*mathematical programming*, cluster methods

*land use competition*, capacity expansion

Report R15:1972 has been supported by Grant Bs 471 from the Swedish Council for Building Research to Anders Karlqvist, Ph. D. (tekn.lic.), Division of mathematics, Royal Institute of Technology, Stockholm, Sweden.

UDC 711.1.001.57  
333.013.4  
SfB A  
ISBN 91-540-2020-4

Summary of:

Karlqvist, A, Snickars, F, Marksjö, B, Lundqvist, L, Holvid, A, Andersson, Å, E & Lönnroth, J, 1972, *Dynamiska lokaliseringsmodeller*. Dynamic location models. (Statens institut för byggnadsforskning) Stockholm. Report R15:1972, 174 p., ill. 35 Sw.Kr.

The report is in Swedish with Swedish and English summaries.

Distribution:

Svensk Byggtjänst  
Box 1403, S-111 84 Stockholm  
Sweden



and some alternative properties of utilities are discussed. A scale is defined as an equivalence relation between utility functionals, and examples of different scales are given. In numerical evaluation of utility functionals with a finite number of arguments, lotteries according to the theoretical work performed by von Neumann and Morgenstern are utilised. The concepts of risk aversion and utility independency are discussed with the aid of similar lottery procedures. In conclusion, problems concerning the determination of utility functionals over time are treated, and it is shown that a decision maker who does not regret decisions made earlier should employ a constant discount rate.

### **Optimization and optimal control problems**

**Lars Lundqvist**

The problem of allocating increasing amounts of activities to given land areas over time is regarded as a control problem and is formulated as an optimization problem. By the introduction of essential land-use limitations, this model can be given the structure of a multistage deterministic decision problem. Different methods of solution in mathematical programming are discussed, with particular reference to methodological and computational restrictions. Some simple results of the problem of locating, over a period of time, activities in an area along a straight line are shown as an illustration.

### **A planning model for capacity expansion**

**Anders Karlqvist & Lars Lundqvist**

The problem of determining optimal expansion strategies for a single-product industry has been dealt with in a doctoral thesis by D. Erlenkotter, Stanford, USA. The problem is to find, for a given number of locations with a known demand for a certain product which increases with time, the time, size and location of the capacity expansions which minimize a total cost function, composed of discounted investment, production and transportation costs. This problem can be formulated and solved by means of dynamic programming. This is usually done by discretizing time, but Erlenkotter demonstrates that the com-

putational volume can be considerably reduced by the retention of continuous time and the use of the capacity levels as stage variables. The non-stationary planning problem with a finite time horizon is dealt with first, and methods of solution are then also developed for the corresponding stationary problem in infinite time. Owing to the expansion of the state space, the size of problems which can be solved by computers of normal size is limited to three places. The stationary problem is illustrated by some examples: the potential gain resulting from regional coordination of investment decisions, the characteristics of optimal expansion strategies and the losses resulting from non-optimal initial decisions.

### **Integer programming models for dynamic location**

**Anders Holvid**

The current distribution of production units is given. This production distribution is intended to satisfy the existing consumption structure. The problem examined here is how the production structure should be changed in order that some specified goals may be achieved in relation to forecasts of future consumption. The proposed model entails an expansion of the simple plant location problem to cover several periods of time, by the formulation of a number of possible strategies for each production unit and the evaluation of these. Technically, the problem becomes a mixed integer programming problem which is broken down into two parts, a strategy choice part and a transport optimisation part.

### **Location analysis for the Stockholm region**

**Åke Andersson**

A number of problems concerning regional dynamic location of economic activities is dealt with to form the basis for a proposed dynamic location analysis for the Stockholm region. The profitability of increased daily commuting of the work force in a region is discussed on the basis of an equilibrium model according to Lefebvre, the work being based, *inter alia*, on production functions which are strictly concave. The realism of this assumption is then discussed with regard to the presence of indivisibilities

and increasing returns to scale.

A transportation model of integer type is shown to make it possible to perform a location analysis of indivisible units, under a suitable choice of objective function. To cope with interdependencies in the economic system a cluster type analysis is presented, which is, in principle, intended to identify the economic activities that have an advantage of being located in the same clusters. Factor-analytical methods are proposed as a way towards solving these problems. Once the cluster analysis has been achieved, the problem of locating the clusters optimally in the urban area can be tackled. To perform a dynamic location analysis in this respect, a recursive programming model is proposed — it might be an extension of the transportation model already discussed, reformulated to treat the housing and transportation problems endogenously. Finally, some related problems concerning economic rate of growth, discount rate and optimal planning horizon are given some attention.

### **About criteria for land use decisions**

**Johan Lönnroth**

To illustrate the part played by the market mechanism in land use decisions a model is discussed which includes land use competition between two activities, one of them productive with an international market (e.g. oil refinement), the other non-productive with a scarcely measurable output (e.g. summer housing). It is shown that the first activity is prepared to pay a location rent of up to the difference between the market price on oil products and their production costs to acquire a desirable location.

The location rents that can most be paid by summer housing could be looked upon as those incomes from competing land users that society is willing to give up in order to clear the ground for a politically determined summer housing project, or the amount with which society is willing to subsidize summer housing in the area under study. Furthermore, it is asserted that it is not possible, in general, to find a consistent set of such location rents and subventions if the activities are mutually exclusive.

In the end the author outlines a research program for studies of land use planning, where an extension of the presented analysis could be included.

Rapport R15:1972

DYNAMISKA LOKALISERINGSMODELLER

Sammanställning av föredrag vid forskar-  
seminarium den 15-16 december 1970

DYNAMIC LOCATION MODELS

Collection of lectures given at a research  
seminar held on 15-16th December 1970

Anders Karlqvist, Folke Snickars, Bertil Marksjö,  
Lars Lundqvist, Anders Holvid, Åke E. Andersson  
& Johan Lönnroth

Denna rapport avser anslag Bs 471 från Statens råd för byggnads-  
forskning till tekn.lic. Anders Karlqvist, matematiska institu-  
tionen, KTH, Stockholm. Försäljningsintäkterna tillfaller fonden  
för byggnadsforskning.

Statens institut för byggnadsforskning, Stockholm

ISBN 91-540-2020-4

Rotobekman Stockholm 1973

## INNEHÅLL

## FÖRORD

## Anders Karlqvist

1	INLEDNING OM DYNAMISKA SYSTEM .....	7
1.1	Vad är dynamiska system .....	7
1.2	Tidsinvarians .....	10
1.3	Jämvikt .....	11
1.4	Tidsfördröjning .....	12
1.5	Kontinuerlig och diskret tid .....	13
1.6	Ändliga system .....	14
1.7	Kontrollteori .....	14
1.8	Litteraturreferenser .....	19

## Folke Snickars

2	BESKRIVANDE MODELLER OCH PREDIKTION - EN ÖVERSIKT .....	20
2.1	Vad är en beskrivande modell? .....	20
2.2	Exempel på beskrivande modeller .....	20
2.3	Trender och prognoser - några synpunkter ....	23
2.4	Regressionsanalys och parameterestimation ...	25
2.5	En statistisk modell med maximum-likelihood- estimering .....	27
2.6	Markovkedjor .....	34
2.7	En trafikmodell, som utnyttjar Markovteknik .	38
2.8	Simulering av lokaliseringsförlopp. En Monte- Carlo-modell för befolkningstäthet .....	40
2.9	Några anmärkningar om "Urban Dynamics" .....	41
2.10	Sammanfattning .....	43
2.11	Litteraturreferenser .....	45

## Bertil Marksjö

3	TEORI FÖR OCH BESTÄMNING AV NYTTOFUNKTIONALER	46
3.1	Grundläggande begrepp .....	46
3.2	Numerisk bestämning av nyttofunktionalen $u$ när antalet attribut är ändligt .....	55
3.3	Kvalitativ bestämning av nyttofunktionalen $u$ , när alternativen beskriver utvecklingar över tiden .....	63
3.4	Litteraturreferenser .....	67

## Lars Lundqvist

4	OPTIMERINGS- OCH STYRPROBLEM .....	68
4.1	Inledning .....	68
4.2	Begränsning av problemet .....	69
4.3	Begreppen styrprogram och styrlag .....	72
4.4	Linjära staden .....	74
4.5	Linjär programmering (LP) .....	75
4.6	Dynamisk programmering (DP) .....	78
4.7	Icke-linjär programmering (ILP) .....	82
4.8	Heltalsprogrammering .....	86
4.9	Sammanfattning .....	88
4.10	Litteraturreferenser .....	89

## Anders Karlqvist, Lars Lundqvist

5	EN PLANERINGSMODELL FÖR KAPACITETSEXPANSION .	90
5.1	Inledning .....	90
5.2	Single-expansion .....	91
5.3	Flera utbyggnader .....	93
5.4	Formulering av dynamisk programmeringsmodell	94
5.5	Beräkningsmässiga aspekter .....	97
5.6	Enkelt exempel på icke-stationära problemet i ändlig tid .....	99
5.7	Stationär planeringsmodell med oändlig tids- horisont .....	102
5.8	Tillämpningar av den stationära modellen i oändlig tid .....	105
5.9	Sammanfattning .....	112
5.10	Litteraturreferenser .....	114

## Anders Holvid

6	HELTALSPROGRAMMERINGSMODELLER FÖR DYNAMISK LOKALISERING .....	115
6.1	Problemformulering .....	115
6.2	Modellförslag .....	116
6.3	Litteraturreferenser .....	120

## Åke E Andersson

7	LOKALISERINGSANALYS FÖR STOCKHOLMSREGIONEN ..	121
7.1	Lokaliseringsteorin som en del av teorin för allmän ekonomisk jämvikt .....	121
7.2	Beräkning av överflyttningsvinster vid ökad rörlighet .....	124
7.3	Icke-konvexa egenskaper hos inomregionala lo- kaliseringsproblem. Förekomst av stordrifts- fördelar .....	130
7.4	Transportmodeller som hjälpmedel för inom- regional lokaliseringsanalys vid odelbarheter	135
7.5	Ömsesidiga beroenden som komplicerande faktor vid odelbarheter .....	140
7.6	En väg att lösa de tekniskt bestämda ömse- sidiga beroendenas problem .....	141
7.7	Värdering med en dynamisk transportmodell ...	147
7.8	Kan prognoser över lokaliseringsmönstret upp- rättas med rekursiv programmering? .....	150
7.9	Dynamiska problem rörande ekonomisk tillväxt, ränta och optimal planhorisont .....	153
7.10	Ett förslag till stegvis genomförande av lo- kaliseringsanalys för Stockholmsregionen ....	154
7.11	Litteraturreferenser .....	156

## Johan Lönnroth

8	OM KRITERIER FÖR MARKANVÄNDNINGSBESLUT .....	157
8.1	Bakgrund och syfte .....	157
8.2	En modell för markanvändningsbeslut .....	158
8.3	Ett forskningsprogram .....	165
8.4	Litteraturreferenser .....	167
	NÅGRA ORDFÖRKLARINGAR .....	168
	MATEMATISKA SYMBOLER .....	175



## FÖRORD

Denna rapport innehåller en samling föredrag som hölls vid ett forskarseminarium i Stockholm den 15-16 december 1970. Temat för seminariet var dynamiska lokaliseringsmodeller och syftet att diskutera olika problem kring tids- och rumsdimensionen i samhällsplaneringsmodeller med tyngdpunkten på matematiska modeller och därmed förknippade metodfrågor.

Materialet skall inte uppfattas som en originalforskningsprodukt. Några föredrag är av översiktskaraktär, andra är referat av tidigare forskningsarbeten och i några fall är det fråga om en diskussion kring egna ej slutgiltigt genomarbetade forskningsidéer.

En del av materialet i denna rapport har arbetats om efter seminariet, bl a med hänsyn till de synpunkter som framkom vid diskussionerna där. Skriften skall dock fortfarande betraktas som ett diskussionsunderlag snarare än en slutprodukt.

Rapporten innehåller en rad matematiska, tekniska och ekonomiska ord och symboler. För att i någon mån underlätta läsningen har de viktigaste samlats och förklarats i en fristående ord- och symbolförteckning som placerats sist i rapporten.

Medverkande med föredrag var, förutom undertecknade arrangörer av seminariet, fil lic Åke Andersson, Generalplaneberedningen, Stockholm, civ ing Anders Holvid, Matematiska institutionen, KTH och fil lic Johan Lönnroth, Nationalekonomiska institutionen, Göteborgs Universitet. Vi tackar för alla värdefulla bidrag.

Projektet har finansierats av Statens råd för byggnadsforskning.

Stockholm den 9 februari 1971  
Matematiska institutionen, KTH

*Anders Karlqvist      Lars Lundqvist*  
*Bertil Marksjö      Folke Snickars*



### 1.1 Vad är dynamiska system

Som framgår av rubriken till rapporten, "Dynamiska lokaliseringsmodeller", har vi velat framhäva två speciella aspekter på modellbyggande: behandlingen av tidsdimensionen och rumsdimensionen.

Det rumsliga fördelningsproblemet har belysts från många håll i lokaliseringsmodeller av olika slag. De flesta av dessa ansatser, liksom f ö även stadsplaner etc, är statiska och ger enbart en ögonblicksbild av ett tänkt framtida tillstånd.

Tidsproblemet finns behandlat, kanske framför allt i den ekonomiska litteraturen, men i den ekonomiska tillväxtteorin tas i allmänhet ingen hänsyn till de rumsliga fördelningsaspekterna.

Att skapa teoribygnader inom samhällsplaneringen, där dessa båda aspekter - rum och tid - är kopplade, är en angelägen uppgift för forskningen (för fysikens del ordnade Einstein denna koppling genom sin relativitetsteori).

I denna inledning kommer vi att hålla oss på ett allmänt, abstrakt plan och exemplen är delvis hämtade från vad som inom matematiken skulle kallas abstract nonsense. Låt oss börja med att skärskåda begreppet dynamiskt system och framför allt då hur man handskas med detta inom fysiken och systemanalysen.

Det finns en rik flora av artiklar om modeller inom samhällsplanering, där man sysslar med att klassificera modeller på olika sätt. En vanlig sådan systematisering är följande:

1. Deterministisk statisk jämviktsmodell
2. Probabilistisk statisk jämviktsmodell
3. Deterministisk diskret tidsmodell
4. Deterministisk kontinuerlig tidsmodell
5. Stokastisk diskret tidsmodell
6. Stokastisk kontinuerlig tidsmodell.

I uppställningen ovan är modellerna ordnade efter växande komplexitet. Samtidigt bör påpekas att nästan samtliga modeller, som på något sätt praktiskt användes idag, är av typ 1. Man kan fråga sig varifrån dessa idéer kommer och om en dylik klassificering kan leda modellutvecklingen vidare för samhällsplaneringens del.

De flesta av oss torde ha sina grundföreställningar om dynamiska system rotade i den klassiska mekanikens världsbild. Man tänker på partiklar som rör sig med rörelsen bestämd av Newtons mekanik t ex.

$$F = m \frac{\partial^2 x(t)}{\partial t^2}$$

I den moderna kvantfysiken har man lämnat detta deterministiska betraktelsesätt och arbetar istället med sannolikhetsfördelningar, vilket i sin tur leder till stokastiska differentialekvationer.

Att representera dynamiken med hjälp av matematiska modeller som t ex 1 - 6 är typiskt inom fysiken och har kanske därför också fått påverka synen på modeller för samhällsplanering. Utan att ta ställning till om denna överstrålningseffekt är av godo eller ej borde det emellertid vara av intresse att se på de underliggande begreppen och idéerna snarare än de matematiska ekvationerna. Dessa idéer har givits en mera abstrakt allmän-giltig form genom systemteorien.

Systemteoriens betraktelsesätt är alltför formaliserat för att direkt tjäna våra mera jordnära syften men kan vara en lämplig referensram för de begrepp och samband, som kommer att ventileras i de kommande avsnitten.

Vi börjar med att ge en definition av dynamiskt system efter Kalman, (1969).

DEFINITION (något förenklad):

Ett dynamiskt system, betecknat  $\Sigma$ , är en uppsättning mängder och avbildningar som kan sammanfattas på följande sätt:

$$\Sigma = (T, X, U, Y, \varphi, \eta)$$

De storheter, som karakteriserar systemet, kan alltså uppdelas i

- a) mängder
- T tiden
- X tillståndet vid tiden  $t = 0$
- U input, där ett ingående element betecknas  $\omega$
- Y output
- b) avbildningar
- $\varphi : T \times U \times X \rightarrow X$  tillståndsöverföringsfunktion
- $\eta : T \times X \rightarrow Y$  resultatfunktion

Därmed förstås att tillståndsöverföringsfunktionen  $\varphi$  avbildar det tillståndsrum som genereras av mängderna T, U och X tillsammans på rummet X enbart. Den kan t ex ange hur ett tillstånd vid tidpunkten 0 via inputpåverkan överförs i ett tillstånd vid en senare tidpunkt,  $\varphi$  uppfyller naturliga villkor som

tiden enkelriktad  $\varphi$  definierad för  $t \geq 0$

konsistens (begränsningsvärdet fixerat)  $(0, \omega, x) = x$

komposition (sammansättning möjlig)  $t_1 \xrightarrow{\varphi} t_2 \xrightarrow{\varphi} t_3 = t_1 \xrightarrow{\varphi} t_3$

kausaltet (gensvar på input)  $\varphi(t, \omega, x) = \varphi(t, \omega', x)$   
om  $\omega = \omega'$  för  $(0, t)$

Dessa något abstrakta definitioner kan belysas med följande från stadsplaneringen hämtade exempel.

#### EXEMPEL

- T tiden från 1970 till 2000
- X den fysiska, ekonomiska, sociala strukturen
- U påverkan utifrån t ex inflyttning och planeringsåtgärder
- Y t ex välfärdsfunktioner
- $\varphi$  stadsbyggnadsprocessen
- $\eta$  människors och företags uppskattning av stadsmiljön.

Det finns många synonyma begrepp till de ovan givna. Några av de mest vanliga är:

$T \times X$  händelserum eller fasrum

$\varphi$  rörelse, trajektoria, lösning (av diff.ekv.)

$U$  input eller kontrollvariabel

$\omega$  överför tillståndet  $x$  till tillståndet  $\varphi(x, \omega)$ .

$\Sigma$  är ett fritt system om  $U$  endast har ett element, dvs omgivning-  
en är det enda input. Exempel: Solsystemet. Den enda påverkan  
är de inre krafter som verkar. Krafterna bestäms av gravitationen.

Samhällsplaneringssystem är ej fria.

Självklart är ovanstående definition alldeles för vid och inne-  
håller för litet struktur för att man skall kunna nå intressanta  
resultat. Det är därför aktuellt att pålägga villkor av olika  
slag som dels är realistiska, dels gör systemen hanterbara.

## 1.2 Tidsinvarians

Invariansbegreppet är fundamentalt. Man kan om man så vill upp-  
fatta sökandet efter invarianta samband och storheter som en  
grundidé i all vetenskaplig verksamhet. Med ovanstående defini-  
tion av dynamiskt system kan en precis tolkning av begreppet  
tidsinvarians ges:

$$a) \quad \omega(t) \in U \Leftrightarrow \omega(t + \tau) \in U$$

$$b) \quad \varphi(t, x, \omega(t)) = \varphi(t + \tau, x, \omega(t + \tau))$$

$$c) \quad \eta(t, x) \text{ oberoende av } t.$$

Dessa villkor, som alltså skall gälla samtidigt betyder att de  
inputrelationer eller den tillståndsöverföringsstruktur och den  
outputavbildning som finns i systemet, ej ändras över tiden.

Invariansegenskaperna är typiska för naturvetenskapliga fe-  
nomen, men är inte alltid tillämpbara i samhällssystem. Nytt-  
och välfärdsbegreppen kan antas vara föränderliga över tiden,  
och en av de viktiga frågor, som diskuteras senare (framför allt  
i Bertil Marksjöes artikel), är hur man skall handskas med dessa  
invariansproblem. Hur skall nyttor mätas och jämföras mellan  
olika punkter i rummet och vid olika tidpunkter? I vilken mån  
är diskontering ett användbart hjälpmedel? Hur skall diskonte-  
ringsräntan bestämmas? Hur skall tidshorisontproblemet behandlas?

I detta sammanhang kan det vara av intresse att beröra det så kallade ergodproblemet. Ergodhypotesen i fysiken uttalar sig om möjligheten att ersätta fördelningar över tiden med fördelningar över rummet (över partiklar etc). Den innebär att man avstår från att i detalj följa ett systems tillstånd i tiden utan i stället studerar en tänkt samling kopior av systemet (en s k ensemble), och postulerar att tidsmedelvärdet för en storhet i systemet är lika med medelvärdet över ensemblen vid en fix tid. De fundamentala och svåra problem, som är förknippade med denna hypotes, har länge sysselsatt fysikerna.

Orsaken till att ergodicitet nämns här är att denna idé faktiskt dyker upp i mer eller mindre förklädd skepnad inom samhällsplaneringen, sannolikt med mycket mindre framgång än i fysiken. En vanlig typ av prediktion är s k trendkorrelation.

#### EXEMPEL

Man ser på en grupp som idag är lågavlönad men i någon framtid förväntas bli högavlönad. Istället för att studera hur dess preferenser förändras över tiden när lönerna ökar, betraktar man de lågavlönade idag tillsammans med en högavlönad grupp idag och drar den ergodiska slutsatsen, att det framtida tillståndet motsvarar preferenserna hos höglönegruppen idag. Ett annat typiskt exempel är beräkning av framtida biltäthet genom jämförelse med förhållandena i USA.

I vilka sammanhang kan det vara vettigt att använda dylika metoder? Ett uppenbart fall när denna princip inte fungerar i fysiken är vid fasomvandlingar (t ex övergång från vatten till is). Lika uppenbart är att detta ergodiska resonemang inte gäller i samhällssystem då nya innovationer introduceras, t ex övergången från "bilfasen" till en ny "transportfas". Dessa och andra prediktionsproblem behandlas i Folke Snickars artikel.

#### 1.3 Jämvikt

Jämvikt och balans är något diffusa begrepp, som (kanske därför) användes flitigt i samhällsplanering. Begreppet jämvikt har ju dessutom en lång tradition i ekonomisk teori. Det är viktigt att notera att jämvikt är ett statiskt begrepp och relaterar till

den inre strukturen i systemet.

Kan detta statistiska betraktelsesätt tillämpas på dynamiska problem? Svaren på den frågan hänger ihop med frågan om hur ett system som inte är i jämvikt reagerar. Fysikerna är även här lyckligt lottade. Ett fysikaliskt system i ojämvikt övergår mycket snabbt i jämvikt. Tiden detta tar, relaxationstiden,  $\tau$ , kan för partikelsystem uppskattas ur formeln

$$\tau \propto d^2/v$$

där  $d$  är den fria medelväglängden och  $v$  medelhastigheten. För heliumgas ger det  $\tau \approx 10^{-7}$  sekunder.

Om en liknande mekanism finnes i urbana system, kan man lätt konstatera att relaxationstiden för t ex ett innovationsförlopp skulle vara dagar eller timmar efter införandet av en ny transportled och kanske år för stora förändringar i lokaliseringsmönstret. Från kemin finns exempel där flera olika processer löper samtidigt och där en av dessa processer har mycket längre relaxationstid än de övriga och därigenom blir styrande. Liknande fall kan sannolikt tänkas förekomma för processer i samhället.

En slutsats av detta resonemang är att komparativa statistiska jämviktsmodeller av typ 1 och 2 inte kan vara tillräckliga för studier av dynamiken i urbana system. Jämviktsresonemangen, om man utgår från sådana, måste kompletteras med studier av transienterna, dvs av de processer som leder till nya jämviktslägen.

#### 1.4 Tidsfördröjning

I den systemanalytiska terminologien kan vi definiera fördröjningen  $s$  i ett system  $\Sigma$ : För alla  $\omega, \omega'$  som är definierade för perioden  $(0, t)$  och sådana att  $\omega = \omega'$  i intervallet  $(0, t-s)$  gäller att  $\varphi(\omega, x) = \varphi(\omega', x)$  dvs de av en viss input orsakade förändringarna i systemet uppträder tidigast tiden  $s$  efter input.

I samhället finns många tidsfördröjningar inbyggda. Forrester diskuterar sådana effekter i "Urban Dynamics", Forrester (1969), och använder i sina ekvationer t ex 20 års tidsfördröjning mellan faktiska förhållanden i ett område och uppfattningen av dessa förhållanden bland grupper inom de lägsta sociala skikten i samhället. Han säger också i sin beskrivning av samhället som ett "counterintuitive" system:



*"From all normal personal experience one learns that cause and effect are closely related in time and space. A difficulty or failure of the simple system is observed at once. The cause is obvious and immediately precedes the consequence. But in complex systems all of these facts become fallacies. Cause and effect are not closely related either in time or in space. Causes of a symptom may actually lie in some far distant sector of a social system. Furthermore, symptoms may appear long after the primary causes."*

Enkla reglertekniska exempel visar också att system med tidsfördröjningar kan ge instabila responser för vissa kontrollåtgärder. Dessa problem kommer att beröras i ett par av de följande avsnitten.

### 1.5 Kontinuerlig och diskret tid

kontinuerlig tid omm  $T = \{\text{reella talen}\}$

diskret tid omm  $T = \{\text{heltalen}\}$

Distinktionen mellan kontinuerliga och diskreta system är i allmänhet inte kritisk utan betingad av tekniska, metodmässiga synpunkter. I klassisk fysik har man en stark tradition för kontinuerliga system, medan diskret tidsbehandling faller sig naturlig när problemen knyts till digitala datorer. Ofta upplevs diskretisering som en approximation och förenkling av ett givet kontinuerligt problem (klassificeringen i början antyder detta). Det finns fall då detta inte gäller och där en kontinuerlig ansats ger direktare lösningsmetoder. En sådan form av kontinuerlig ansats är s k händelsestyrda modeller i motsats till tidsstyrning. Termerna är hämtade från datorsimulering. I en händelsestyrd modell låter man tidsförloppet vara kontinuerligt men noterar enbart de diskreta tidpunkter när något inträffar.

#### EXEMPEL

Stockholms regionplan är en tidsstyrd modell därför att man i förväg bestämt sig för att studera tidsintervallen 1970-85 och 1985-2000. Det vore möjligt i princip att göra planen händelsestyrd, t ex göra utbyggnadsplaner för Stockholm vid 1,7 miljoner invånare, 2,1 miljoner invånare etc. I en senare uppsats kommer en matematisk modell för utbyggnad av en industri på olika platser

att behandlas. Modellen är formulerad som ett dynamiskt programmeringsproblem och har denna händelsestyrda karaktär.

## 1.6 Ändliga system

Det viktigaste måttet på komplexiteten hos ett system  $\Sigma$  ges av strukturen hos tillståndsrummet  $X$ .

### DEFINITIONER

$\Sigma$  är ändligt dimensionell omm  $X$  är ett ändligt dimensionellt vektorrum

$\Sigma$  är ändlig omm  $X, U, Y$  är ändliga mängder och om dessutom  $\Sigma$  är tidsinvariant med diskret tid.

Ändliga system är den enklaste generella klassen av system som har blivit ordentligt studerade. Från en teoretisk ståndpunkt vore det naturligare att arbeta med funktionalanalysens verktyg och t ex låta  $X, U, Y$  vara oändligt dimensionella topologiska rum (Banach-, Hilbertrum) med  $\varphi$  och  $\eta$  som kontinuerliga avbildningar. Dyliga teoretiska utvidgningar har också gjorts.

Den vanligaste och viktigaste restriktionen på dynamiska system är linearitet, som vi definierar sålunda:

$\Sigma$  är lineärt om  $X, U, Y$  är vektorrum och  $\varphi, \eta$  är lineära avbildningar för alla  $t$ .

Ändlighet och linearitet ger oss tillträde till den stora mängd metoder och resultat, som den linjära algebran har att ge. Linjärprogramering är ett lysande exempel på denna teoris möjligheter. Vidare är den linjära systemteorien av stor betydelse för studiet av lokala egenskaper hos icke-linjära system.

## 1.7 Kontrollteori

Ett viktigt problem vid studier av dynamiska system är frågan om kontroll. Situationen är följande:

Givet ett system att kontrollera och en mängd av tillåtna kontrollåtgärder (inputs)  $\omega$ . Vidare förutsättes en objektfunktion eller mål (output) och mått på kostnaden eller effektiviteten av vidtagna kontrollåtgärder. Frågor man vill ha svar på är t ex

existerar det någon optimal kontroll? Om det existerar optimala kontroller, hur skall dessa i så fall bestämmas?

För ett linjärt system  $\Sigma$  kan kontrollproblemet visas vara ekvivalent (i princip) med att lösa ekvationerna

$$\frac{d x(t)}{dt} = F x(t) + G u(t)$$

$$y(t) = H x(t)$$

där  $F$ ,  $G$  och  $H$  är konstanta matriser som bestämmer systemstrukturen.

Det finns ett stort antal matematiska metoder utvecklade för sådana problem bl a olika programmeringsmetoder, Pontryagins maximumprincip, variationskalkyl etc. Lars Lundqvist uppehåller sig i sin artikel vid några av dessa metoder, och de mera praktiskt inriktade modeller som behandlas i övriga artiklar är också formulerade som kontrollproblem. Jag skall i detta sammanhang peka på ett par viktiga principiella problem, som inte bara är av akademiskt intresse, nämligen frågorna om kontrollerbarhet och observerbarhet.

Den första gäller problemet om man kan styra systemet dit man vill. Den andra frågan avser om man från systemets sätt att reagera kan sluta sig till dess inre struktur. Problemen visar sig vara duala ur matematisk synpunkt.

En händelse  $(\tau, x)$  är kontrollerbar om

$\exists t \geq \tau, \omega \ni (\tau, x) \xrightarrow{\omega} (t, x_0)$  dvs om det finns en styråtgärd  $\omega$  som överför tillståndet  $x(\tau)$  till  $x_0(t)$  inom ändlig tid.

#### EXEMPEL 1

$\tau =$  idag,  $x =$  Järvafältet,  $x_0$  en plan för Järva utbyggt med 100.000 invånare. Finns det någon planering  $\omega$  som kan förverkliga denna plan  $x_0$  tills Järva har hunnit växa till 100.000 (vid tiden  $t$ )?

En händelse  $(\tau, x)$  är uppnåelig om

$\exists s \leq \tau, \omega \ni (s, x_0) \xrightarrow{\omega} (\tau, x)$

## EXEMPEL 2

$x_0$  = Stockholm idag,  $\tau$  = år 1985,  $x$  = regionplanen för 1985.  
Finns det ett utbyggnadsprogram  $\omega$  som leder oss dit?

Exemplen ovan antyder, att dessa frågor inte alls är triviala att besvara.

I planeringssammanhang kan mängden tillåtna kontrollåtgärder (inputs)  $\omega$  vara mycket begränsad och möjligheten att påverka systemet kan i hög grad bero på beslutsprocessens sätt att fungera. Johan Lönnroth anknyter till dessa frågor i en av sina lokaliseringsmodeller.

Den teoretiska analysen visar, att ett nyckelbegrepp i detta sammanhang är s k feedback. Denna princip säger att inputs  $\omega$  skall bestämmas av tillståndet  $x$ . Detta är den fundamentala idéen i kontrollteorin och den formulerades bl a av Bellman på 50-talet.

## DEFINITION

En kontrolllag eller feedback är en funktion  $k$

$$k : T \times X \rightarrow U$$

som för varje tillstånd  $x(t)$  vid tiden  $t$  tillordnas ett värde  $\omega(t) = k(t, x(t))$ .

Grundidéen är alltså att tillståndet innehåller all information tillräcklig för att bestämma de kontrollåtgärder som skall vidtagas, eftersom (vilket följer av definition på dynamiskt system) den framtida utvecklingen bestäms fullständigt av nuvarande tillstånd och kommande inputs.

Feedbackmetoden är ytterligt kraftfull om den användes med omsorg. Om den användes oförsiktigt kan den istället förstärka den effekt den var avsedd att dämpa. Detta gäller speciellt när tidsfördröjningar finns inbyggda. Om feedback användes ur fas kan större avvikelser och oscillationer inträffa. Exempel på detta är vanliga i den ekonomiska världen. Liknande gäller också för mekaniska och elektriska system. Detta paradoxala fenomen är

intimt förknippat med studier av stabilitet. Ämnet är associerat med en extensiv matematisk teoribyggnad, men fortfarande finns många viktiga och svåra problem att lösa inom detta fält.

Implicit i ovanstående definition antas att vi alltid kan utläsa tillståndet  $x$  hos systemet. Denna information är i själva verket av två slag:

1. Kunskap om strukturen  $X, \varphi, \eta$ .
2. Kunskap om de faktiska inputs och outputs till systemet.

För en maskin, processindustri etc antas i allmänhet information av typ 1 given a priori och data av typ 2 ges av real-time-mätningar medan systemet är igång.

När data av den första typen inte är tillgängliga utan på något sätt måste bestämmas ur observerade inputs och outputs har vi ett sk identifieringsproblem. I samhällsplaneringen saknas ofta information av båda typerna. Den stora populariteten hos databanker och informationssystem är ett tecken på detta. Sådana mätningar skulle ge information av typ 2, men fortfarande återstår det svåra problemet att klarlägga de mekanismer av typ 1, som verkar inom samhällssystemet. Teoretiskt formulerat står vi inför ett adaptivt kontrollproblem. Mycket har talats om adaptiv kontroll men lite har utträttats. Kalman säger i inledningen till sin bok om systemteori: "We regard adaptive control as a problem for the future". Med dessa något pessimistiska funderingar lämnar vi här kontrollteorin åt sitt öde. För att avrunda detta inledande avsnitt sammanfattas några av de systemteoretiska begreppen i TAB. 1.1.

TAB. 1.1. Jämförelse mellan modeller av skilda dynamiska system (1,2,3) med avseende på vissa karakteristiska egenskaper som definieras i denna artikel. Klassificeringen är gjord mer eller mindre på måfå och är avsedd enbart som underlag för diskussion.

Jämförda egenskaper	Dynamiska system		
	1. Samhällsplanering lokalisering	2. Klassisk fysik	3. Kontrollteori
ett fritt system	nej	ja	nej
output	"välfärd"	trivialt $\eta(x(t)) \equiv x$	kostnad, ofta kvadratisk avvikelse från idealvärde
tidsinvarians	ibland	ja	ja
jämvt	nej	ja	-
tidsfördröjning	ja	nej	ibland
kontinuerlig tid	ibland	ja	ibland
diskret tid	ibland	nej	ofta
ändlig	ja?	ja	ja
lineärt	nej	ibland	ibland
kontrollerbart	ibland	nej	ja
feedback	ja	nej	ja
adaptiv	ofta	nej	nej
stabilitet	nej	ja	ja

1.8 Litteraturreferenser

Kalman, R E, Falb, P L, Arbib, M A, 1969, *Topics in Mathematical System Theory*. (Mc Graw Hill). New York.

Forrester, I, 1969, *Urban Dynamics* (M.I.T. Press) Cambridge, Massachusetts.

## 2 BESKRIVANDE MODELLER OCH PREDIKTION - EN ÖVERSIKT - Folke Snickars

### 2.1 Vad är en beskrivande modell?

Man brukar i modelltekniska sammanhang skilja på tre slags modeller, nämligen beskrivande, förklarande och styrande modeller.

En beskrivande modell har det syftet att ge en så god beskrivning av den studerade processen som möjligt. Därvid är det ett mål att reducera mängden primärt inblandade variabler till ett mindre antal, som ändå ger god anpassning mellan modell och data. De rent beskrivande modellerna kan dock inte alltid tillfredsställa modellbyggarens krav på sådan kunskap om processen, som skall ligga till grund för förutsägelse och planering av dess framtida utveckling. Försök att komma tillrätta med sådana svagheter leder till studium av förklarande och styrande modeller.

I en förklarande modell nöjer man sig inte med att så väl som möjligt beskriva ett samband mellan två variabler utan vill även studera orsakssambanden, så att man kan göra troligt att just den använda formen hos relationen är den rätta. En sådan modelltyp underlättar framskrivning av processens målvariabler, genom att exempelvis koppla samman deras tidsutveckling med den hos mera lättprognostiserbara variabler. I fortsättningen kommer denna modelltyp att hänföras till den utvidgade klassen beskrivande modeller.

Vad de beskrivande modellerna i den ovan angivna tappningen inte innehåller är explicita målformuleringar och vägar att uppnå dessa mål. Sådana aspekter beaktas i de styrande och planerande modellerna, vilka faller utanför ramen för detta ämne. Därmed anses begreppet beskrivande modell tillräckligt definierat, och ämnet för den fortsatta framställningen avgränsat.

### 2.2 Exempel på beskrivande modeller

Den välkända gravitationsmodellen i trafikplanering är ett typiskt exempel på en beskrivande modell. Framtida trafikflöden mellan två områden uppskattas genom prognoser för antalet boende



och arbetande i de båda områdena och genom gissning av det framtida värdet på den avståndskänslighetsparameter för resval som ingår i modellen.

Ett exempel på en enkel beskrivande lokaliseringsmodell är den s k Clarks formel för befolkningstätheten i ett stadsområde, som bygger på den empiriskt iakttagna regulariteten i boendetätheten för nattbefolkningen som funktion av avståndet till stadens centrum. Formeln anger att boendetätheten på avståndet  $r$  från centrum (variation i vinkelled ignoreras) är

$$D(r) = A e^{-br} \quad (2.1)$$

där  $A$  anger centrumtätheten och  $b$  bestämmer hur snabbt tätheten avtar när man avlägsnar sig från centrum. Vanligtvis gäller formeln dåligt för den centralaste delen av staden, men förvånansvärt bra för längre avstånd. För prognostisering av den framtida fördelningen krävs framskrivning av både  $A$  och  $b$ .

Ett tredje, och något mera omfattande, exempel på en beskrivande modell är den ekonometriska modellen för lokalisering av nya bostäder, som utvecklats vid det franska planeringsinstitutet CERAU. I följande schema (FIG. 2.1) anger pilarna påverkan mellan variablerna.

De endogena variablernas värden beror alltså dels på globala bivillkor rörande marktillgång och bostadsbestånd i regionen under den betraktade tidsperioden, dels på exogena data rörande den aktuella zonens urbaniseringsgrad och geografiska egenskaper.

Försäljningstakten exempelvis, för bostäder av en viss kategori i zonen, är en funktion av försäljningspriset hos bostadstypen i både den aktuella zonen och intilliggande zoner, bostadstypens kvalitet, dess lägeskaraktistika, tillgängligheten till arbetsplatser från zonen och tillgängligheten till näraliggande, redan existerande, bostäder av samma kategori. Alla de i tabellen angivna storheterna har kvantifierats och ett stort antal ekvationer, av den typ som ovan med ord beskrivits, har uppställts i modellen. Det så erhållna ekvationssystemet har sedan lösts med ekonometrisk teknik.

Huvudsyftet med modellen är att förutsäga den allokering av nya bostäder, som motsvarar ett givet ramprogram för regionen i

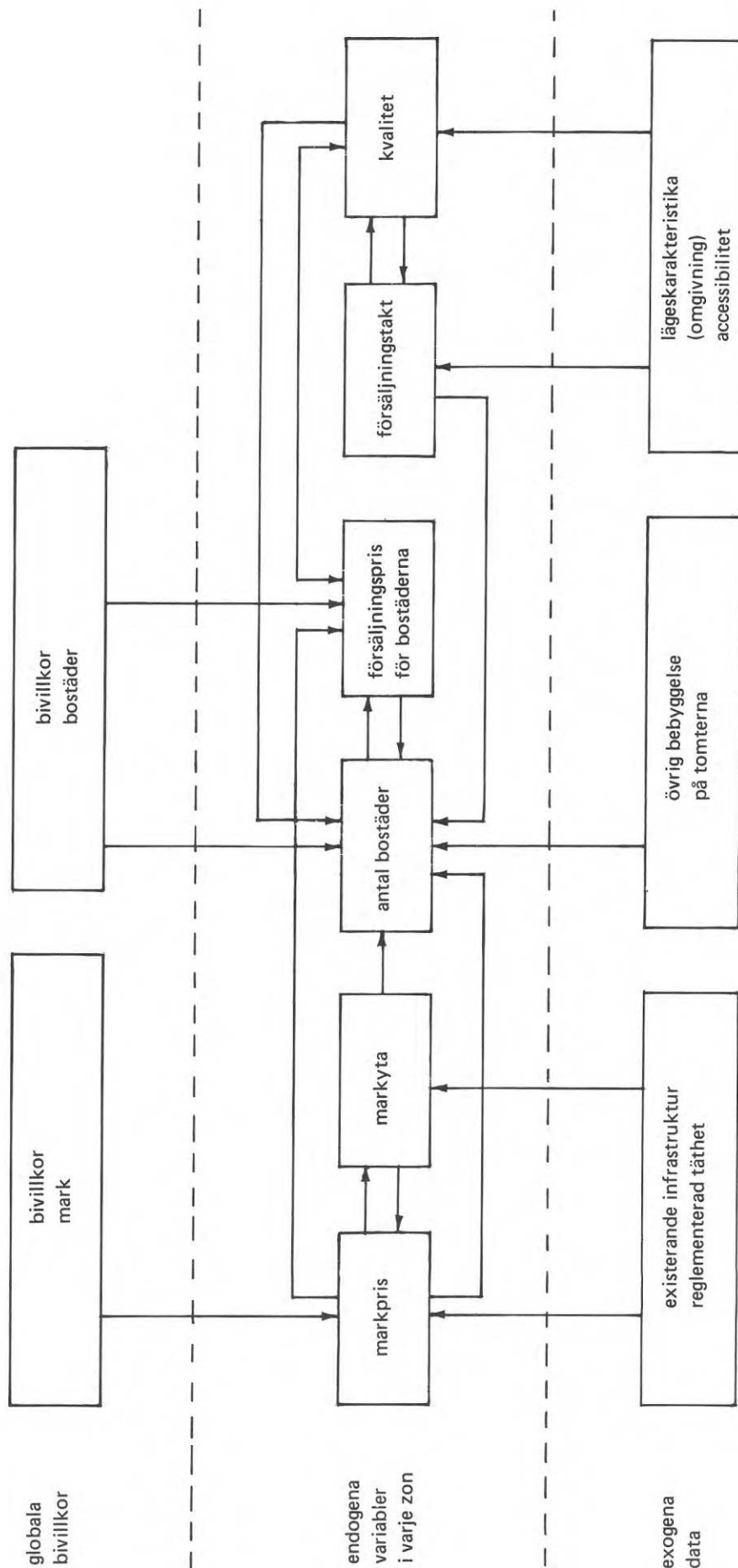


FIG. 2.1 Ekonometrisk modell, för lokalisering av nya bostäder, utvecklad av det franska planeringsinstitutet CERAU.

stort, med hänsyn till prognostiserade värden på de exogena variablerna. Problem uppstår när man vill använda modellen för andra tidsperioder och städer än de för vilka modellen kalibrerats.

### 2.3 Trender och prognoser - några synpunkter

De tre givna modellexemplen ger alla problem, när man vill använda dem för prediktion. Då ett av de viktigaste skälen till detta slag av modellbyggen brukar vara att man vill skaffa sig en bild av framtida förhållanden, kan det vara av intresse att något mera ingående behandla dessa framskrivningsproblem. En idealisk prediktiv lokaliseringsmodell skulle vara en där rums- och tidsaspekterna helt separerats, så att modellens parametrar var konstanta över tiden och allt tidsberoende fördes in i modellen via de exogena variablerna. I en sådan modell - den torde vara ovanlig - skulle godheten i prognosen främst hänga på osäkerheten i prediktionen av de exogena variablerna. Någon längre utredning av allmänna prognosmetoder för sådana exogena variabler kommer inte här att göras, utan framställningen kommer att inskränkas till en genomgång av några vanliga trendanalysmetoder.

I fortsättningen tänkes  $y(t)$  vara en exogen variabel som skall prognostiseras. Beroende på hur mycket information som finns om  $y(t)$  och dess tidigare utveckling kan någon av nedanstående metoder utnyttjas.

1. Ingen förändring
2. Linjär trend
2. Polynomtrend
4. Exponentiell trend
5. Mättnadstrend
6. Glidande medelvärden
7. Differensekvationer
8. Trendkorrelation

Samtliga dessa trender grundar sig på data för en tidsserie av observationer på  $y(t)$ . Föreligger ingen ytterligare kännedom om egenskaperna hos  $y(t)$ , kan någon av de fyra första metoderna tillämpas, beroende på tidsseriens information om förloppet

hos  $y(t)$ . Den andra metoden bygger t ex på att tillväxt med konstant belopp observerats i tidsserien, den fjärde på tillväxt med konstant andel. Har man dessutom anledning att anta att  $y(t)$  går mot ett mättnadsvärde är det vanligt att den exponentiella trenden modifieras till en mättnadstrend genom att anta att

$$\frac{dy}{dt} \cdot \frac{1}{y} = l(k - y).$$

Direkt integration ger  $y(t) = \frac{k}{1+m \cdot e^{-kl(t-t_0)}}$ ,  $t \geq t_0$ .

$m = \frac{k-y(t_0)}{y(t_0)}$ .  $y(t)$  brukar i detta fall kallas en logistisk kurva.

Glidande medelvärden används t ex om observationerna på  $y(t)$  fluktuerar kraftigt och misstänks innehålla en stor stokastisk komponent. Ett sätt att bilda sådana medelvärden utgörs av exponentiell smoothing. Om  $S(t-1)$  är det utjämnade värdet vid tiden  $t-1$  och  $y(t)$  är observationen vid tiden  $t$  är - i det enklaste fallet - det utjämnade värdet vid tiden  $t$ ,  $S(t) = \alpha y(t) + (1-\alpha)S(t-1)$ . Om man anser observationen  $y(t)$  säker kan detta beaktas genom att välja andelen  $\alpha$  nära ett, och om den är osäker ges  $S(t-1)$ , dvs hela tidsserien av observationer, större vikt genom att minska  $\alpha$  mot noll.

Tror man sig att med någon säkerhet kunna härleda en differensekvation för  $y(t)$ , dvs om man explicit lyckas erhålla  $y(t) = f(y(t-1), \dots, y(t-k))$ , ( $t$  är diskret), kan man ibland analytiskt lösa ut  $y = g(t)$ . Det går naturligtvis också att utnyttja differensekvationen direkt. Metoden har den svagheten att man inte vet om parametrarna i differensekvationen är konstanta över tiden.

Med trendkorrelation menas i detta sammanhang att om  $y(t)$  är svårprognostiserad, kan man göra en analys av tidsutvecklingen av en med  $y(t)$  positivt korrelerad variabel, vars tidsberoende är enklare att analysera. Förutsättningen är alltså att man har påvisat hög korrelation mellan variablerna under tidigare tidsperioder.

De ovan gjorda anmärkningarna är huvudsakligen av teknisk natur och självfallet är exempelvis en prognos av den framtida

totalbefolkningen i Stockholmsregionen betydligt mera komplicerad. I en sådan flerdimensionell prognos kan någon av metoderna ovan möjligen ingå som en liten del.

#### 2.4 Regressionsanalys och parameterestimation

En av de viktigaste uppgifterna vid modellarbete är estimation (uppskattning) av modellens parametrar. Först kan den anmärkningen göras, att detta är en välutvecklad gren av statistiken - det är lätt att med regressionsanalytiska metoder få fram korrekta parameterestimater under ganska allmänna antaganden om systemstrukturen. Problem uppstår speciellt för tids- och rumsserier där gjorda observationer inte alltid kan anses oberoende av varandra. Några synpunkter på dessa problem lämnas nedan med utgångspunkt från teorin för stokastiska processer.

Med en stokastisk process  $Y(\omega, t)$  menas en process som är en funktion av dels  $\omega \in \Omega$ , utfallsrummet för processen, dels  $t$ , som är en parameter, vanligen tiden. Ett visst utfall av den stokastiska komponenten genererar en deterministisk kurva som funktion av tiden - en realisering av processen. För fixa  $t$  är  $Y$  en stokastisk variabel med någon given fördelning. Intressanta storheter är medelvärdet ( $\omega$  undertrycks)

$$E \{ Y(t) \} = m(t)$$

och autokorrelationen

$$E \{ Y(t+\tau)Y(t) \} = \varphi(t+\tau, t)$$

I en diskret stokastisk process antar  $t, \tau$  bara diskreta värden. Om processen är svagt stationär, betyder detta att  $m(t) = m$  och  $\varphi(t+\tau, t) = \varphi(\tau)$ , dvs medelvärdet är konstant över tiden, och autokorrelationen beror bara på tidsskillnader mellan de betraktade tidpunkterna.

En regressionsekvation av enklaste slag,  $Y_t = \alpha + \beta X_t + u_t$ , representerar en stokastisk process.  $X_t$  är vanligen deterministisk (speciellt i tidsserier) eller också brukar regressionen utföras betingat av att  $X_t$  är givet, varför  $u_t$  är den enda stokastiska komponenten i ekvationen. Gör man om  $u_t$  de tre antagandena

$$E \{ u_t \} = 0, E \{ u_{t+\tau} \cdot u_t \} = 0 \text{ för } \tau \neq 0 \text{ och } E \{ u_t^2 \} = \sigma_u^2,$$

erhålles fullständigt oberoende mellan observationerna, som kan ses som en svit oberoende, likafördelade stokastiska variabler. I så fall kan vanlig minsta-kvadratregression användas för estimation av parametrarna  $\alpha$  och  $\beta$ .

Ett exempel på en rumsserie, där sådan estimation är möjlig, utgöres av befolkningstäthetsobservationer för testning av Clarks formel. Genom att införa  $Y_t = e^{\log D_r}$  och  $X_t = r$  i ekv. (2.1) överförs formeln till den ovan givna formen. I allmänhet är rums-serier svårare att handskas med än tidsserier p g a rummets flerdimensionalitet.

Om man släpper på det tredje av antagandena, kommer man att få  $\sigma_u^2 = \sigma_u^2(t)$ , medan observationerna fortfarande är oberoende. Resultatet blir förändringar av variansen i parameterestimatet. Om det i stället antas att det andra antagandet ovan inte är giltigt, så innebär detta att  $E \{ u_{t+\tau} \cdot u_t \} = \varphi(\tau)$ , dvs observationerna är inte oberoende, och man har att göra med en svagt stationär stokastisk process. I exemplet om befolkningstätheten skulle detta innebära att ett högt täthetsvärde på ett visst avstånd hade en tendens att medföra höga tätheter också i närheten. En sådan effekt torde inte vara osannolik. Ytterligare en allmän anledning till att en sådan autokorrelation förekommer kan vara att regressionsekvationen är felaktig, dvs inte ger en rättvisande bild av variationen i data.

Tillämpning av vanlig regressionsanalys på autokorrelerade data ger inga systematiska fel, vilket betyder att parameterestimatet fortfarande blir teoretiskt medelvärdesriktiga, men däremot kommer variansen i estimaten att underskattas. Känner man autokorrelationsfunktionens form kan man korrigera variansen.

Ett ytterligare exempel på ett fall när autokorrelation förekommer i rumsserier finns i den ovan presenterade modellen från CERAU, där försäljningstakten för en viss kategori bostäder i en zon även påverkades av försäljningspriset hos bostäderna i näraliggande zoner.

De hittills behandlade komplikationerna med minsta-kvadratregression leder inte till felaktiga estimat i den meningen att dessa inte skulle vara teoretiskt medelvärdesriktiga. Om man

emellertid vill dynamisera en regressionsekvation genom att införa tidsfördröjda variabler i högerledet, kan man råka ut även för detta. Konsekvensen blir att parameterestimaten blir systematiskt felaktiga, speciellt om man har små datamängder.

Ett sätt att glida undan estimationsproblemen i tids- och rumsserier bygger på ergodhypotesen. Med hjälp av denna hypotes kan man ersätta en tidsserie av observationer på en stokastisk process med en svit statistiska observationer som består av en samling utfall av processen vid en viss tidpunkt. Eftersom de statistiska parametrarna i många sammanhang uttrycks som tidsmedelvärden, brukar hypotesen ofta formuleras: En stokastisk process  $Y(t)$  är ergodisk om tidsmedelvärden är lika med ensembleredelvärden (vid en viss tidpunkt).

Ett exempel på en ergodisk filosofi skulle vara att i stället för att studera parametrarnas variation över tiden i Clarks formel med hjälp av tidsseriedata, göra en tvärsnittsanalys över en ensemble av städer av olika storlek, och prediktera variationer i tiden genom att gå in på en annan nivå i ensemblen. Metoden lider av bristen att man måste känna till något om anpassningshastigheten vid övergångar mellan olika nivåer (invånarantal) och begränsas av att ensemblen inte får innehålla städer med helt annorlunda transportstruktur än den som finns i den betraktade staden.

Ovan har en översikt gjorts av några tekniska problem vid parameterestimation i tids- och rumsserier, varvid framställningen hela tiden rört de klassiska regressionsanalytiska metoderna. Härnäst kommer en modifierad estimationsmetod att behandlas, med tillämpning på ett speciellt, redan tidigare introducerat, problem.

## 2.5 En statistisk modell med maximum-likelihoodestimering

Den statistiska modell som skall beskrivas bygger på den modellteori som introducerats av Martin-Löf (1970). Tanken är, att härleda en modell för befolkningstätheten i ett stadsområde, som har ett något högre förklaringsvärde än den empiriska Clarks formel.



Modellen skall beskriva befolkningstäthetens variation i ett stadsområde och dess tidsutveckling under  $m$  år. Staden antas indelad i  $n$  områden. I det  $j$ :te området, som har yta  $s_j$ , är den förväntade boendetätheten under det  $i$ :te året lika med  $D_{ij}$ . Stadens totala invånarantal under år  $i$  är  $N_i$ . Det antas dessutom att den viktigaste faktorn för täthetsmönstrets utseende är fågelvägsavståndet till centrum, betecknat  $r_j$  för det  $j$ :te området. Dessutom antas medelavståndet till centrum för hela befolkningen vara  $\bar{r}_i$  det  $i$ :te året. Det gäller att

$$\sum_{j=1}^n D_{ij} s_j = N_i, \quad \sum_{j=1}^n r_j D_{ij} s_j = \bar{r}_i \cdot N_i = w_i$$

Man vill studera i vilken utsträckning det räcker att beskriva befolkningstäthetens variation genom att ange paren  $(N_i, w_i)$  i stället för att ange hela sviten tätheter  $(D_{11}, \dots, D_{mn})$ .

Genom utnyttjande av den s k Boltzmanns lag kan man visa, att den förväntade tätheten i det  $j$ :te området under det  $i$ :te året blir

$$D_{ij} = \frac{N_i e^{-b_i r_j}}{\sum_{j=1}^n s_j e^{-b_i r_j}} = A_i e^{-b_i r_j}$$

(För en närmare förklaring se Martin-Löf (1970)).

Om summan approximeras med en integral blir resultatet

$$D_i(r) = A_i e^{-b_i r}, \quad \text{där } A_i = \frac{b_i^2 N_i}{2\pi(1 - (1 + b_i R_M)e^{-b_i R_M})}$$

vilket inte är något annat än Clarks formel (sid.21).

$R_M$  betecknar den godtyckligt valda stadsgränsen. Man kan dessutom visa att

$$\bar{r}_i + \frac{d}{db_i} \log \frac{N_i}{A_i} = 0, \quad \text{varav } b_i \text{ kan uppskattas.}$$

För övrigt kan det påpekas, att detta estimat är liktydigt med det klassiska maximumlikelihoodestimaten för  $b_i$ .



Denna härledning av Clarks formel visar kopplingen mellan tidsutvecklingen av parametrarna  $A$  och  $b$  och förändringarna i totalbefolkningen och dennas medelavstånd till centrum. Formeln ovan kan naturligtvis också skrivas som

$$D(r,t) = A(t) e^{-b(t)r} \quad (2.2)$$

Eftersom denna formel inte beaktar någon variation i vinkelled, kan man multiplicera med  $2\pi r$  och erhålla tätheten per radiell längdenhet.

$$G(r,t) = 2\pi A(t) r e^{-b(t)r} \quad (2.3)$$

Lika enkelt är det att integrera över en cirkel med radien  $R$  och få den förväntade totalbefolkningen därinnanför.

$$P(R,t) = \frac{2\pi A(t)}{b^2(t)} \left[ 1 - (1+b(t)R)e^{-b(t)R} \right] \quad (2.4)$$

Genom att helt enkelt tidsderivera ekvationerna (2.2)-(2.4) kan en del intressanta iakttagelser göras. Under antagandena att  $\frac{db}{dt} < 0$  och  $\frac{dA}{dt} > 0$  kommer tidsderivatornas utseende, som funktion av  $r$ , i princip att bli enligt FIG. 2.

De båda första kurvorna anger, att tätheten minskar innanför avståndet  $\frac{dA}{db} \cdot \frac{1}{A}$  under ett tidsintervall runt den betraktade tidpunkten  $t$  och ökar utanför. För Stockholm 1957 ligger avståndet vid ungefär 4 km, och fram till 1967 har det rört sig ut till ungefär 6 km. Den tredje kurvan visar hur totalbefolkningen i regionen omfördelas.  $R_0$  betecknar det avstånd innanför vilket totalbefolkningen är oförändrad under ett litet tidsintervall runt  $t$ .  $R_0$  varierar för Stockholm under tiden 1957-1967 från 8 till 15 km, vilket är en avsevärd ändring.

En förutsättning för att kunna göra kvantitativa uttalanden om egenskaperna hos kurvorna är att man känner  $A(t)$  och  $b(t)$ . Studier av tidsutvecklingen av parametrarna för en 40-årsperiod mellan 1960-2000 har utförts för Stockholm, varvid data för 1985 och 2000 har tagits från de framlagda regionplaneförslagen. Några exempel på resultat från denna studie redovisas i FIG. 2.3-2.7. Beteckningen "utsträckt" och "kompakt" syftar på de två

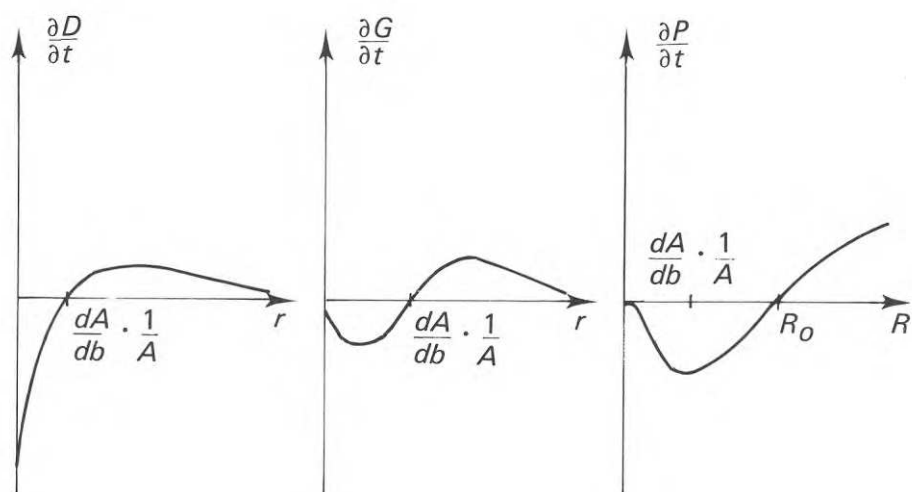


FIG. 2.2 Tidsderivator av ekvationerna (2.2) - (2.4).

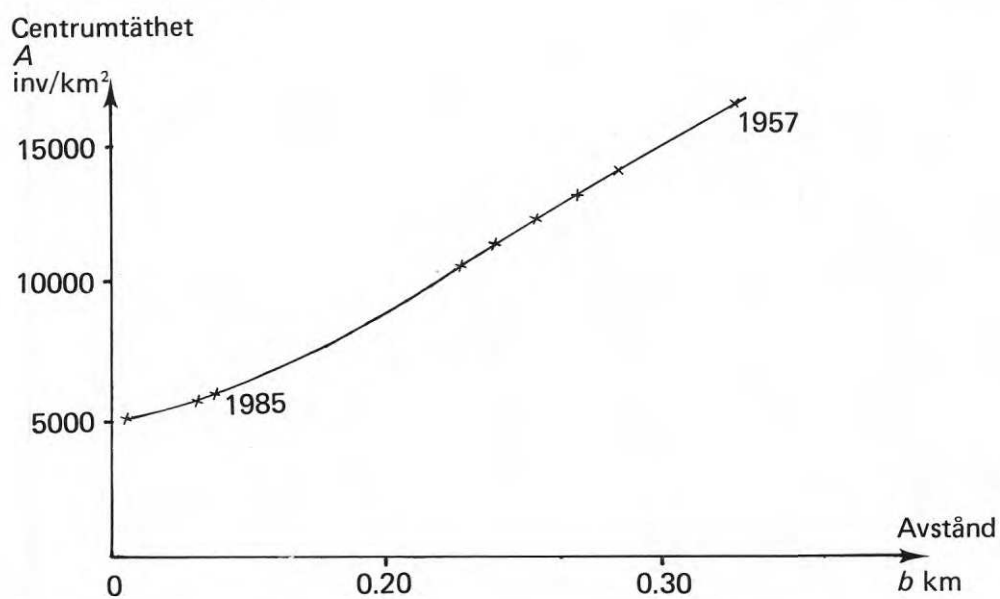
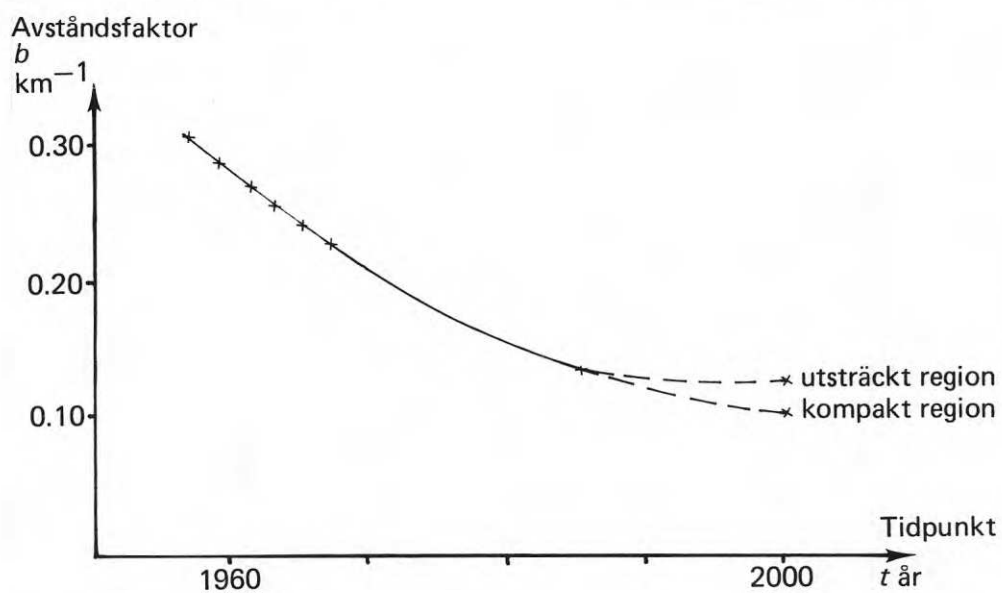
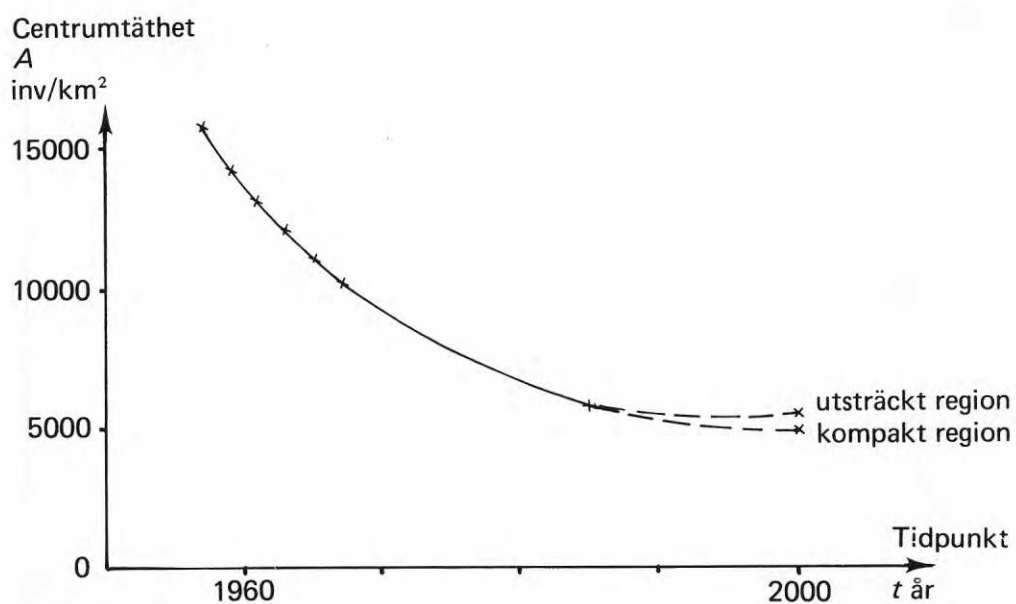


FIG. 2.3 Parametrarnas tidsberoende i Clarks formel.  
Stockholm år 1955-2000, nattbefolkning.

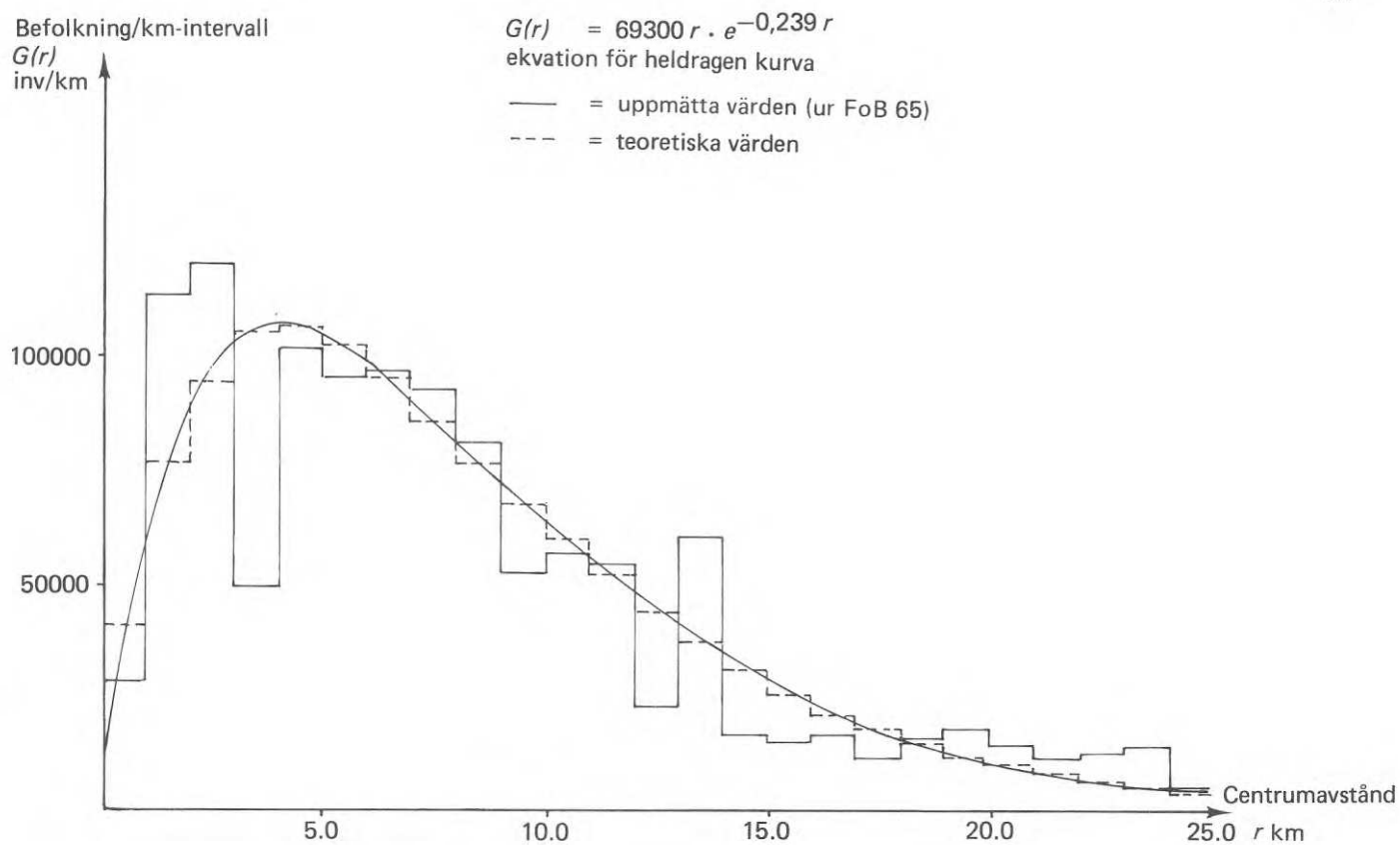


FIG. 2.4 Nattbefolkningens fördelning i Stockholm år 1965 med hänsyn till avståndet från centrum.

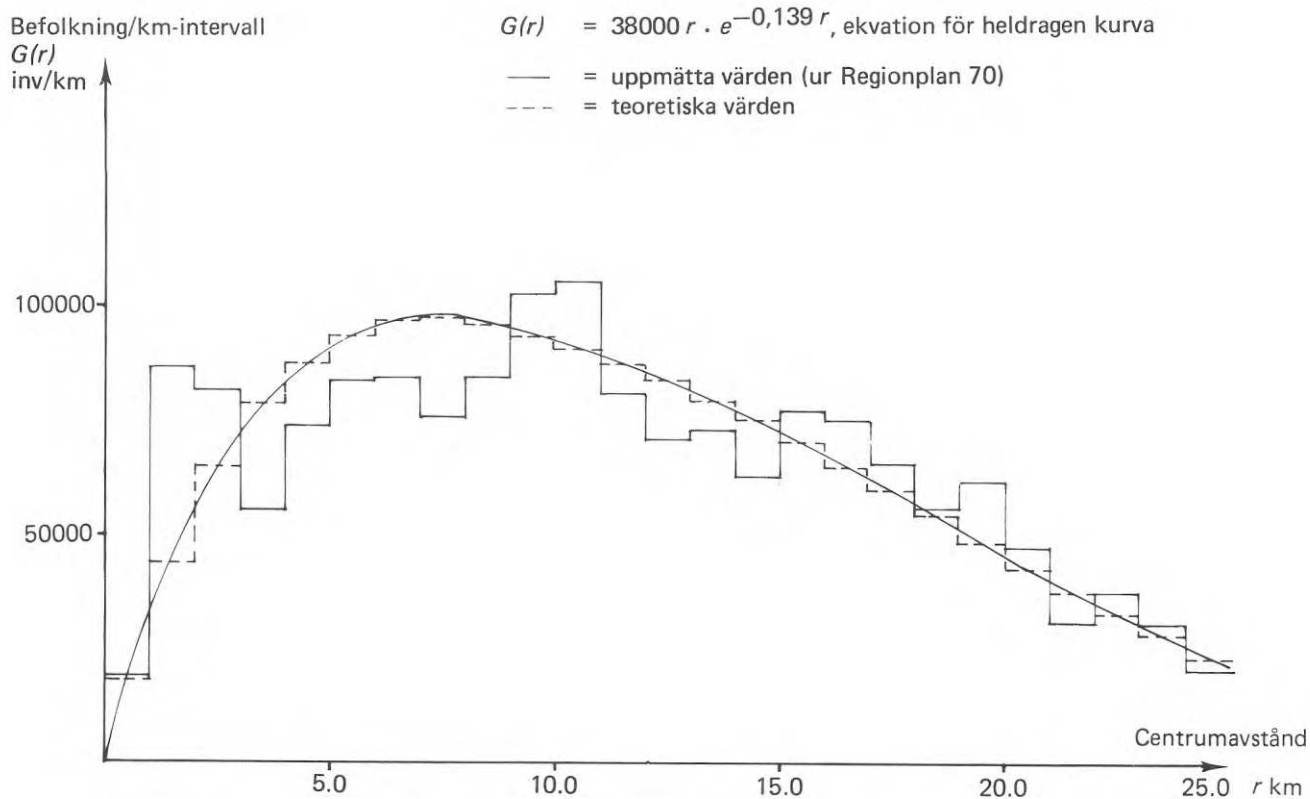


FIG. 2.5 Nattbefolkningens beräknade fördelning i Stockholm år 1985 med hänsyn till avståndet från centrum.

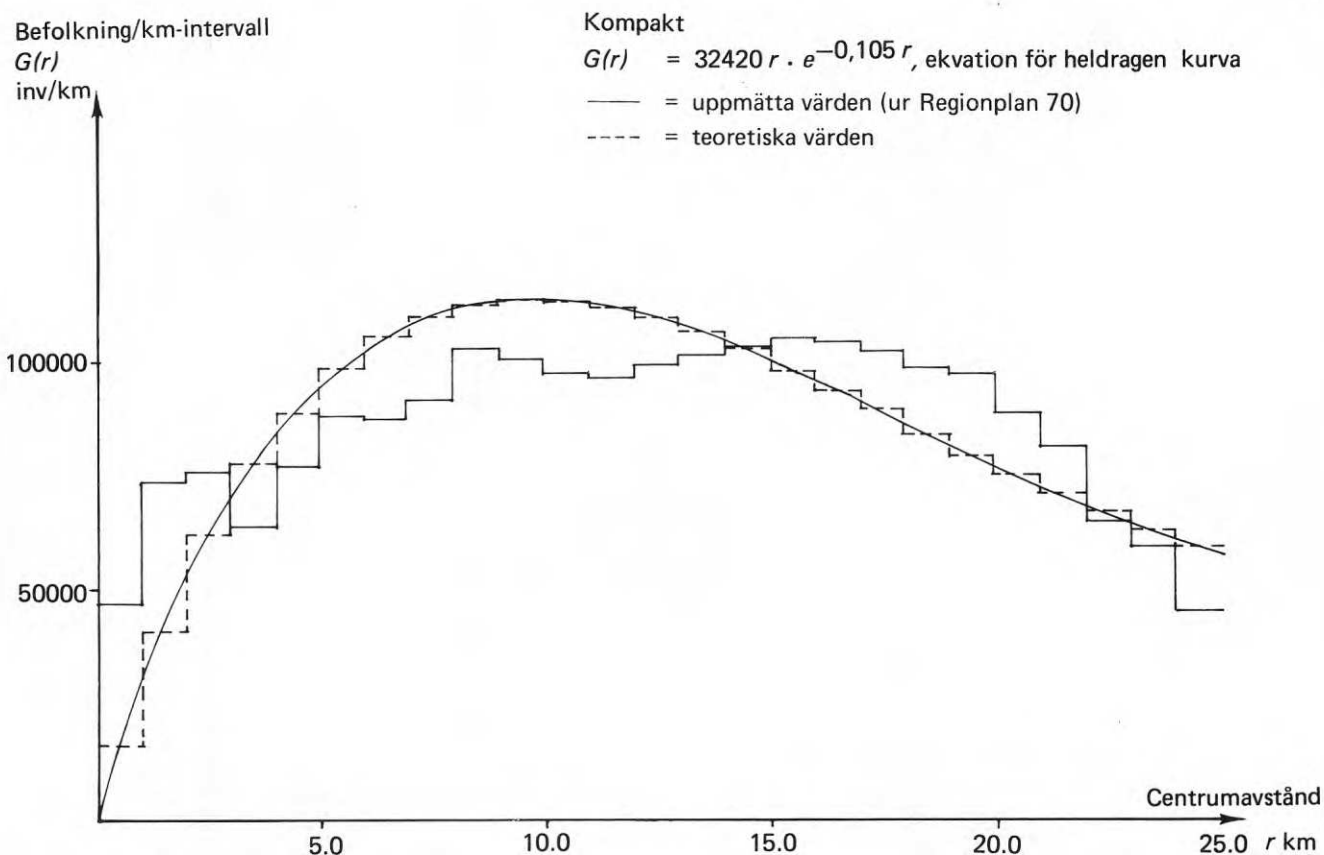


FIG. 2.6 Nattbefolkningens beräknade fördelning i Stockholm år 2000 med hänsyn till avståndet från centrum för alternativet "Kompakt stad".

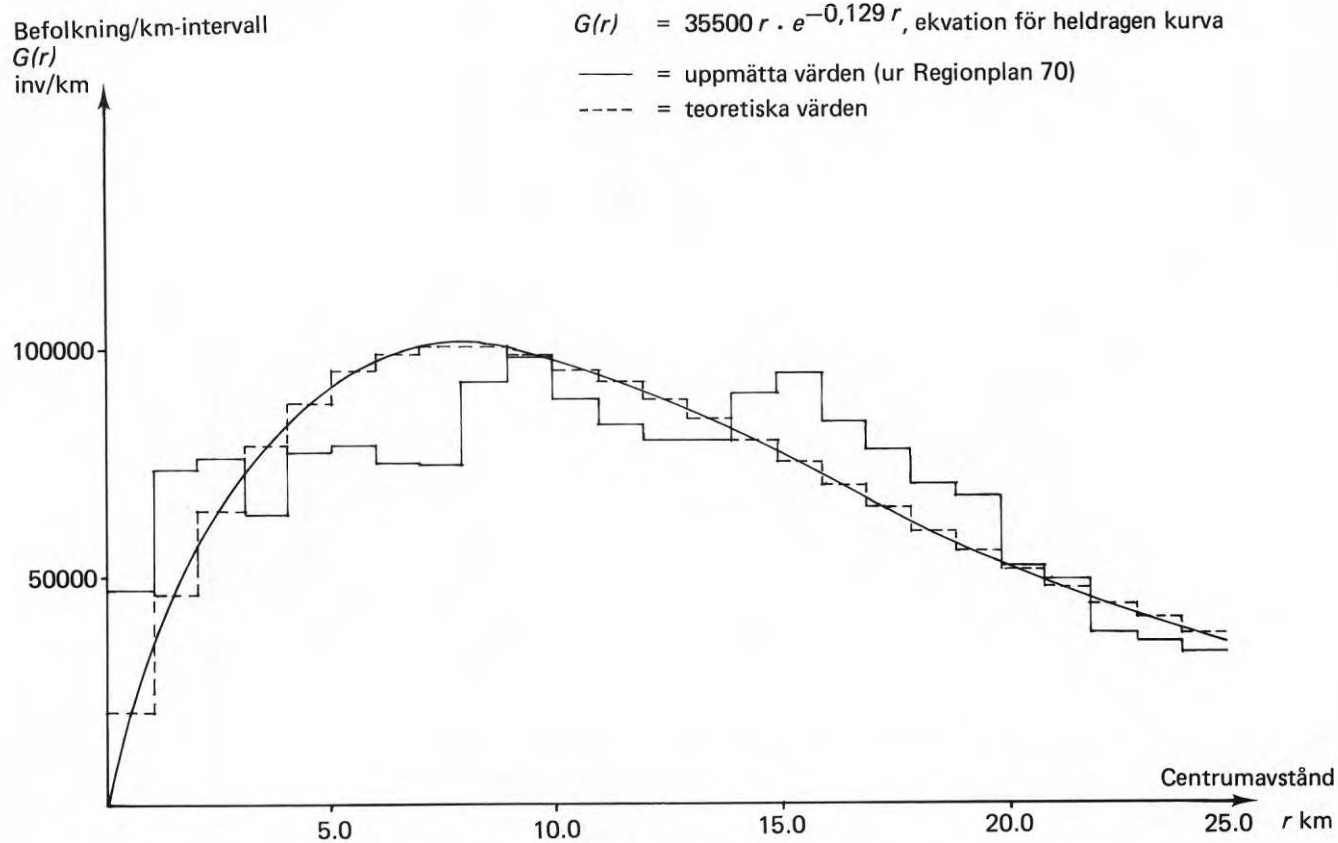


FIG. 2.7 Nattbefolkningens beräknade fördelning i Stockholm år 2000 med hänsyn till avståndet från centrum för alternativet "Utsträckt stad".

huvudvarianterna av perspektivplaner för Stockholm år 2000 som presenteras i Förslag till regionplan 1970 för Stockholmstrakten. Man ser att tidsutvecklingen av A och b sker relativt jämnt under perioden, men kurvorna döljer delvis det faktum att anpassningen till Clarks formel försämras avsevärt under perioden. Detta förhållande, som kan studeras i de följande kurvorna, beror huvudsakligen på att en av bashypoteserna för modellen satts ur spel - Stockholm kommer år 2000 att vara en trekärnig stad om regionplaneförslaget följs.

## 2.6 Markovkedjor

Teorin för Markovkedjor kan ibland vara värdefull vid studium av dynamiska förlopp. I detta kapitel ges därför en kort introduktion till detta ämne, åtföljd av några enkla exempel.

Låt  $X(t)$  vara en stokastisk process. Om processen uppfyller sannolikhets sambandet

$$P(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) = \\ = P(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1})$$

dvs om all information om tillstånden före tiden  $t_{n-1}$  är betydelselös för sannolikheten att processen skall befinna sig i tillståndet  $x_n$  vid tiden  $t_n$  betingat av att hela processens förhistoria är känd, så sägs den ha Markovegenskapen. En sådan process har alltså inget minne i den bemärkelsen att endast nuvärdet spelar någon roll för den framtida utvecklingen. Följande indelning är vanlig:

tiden	$X(t)$	
	diskret	kontinuerlig
diskret	diskret M-kedja	diskret M-process
kontinuerlig	M-kedja med kontinuerlig parameter	M-process med kontinuerlig parameter

Som ett enkelt exempel på en diskret Markovkedja betraktas en person som vill byta bostad i en stad. Denna indelas i tre områden:

1. city
2. halvcentralt
3. förort

Personen antas från början bo t ex halvcentralt. Om

$\bar{q}_i(n)$ ,  $1 \leq i \leq r$ ,  $n \geq 0$ , betecknar en sannolikhetsfördelning över  $r$  tillstånd efter  $n$  steg av processen, innebär antagandet att radvektorn  $\bar{q}(0) = (0,1,0)$ .

Bostadsvalet antas ske med vissa fasta sannolikheter enligt följande matris (kolumnerna anger flyttningsort, raderna bostadsort).

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \begin{array}{l} \text{bor i city (c)} \\ \text{bor halvcentralt (hc)} \\ \text{bor i förort (f)} \end{array}$$

c    hc    f  
(flyttar till)

dvs om personen bor t ex i centrum, är sannolikheten att bo kvar  $3/4$ , sannolikheten att flytta till halvcentralt läge  $1/4$  och sannolikheten att flytta till förort  $0$ . Elementen i matrisen utgörs alltså av övergångssannolikheter av typen  $p_{ik} = P(x(n+1)=k | x(n)=i)$ , där  $\sum_{k=1}^r p_{ik} = 1$ .  $P$  kallas en stokastisk matris.

Om lägenhetsutbudet är slumpvis fördelat, kommer sannolikheten att personen accepterat ett anbud i de olika områdena efter en omgångs sökande att bli

$$\bar{q}(1) = (0 \ 1 \ 0) \cdot \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \left( \frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{4} \right)$$

Allmänt transformeras sannolikhetsfördelningen varje omgång som  $\bar{q}(n+1) = \bar{q}(n)P$ . Om personen söker bostad i många omgångar kommer till slut  $\bar{q}(n)$  inte att förändras, utan övergå i en stationär fördelning över tillstånden. Denna kan fås som lösningen till  $\bar{a} = \bar{a}P$ , där  $\bar{a}$  är den stationära fördelningen.



I detta enkla exempel blir  $\bar{a} = (\frac{7}{11}, \frac{3}{11}, \frac{1}{11})$ , vilket betyder att chansen är 7/11 att personen bosätter sig i city, 3/11 att han bor kvar och 1/11 att han flyttar till förort.

Markov-kedjor med kontinuerlig tidsparameter är vanliga bl a i samband med befolkningsprognoser. Här måste man i stället för övergångssannolikheterna  $p_{ik}$  använda övergångsintensiteter  $a_{ik}$ . Under ett litet tidsintervall  $dt$  är sannolikheten att hoppa till ett annat tillstånd  $a_{ik} dt$  ( $i \neq k$ ). Sannolikheten för två eller flera övergångar försummas. Om sannolikhetsfördelningen över tillstånden vid en viss tidpunkt är  $q_k(t)$  kommer övergångsekvationerna att bli

$$\frac{dq_k}{dt} = \sum_{i=1}^r q_i(t) \cdot a_{ik}$$

$A = \{ a_{ik} \}$  kallas en differentiell matris och har  $\sum_{k=1}^r a_{ik} = 0$ .

I en födelse- och dödsprocess antas  $\sum_{|i-k| \geq 2} a_{ik} = 0$ , varför endast

övergångar mellan två näraliggande tillstånd blir möjliga.

En tillämpning av denna teori på en befolkningstäthetsmodell är följande. Dela in den studerade staden, som har oändlig utsträckning, i ringar numrerade 0, 1...

$X_i(t)$  = antalet boende i ring  $i$  vid tiden  $t$ .

$E \{ X_i(t) \} = m_i(t)$  = förväntat antal boende i ring  $i$ .

Staden har  $N$  innevånare vid tiden  $t=0$ , så att  $\sum_{i=0}^{\infty} m_i(0) = N$ .

För ett litet tidsintervall  $dt$  gäller

$\lambda_i dt$  = födelsesannolikhet i ring  $i$

$\mu_i dt$  = dödssannolikhet i ring  $i$

$\gamma_i dt$  = sannolikhet för inflyttning utifrån till ring  $i$

$\nu_i dt$  = sannolikhet för flyttning till ring  $i-1$  från ring  $i$

$\eta_i dt$  = sannolikhet för flyttning till ring  $i+1$  från ring  $i$

Det visar sig att

$$\frac{dm_i}{dt} = (\lambda_i - \mu_i + \gamma_i - \nu_i - \eta_i) m_i(t) + \nu_{i-1} m_{i-1}(t) + \eta_{i+1} m_{i+1}(t)$$

Följande antaganden görs för att förenkla modellen:

$$\lambda_i - \mu_i + \gamma_i = \alpha \quad \text{oberoende av } t$$

$$v_i = \frac{v}{\pi(2i+1)} \quad \eta_i = \frac{\eta}{\pi(2i+1)}$$

Det första antagandet är ett grovt antagande, som säger att det befolkningstillskott som härrör från födelser, dödsfall och inflyttning från andra områden är likformigt fördelat på ringarna. Antagandena om  $v_i$  och  $\eta_i$  innebär, att sannolikheterna att flytta inåt och utåt från en ring båda är omvänt proportionella mot ringens yta.

Det är i princip möjligt att lösa  $m_i(t)$  numeriskt direkt ur differentialekvationen ovan. Om man dock antar att  $\alpha = 0$ , dvs stadens invånarantal inte ändras, kommer  $m_i(t)$  att gå mot en stationär fördelning efter lång tid, som resultat av flyttningsmekanismerna mellan ringarna. Om  $p_i = m_i/N$  kan man visa att den stationära fördelningen blir

$$p_i = \frac{\eta_{i-1} \eta_{i-2} \cdots \eta_0}{v_i v_{i-1} \cdots v_1} \cdot p_0, \quad \sum_{i=0}^{\infty} p_i = 1$$

Insättning av antagandena ovan ger att

$$m_i = \frac{N(1-\rho)^2}{1-\rho} (2i+1) \rho^i \quad \text{där } \rho = \frac{\eta}{v}.$$

Under samma antaganden som ovan om stadens utsträckning och indelning kan man visa, att Clarks formel i maximum-likelihoodtappningen som estimat för antalet boende per ring ger

$$n_i = \frac{N(1-e^{-b})^2}{1-e^{-b}} \cdot (2i+1) e^{-bi}$$

Man ser direkt att fullständig överensstämmelse mellan de båda resultaten fås om  $\rho = e^{-b}$ .

Därmed har visats att det är möjligt att härleda Clarks formel även genom resonemang om flyttningsmekanismer i en stad. Härledningarna är lika i så måtto, att det som styr befolknings-

täthetens utseende antages vara avståndet till centrum. Uppritas det ovan gjorda antagandet i diagramform (kvoten  $\frac{n_{i-1}}{v_i}$  är den intressanta storheten) får man FIG. 2.8.

Huruvida överensstämmelsen mellan de båda resultaten är en ren tillfällighet är det svårt att uttala sig om. Dock bör erinras om att den senare härledningen bygger på ganska många för-  
enklingar och antaganden.

## 2.7 En trafikmodell, som utnyttjar Markovteknik

Då parametrarna i gravitationsmodellen är tidsberoende, och kalibrering kräver omfattande arbete, har försök gjorts att upp-  
ställa alternativa modeller.

I en sådan modell, behandlad i Horstmann (1970), beskrivs trafikflödet med hjälp av Markovteknik. Modellen bygger på antagandet, att en bilist som befinner sig i ett visst område i en stad, väljer nästa resmål efter vissa fasta sannolikheter, oberoende av tidigare val av resmål. Övergångssannolikheterna  $p_{ij}$  antages alltså kända, och om man dessutom känner totala antalet bilar i staden vid en viss tidpunkt kan trafikströmmarna bestämmas.

Bilarnas initialfördelning  $\bar{q}(0)$  på områdena antages given. Komponenten  $q_i(0)$  betyder andelen bilar i område  $i$ .  $T$  är totalantalet bilar i de  $n$  områdena. Utvecklingen av processen tänkes så, att samtliga bilar gör en resa. Härfter beräknas det antal resor som företagits mellan olika områdespar och bilar-  
nas nya fördelning över områdena bestäms. Detta förfarande tänks genomföras  $N$  gånger, där  $N$  är medelantalet resor per bil och dag.

$$n_{ij}(k) = \text{antalet resor mellan } i \text{ och } j \text{ i omgång } k$$

$$q_i(k) = \text{andelen bilar i område } i \text{ efter omgång } k$$

$$N_{ij} = \text{totalantalet resor mellan } i \text{ och } j$$

$$n_{ij}(k) = q_i(k-1) \cdot p_{ij}$$

$$q_j(k) = \sum_{i=1}^n n_{ij}(k) = \sum_{i=1}^n q_i(k-1)p_{ij}$$

$$N_{ij} = T \cdot \sum_{k=1}^N n_{ij}(k)$$

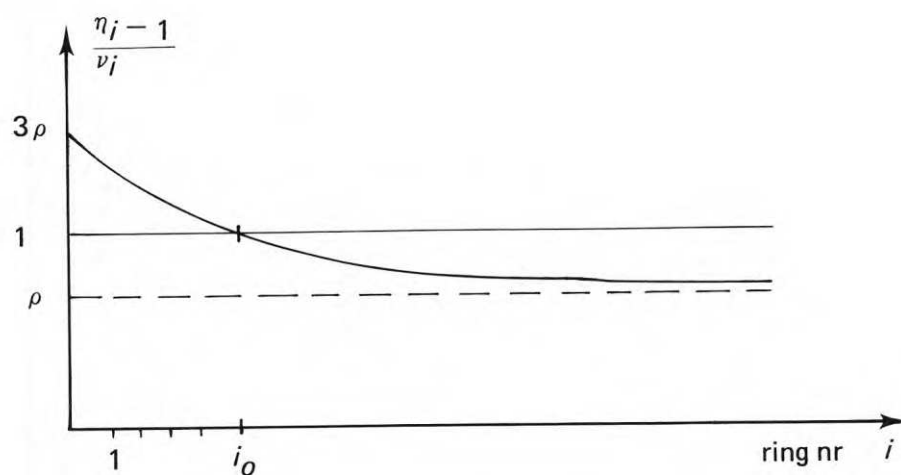


FIG. 2.8 Kvot mellan sannolikhetstätheterna för flyttning till ring  $i$  från ring  $i-1$  och för flyttning till ring  $i-1$  från ring  $i$  som funktion av avståndet till centrum (representerat av ringnumret).

Under allmänna antaganden om  $p_{ij}$  kommer  $\bar{q}(k)$  att gå mot en stationär fördelning  $\bar{q}$  efter relativt kort tid. Känner man den stationära fördelningen behöver man bara genomföra schemat ovan en gång och får direkt resultatet att  $N_{ij} = NT \cdot q_i p_{ij}$ , där  $q_i$  är den stationära fördelningen. Jämförelser mellan observerade resefrekvenser och Markovmodellen ger godtagbar överensstämmelse.

För att kunna göra en prognos för trafikflödet i framtiden krävs prognostisering av  $N$  och  $T$ , som kan fås med vanliga prognosmetoder. Matrisen  $P$  är svårare att uppskatta, speciellt om man vill göra det utan tillgång till resematriser. I artikeln föreslås en metod som använder uppskattningar av restider för att söka en uppsättning  $p_{ij}$ , som gör kvoten mellan entropifunktionen  $H = - \sum_i \sum_j q_i p_{ij} \log p_{ij}$  och medelrestiden  $\bar{t} = \sum_i \sum_j q_i p_{ij} t_{ij}$  så stor som möjligt. Man blir för att uppnå godtagbar överensstämmelse med observationerna tvingad att korrigera restiderna för korta avstånd. Då denna del av arbetet faller utanför ramen för denna framställning kommer ingen ytterligare diskussion därom att föras.

## 2.8 Simulering av lokaliseringsförlopp. En Monte-Carlo-modell för befolkningstäthet

Simuleringsmodeller utgör ett, åtminstone utomlands, vanligt hjälpmedel vid studier av lokaliseringsförlopp i stadsområden. Ytterst är denna popularitet ett uttryck för inställningen att de urbana lokaliseringsförloppen är så komplicerade att analytiska lösningar inte kan komma i fråga. I en simulering försöker man i stället, efter en teoretisk analys av de faktorer som påverkar lokaliseringsmönstret, att studera resultatet av samspelet mellan dessa genom av slumpvalsdragningar styrda datamaskinkörningar. I en äkta Monte-Carlo-modell avgörs alla beslut genom slumpningar ur skilda sannolikhetsfördelningar och för att få pålitliga resultat krävs vanligen ett stort antal upprepningar.

Ett enkelt exempel på en simulering är ett försök av Pedersen (1967) att förklara Clarks formel för befolkningstätheten med hjälp av flyttningsmekanismer liknande dem som tidigare behandlats här. Han gör följande - på studier av Köpenhamn byggda - an-

taganden om migrationerna inom en stad:

1. Flyttningsfrekvenserna avtar kvadratisk med flyttningens längd.
2. Totalantalet flyttningar från ett visst område är proportionellt mot antalet boende (alla områden lika stora).
3. Sannolikheten att flytta till ett område är proportionellt mot ytan och oberoende av befolkningstätheten i området.

Frågan som skall belysas är huruvida befolkningstätheten kommer att utveckla sig till en negativt exponentiell täthet om man tillåter bosättning utanför stadsgränsen på en tidigare sluten stad med hög täthet. Modellen utgör ett Monte-Carlospel med reglerna:

1. Staden har  $N$  innevånare, från början bosatta i de fyra mittersta rutorna. Befolkningen ändras inte.
2. Varje invånare förses med ett flyttningsfält, sådant att sannolikheterna att flytta 1, 2 respektive 3 rutor bort förhåller sig som  $9 : 4 : 1$ . Man flyttar högst 3 rutor varje gång. Varje ruta i flyttningsfältet numreras och tilldelas 1, 4 eller 9 numrerade brickor allteftersom den ligger i ring 3, 2 eller 1.
3. Spelet uppdelas i tidsperioder, så att 5% av befolkningen flyttar i varje period. Var tjugonde person märks, och flyttningsdestinationen bestäms enligt 2.

Resultatet visar, att det finns en tendens att tätheten efter många omgångar närmar sig en negativ exponentialkurva. Det intressanta med modellen är att den antyder att man kan få fram exponentialkurvan utan de antaganden om centrums attraktion som görs i andra modeller, och i stället grunda härledningen på flyttningsmekanismer av allmän karaktär.

## 2.9 Några anmärkningar om "Urban Dynamics"

Den hittills mest ambitiösa simuleringsmodellen för det urbana systemet, utvecklad av Forrester (1969) i Urban Dynamics, gör i själva verket inte skäl för beteckningen - om man med en simuleringsmodell menar att slumtalsstyrning förekommer. Däremot är

den en simuleringsmodell i den meningen, att den försöker avbilda hela det komplexa urbana systemet i laboratorieskala, och med datamaskinens hjälp studera systemets tillväxt, förfall och återupplivning, och den är alltså en simuleringsmodell i ordets reglertekniska betydelse.

I reglersystemet ingår 20 tillståndsvariabler (eller nivå-variabler) och ungefär lika många flödesvariabler. Dessutom finns ett stort antal konstanter, och det hela knyts ihop av ungefär 300 ekvationer. I systemet finns dessutom inbyggt tidsfördröjningar på upp till 20 år och ett antal positiva feedback-slingor, vilka huvudsakligen är belastade vid systemets tillväxt. Också tidsfördröjningarna förlorar i betydelse i jämviktstillståndet.

Alla variabler i området antas ha vissa börvärden, som beror på stadsområdets storlek. På grund av systemets dynamik kommer avvikelser från dessa att förekomma beroende på det aktuella tillståndet. En grundtanke är, att ekvationerna som relaterar de olika tillståndsvariablerna till varandra, i största möjliga mån bör vara multiplikativa. För detta anges två skäl. Det första är att ändringar i multiplikatorerna bör förstärka varandra, så att den samlade effekten av en liten ökning i alla de i ett samband ingående variablerna är större än summan av ökningarna. För det andra sägs det vara av vikt att om en variabel i ett samband blir nära noll skall hela uttrycket följa med. Detta stora reglersystem utgör enligt Forrester ett så komplext system att det inte kan bedömas efter samma måttstock som de enklare reglersystem, som vi är bekanta med från naturvetenskaperna. Följande sammanställning av synpunkter görs.

Komplexa system är sådana att

1. de reagerar mot intuitionen
2. är förvånansvärt okänsliga för ändringar i åtskilliga systemparametrar
3. motverkar policy-ändringar
4. innehåller viktiga omslagspunkter från vilka kraftiga ändringar strålar ut i systemet
5. motverkar och kompenserar externt pålagda korrektioner genom att reducera den internt genererade korrektionen
6. reagerar ofta på motsatt sätt gentemot en policy-ändring på kort och lång sikt
7. tenderar att vara ineffektiva.



Huvudanledningarna till dessa negativa egenskaper hos ett komplext system sägs vara att

1. det finns både positiva och negativa feedbackslingsor som kan vara multipla
2. det är fråga om ett olinjärt system. Dessa olinjariteter kan göra en feedback-slinga styrande för en stor del av systemet vid en viss tidpunkt, varefter den snabbt kan förlora sin styrande verkan.

Forrester slår slutligen ett slag för simuleringsteknik och hävdar att olinjariteter blir lätthanterliga bara man slutar att kräva analytiska lösningar till ett problem. Detta skulle medföra att uppmärksamheten fördes bort från de hopplösa uppgifterna att estimerar parametrar i sociala system och fokusera intresset på de viktigare frågorna om struktur och dynamik.

#### 2.10 Sammanfattning

Denna framställning har i huvudsak bestått av tre delar. I den första gjordes några försök att göra en översikt över prediktions- och prognosproblem i beskrivande (inklusive förklarande) modeller. Ambitionen var att försöka ge en bild av vilka metoder som användes i prognos- och estimationssammanhang och något beröra de problem som uppkommer när både rums- och tidsdimensionen bör beaktas. Problemet är relevant därför att även i planerande modeller och optimeringsmodeller - vilka måste anses mera lämpliga att använda vid studium av dynamiska förlopp - krävs att vissa variabler (kapaciteter, arbetstillfällen, befolkningssiffror, kostnader) prognostiseras med hjälp av någon av de diskuterade metoderna.

I den andra delen gavs några exempel på alternativa angreppssätt på lokaliseringsproblem över tiden, grundade på andra metoder än de klassiska regressionsanalytiska. Framställningen är i detta avsnitt främst avsedd att visa hur ett speciellt problem, nämligen det att beskriva befolkningstätheten i stort i stadsområden, kan angripas på olika sätt.

Den avslutande delen redovisar några synpunkter och exempel på simulering av lokaliseringsförlopp, med dels en simuleringsmodell för ett begränsat problem, dels en total modell av det urbana systemet sett som ett komplext icke-linjärt reglersystem. Denna modell leder till en del anmärkningsvärda resultat, vars tolkning i hög grad kan diskuteras. Man frågar sig om det är rimligt att bygga modeller av denna storleksordning, där man inte kan överblicka de förenklingar som modellbyggaren fört in i sina ekvationer och därför inte kan vara säker på, att modellen reagerar mot intuitionen därför att verkligheten gör det, eller att orsaken trots allt ligger i implicita förenklingar i modellen.

2.11 Litteraturreferenser

- Bussière, R, Snickars, F, 1970, *Derivation of the Negative Exponential Model by an Entropy Maximizing Method*. Environment and Planning, Vol. 2, p. 295-301. London.
- Forrester, J, 1969, *Urban Dynamics*. (M.I.T. Press.) Cambridge, Massachusetts.
- Förslag till regionplan 1970 för Stockholmstrakten*, 1970, Stockholms läns landsting, Regionplanekontoret.
- Horstmann, J, 1970, *Den fremtidige biltrafik mellem zonerne i et byområde bestemt uden forudgående O-D-analyse*. Dansk Vejtidskrift, nr 2, p. 21-28. Köpenhamn.
- Howard, R, 1960, *Dynamic Programming and Markov Processes*. (M.I.T. Press.) Cambridge, Massachusetts.
- Ivakhnenko, A, 1967, *Cybernetics and Forecasting Techniques*. (Elsevier.) New York.
- Johnston, J, 1963, *Econometric Methods*. (Mc Graw-Hill.) Tokyo.
- Martin-Löf, P, 1970, *Statistiska modeller*. Seminarier vid Stockholms Universitet 1969-70. Stockholm.
- Pedersen, P, 1967, *Modeller for befolkningsstruktur og befolkningsudvikling i storbyområder - specielt med henblik på Stor-københavn*. (Statens Byggeforskningsinstitut.) p. 127-129. Köpenhamn.

### 3.1 Grundläggande begrepp

Om man låter en person  $B$  - beslutsfattaren - välja bland några bestämda alternativ och observerar vilket han tar, anses detta avslöja  $B$ 's preferenser och därmed hans nytta av alternativen. Omvänt vill man med hjälp av nyttobegreppet kunna förutsäga  $B$ 's handlande i liknande valsituationer.

I det följande skall vi förutsätta att mängden  $C$  av alternativ  $c$  är stor, vilket gör det lämpligt av matematiska skäl att anta, att alternativen är kontinuerligt många. Om  $B$  har syftet med valet klart för sig, kan han konstruera en modell av valsituationen och därmed specificera en uppsättning attribut  $a_\alpha$ ,  $\alpha \in I$ , som han anser är särskilt viktiga hos alternativen. Indexmängden  $I$  innehåller minst ett element och kan ha upp till kontinuerligt många. Med andra ord gäller att modellen ger en avbildning elementvis definierad som:

$$c \longmapsto \prod_{\alpha \in I} a_\alpha$$

där pilen ( $\longmapsto$ ) skall uppfattas som en elementvis tillordning och där symbolen  $\prod$  är ett produkttecken som anger att alla attribut skall tagas med.

Vidare antas att varje attribut  $a_\alpha$  är kvantifierbart - enligt modellen - med ett reellvärt måttetal  $x(\alpha)$ . Elementvis förutsätter detta en avbildning:

$$a_\alpha \longmapsto x(\alpha) \in R,$$

där  $R$  betecknar de reella talen.

Sammansättning av avbildningarna ger:

$$\begin{array}{ccc} & \prod_{\alpha \in I} a_\alpha & \\ & \nearrow & \searrow \\ c & \dashrightarrow & x_c = \prod_{\alpha \in I} x(\alpha) \end{array}$$

<sup>1</sup> En funktions argument är en funktion och motsvarande värde är ett reellt tal.

Modellen anger således för varje alternativ  $c$  en reellvärd funktion  $x_c$ .

$$x_c : I \ni \alpha \longmapsto x_c(\alpha) \in \mathbb{R}$$

där med  $I \ni \alpha$  menas samma sak som  $\alpha \in I$ . Funktionen  $x_c$  tillordnar numret hos ett attribut det numeriska värdet  $x_c(\alpha)$ .

Mot alternativmängden  $C$  svarar sålunda funktionsmängden

$$X = \{ x_c \mid c \in C \}$$

Låt oss förutsätta att beslutsfattaren  $B$ 's valmekanism kan beskrivas av en reellvärd nyttofunktional  $u$ .

$$u : X \longrightarrow \mathbb{R}$$

vilket innebär, att varje alternativ  $c$  tillordnas ett reellt tal  $u(x_c)$  som anger  $B$ 's nytta. En pil ( $\rightarrow$ ) anger en inte närmare specificerad tillordning mellan två mängder.

Det kan mycket väl hända att flera nyttofunktionaler ger samma beskrivning av  $B$ 's val i enlighet med vårt förutsäggande syfte. Av detta skäl bildar vi en lämplig klass  $U$  av nyttofunktionaler, i vilken vi senare skall införa ett antal ekvivalensrelationer  $S$  - eller skalor. Dessa kan representeras med en motsvarande mängd transformationer  $F$  med element  $f$ .

$$f : U \longrightarrow U$$

under vilka ekvivalenta nyttofunktionaler överförs i varandra.

Sammanfattningsvis visas ett diagram i vilket de viktigaste begreppen ingår:



$(1-\lambda)(x_2, s_2)$ ,  $0 \leq \lambda \leq 1$  det. I detta fall krävs uppenbarligen kontinuerligt många alternativ svarande mot  $\lambda x_1 + (1-\lambda)x_2$ ,  $0 \leq \lambda \leq 1$ . Komplementet till en konvex mängd är konkav.

Ett vanligt antagande om nyttofunktionaler är att de är konkava.

### Exemplifieringar

I exemplen 1 och 2 nedan innehåller indexmängden  $I$  endast ett element  $\alpha_0$ . En naturlig inbäddning av funktionsmängden  $X$  i  $R$  fås genom identifiering av avbildningen  $x$ .

$$x : \alpha_0 \longrightarrow x(\alpha_0) \in R$$

med dess bild  $x(\alpha_0)$ . Geometriskt representeras alternativet  $c$  via funktionen  $x_c$  med en punkt  $x_c(\alpha_0)$  på reella talaxeln.

1. Antag att  $u$  är 2 ggr kontinuerligt deriverbar.

Nyttans omättlighet uttrycks i detta fall med att marginalnyttan  $\frac{du(x)}{dx} > 0$ . Om dessutom  $\frac{d^2u(x)}{dx^2} < 0$ , dvs fallet att marginalnyttan är avtagande, är  $u$  konkav.

2. Antag att  $B$  är en kommun, syftet med väljandet är att långtidsplanera bostadssektorn, endast attributet markreserv är relevant och kommunens nytta är en långsiktig valfrihet i markdispositionen. Det är klart att omättlighet gäller inom ramen för given markareal, eftersom större markreserv ger större valfrihet på lång sikt. Antag att kommunen  $B$  ställs inför ett lotteri, där markreserven  $x_1$  har sannolikheten  $\lambda$ , och markreserven  $x_2 > x_1$  har sannolikheten  $1-\lambda$ , där  $0 \leq \lambda \leq 1$ .

Det är svårt att tänka sig en kommun, som skulle föredra lotteriet framför det säkra alternativet  $x_3 = \lambda x_1 + (1-\lambda)x_2$ , vilket medför att nyttans epi-graf är konkav, och därmed att nyttofunktionalen är konkav. I annat fall skulle  $B$  vara villig att betala för spänningsmomentet vid lottningen, se Raiffa (1968) p. 94-96, för en närmare diskussion.

Man kan se lotteriförfarandet som ett praktiskt test på konvexitet.





3. Låt alternativet  $c$  vara ett samhälles utveckling under en femårsperiod. Attributet  $a_\alpha$  kan väljas som det ekonomiska tillståndet vid tidpunkten  $\alpha$ ,  $0 \leq \alpha \leq 5$  år. Mätetalet för  $a_\alpha$  definieras som konsumtionsintensiteten  $x(\alpha)$  i kr/sek vid tidpunkten  $\alpha$ . Mot varje alternativ  $c$  svarar således en tidsutveckling av konsumtionsintensiteten  $x$ . Nyttofunktionalen  $u$  avbildar tidsutvecklingen  $x$  på välfärden  $u(x)$ ; i detta fall kan vi välja

$$u(x) = \int_{\alpha \in I} \log x(\alpha) d\alpha = \int_0^5 \log x(\alpha) d\alpha.$$

Alternativt kan tidpunkten  $\alpha$  beteckna år nr  $\alpha$  under femårsperioden. Indexmängden  $I$  innehåller i så fall 5 element.

Vi kan tolka  $x(\alpha)$  som konsumtion i kr under år nr  $\alpha$ .

Ett möjligt mått på välfärd är  $u(x) = \min_{\alpha \in I} x(\alpha)$ .

### Skalor

I mängden  $U$  av nyttofunktionaler inför vi nu en ekvivalensrelation  $S$  med hjälp av en mängd  $F$  av transformationer som överför nyttofunktionaler i varandra.  $u$  är delement i  $U$ . Axiomen för ekvivalensrelationer ger följande krav på  $F$ :

- 1) Reflexivitet :  $u S u$  gäller för alla  $u \in U$ , eller annorlunda uttryckt: Varje nyttofunktional  $u$  skall stå i relationen  $S$  till sig själv. Det finns således en identisk transformation,  $i \in F$ , sådan att sammansättningen  $i \circ u$  där  $\circ$  är en symbol som betecknar sammansättning är definierad och lika med  $u$  för alla  $u \in U$ .

Det överlåtes åt läsaren att bevisa, att  $i$  är den enda identiska transformationen i  $F$ .

- 2) Transitivitet : för alla  $u, v, w \in U$  medför  $u S v$  &  $v S w$  att  $u S w$  gäller.

Om  $f, g \in F$  och  $v = f \circ u$  &  $w = g \circ v$  gäller fås genom upprepade sammansättningar  $w = g \circ v = g \circ (f \circ u) = (g \circ f) \circ u$ . Transitiviteten kräver att  $g \circ f \in F$ .

3) Symmetri : för alla  $u, v \in U$  medför  $u S v$  att  $v S u$  gäller.

Om  $f \in F$  och  $v = f \circ u$ , så existerar en transformation  $h \in F$  sådan att  $u = h \circ v$ .

Härav följer att  $v = f \circ u = f \circ (h \circ v) = (f \circ h) \circ v$  och analogt  $u = (h \circ f) \circ u$ .

Således gäller  $f \circ h = h \circ f = i$ , eftersom  $i$  är unik.

Anmärkning: I allmänhet gäller inte att  $f \circ h = h \circ f$ .

För att kunna åskådliggöra transformationerna grafiskt gör vi följande konstruktion: Nyttofunktionalen  $u$ 's bild av funktionsmängden  $X$  skrives  $u(X)$ , definieras som  $u(X) = \{ u(x) \mid x \in X \}$ , och är en delmängd till  $R$ . Mot varje alternativ  $c$  svarar en punkt  $u(x_c)$  på reella talaxeln.

När två nyttofunktionaler  $u$  och  $v$  är givna, läggs deras bilder av  $X$  vinkelrätt mot varandra längs var sin reella talaxel. Transformationen  $f$ :

$$f : u \longmapsto v ; u, v \in U$$

representeras således av avbildningen  $\varphi$  :

$$\varphi : R \ni u(x) \longmapsto v(x) \in R.$$

Eftersom  $\varphi$  är en reellvärd funktion av en reell variabel, kan den lätt åskådliggöras med hjälp av en graf  $G_f = \{ (u(x), v(x)) \mid x \in X \}$  i det sålunda uppkomna reella talplanet.

Nedan följer definitioner på några använda skalor såsom absolut-, kvot-, intervall- och ordinalskala. Medtagna alternativ kommer för enkelhets skull att numreras i stället för att namnges med  $c, d$ , osv. Till alternativ 1 hör således funktionen  $x_1$ , vars värde är  $u(x_1)$ .

Absolutskalan tillåter ingen frihet, endast den identiska avbildningen  $i \in F$  duger. Enligt FIG. 3.3 kan detta åskådliggöras av att  $u(x)$  avbildas på  $v(x)$  via en  $45^\circ$  rät linje

$$G_a = \{ (u(x), u(x)) \mid x \in X \}$$

genom origo. Här gäller tydligen att  $v(x_1) = u(x_1), v(x_2) = u(x_2)$  och  $v(x_3) = u(x_3)$ . Ett exempel är penningmängd mätt i kr.

Kvotskalan är invariant under multiplikation med positiva reella tal  $a$ . Avbildningen sker via den räta linjen

$$G_k = \{ (u(x), a \cdot u(x)) \mid x \in X \text{ \& } a > 0 \} ; \text{ se FIG. 3.4.}$$

Härav inses, att  $\frac{v(x_1)}{v(x_2)} = \frac{a u(x_1)}{a u(x_2)} = \frac{u(x_1)}{u(x_2)}$ ; dvs att kvoten är oförändrad. Ett exempel är penningmängd. Vid övergång från kr till öre är  $a = 100$ .

Intervallskalan håller kvoten mellan intervalllängder konstant. Avbildningen sker via den räta linjen

$$G_i = \{ (u(x), a \cdot u(x) + b) \mid x \in X \text{ \& } a > 0 \} . \text{ se FIG. 3.5.}$$

Härav följer att:

$$\begin{aligned} \frac{v(x_3) - v(x_2)}{v(x_2) - v(x_1)} &= \frac{a u(x_3) + b - (a u(x_2) + b)}{a u(x_2) + b - (a u(x_1) + b)} = \frac{a u(x_3) - a u(x_2)}{a u(x_2) - a u(x_1)} = \\ &= \frac{u(x_3) - u(x_2)}{u(x_2) - u(x_1)} \end{aligned}$$

Skalan är bestämd så när som på en positiv linjär avbildning (se definition av  $G_i$  i FIG. 1.5). Som exempel kan tas subjektiv ljudstyrka. Vi upplever kontraster i ljudnivån men förnimmer ingen absolut nollnivå. Buller mäts i dB, dvs i absolut skala, vilket medför att sådana mätningar knappast är det relevanta måttet på personligt obehag.

För en enskild beslutsfattare antages i litteraturen vanligen att nyttan kan mätas i intervallskala.

Ordinalskalan (rangordning) är invariant under strikt växande avbildningar, se FIG. 3.6. Således medför

$$u(x_3) < u(x_2) < u(x_1) \text{ att}$$

$$v(x_3) < v(x_2) < v(x_1) \text{ gäller.}$$

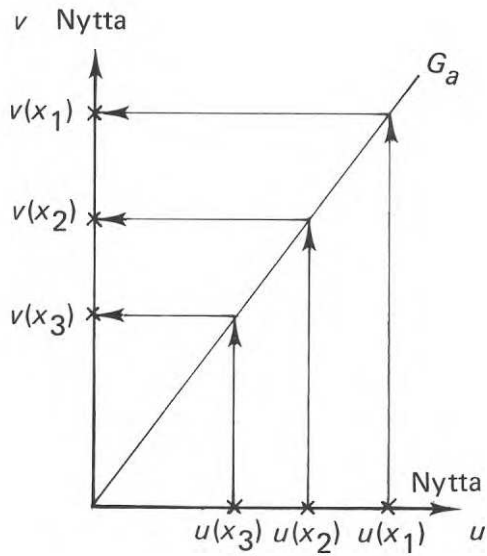


FIG. 3.3 Absolutskala, invarians under identisk avbildning.

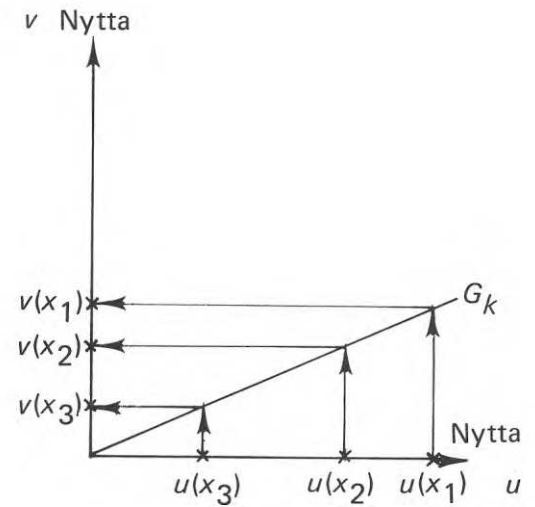


FIG. 3.4 Kvotskala, invarians under positiva linjära avbildningar med origo som fixpunkt.

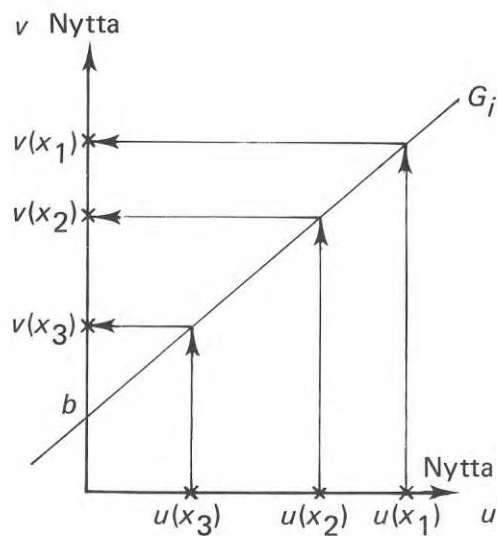


FIG. 3.5 Intervallskala  $H$ , invarians under positiva linjära avbildningar.

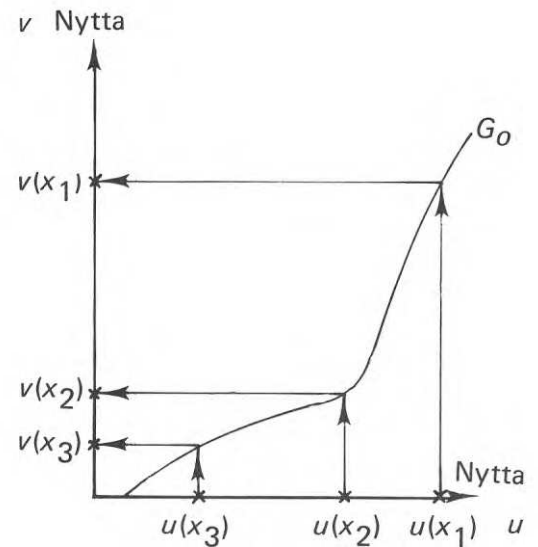


FIG. 3.6 Ordinalskala, invarians under strikt växande avbildningar.

Grafen  $G_0$  är

$\{ (u(x), \varphi(u(x))) \mid x \in X \}$ , där  $\varphi$  är en avbildning.

$\varphi : R \ni z \longrightarrow \varphi(z) \in R$ , sådan att  $z_1 < z_2$  medför  $\varphi(z_1) < \varphi(z_2)$ .

#### ANMÄRKNING

Om funktionen  $\varphi$  i exemplet ovan vore konkav, t ex  $\varphi(z) = \sqrt{z}$ ,  $z > 0$ , skulle dess omvändning  $\varphi^{-1}(z) = z^2$  vara konvex. Det går således inte att konstruera en skala byggd på invarians med avseende på konkava funktioner, eftersom symmetriaxiomet för ekvivalensrelation inte är uppfyllt.

Steget mellan intervall- och ordinalskala kan tyckas långt. Man kan fråga sig vilken skala som uppnås, om man i stället för kvoten vill bevara rangordningen mellan längderna på intervall. Figuren som förklarar ordinalskalan (FIG. 3.6) kan användas på följande sätt: Grafen får inte knycka av som  $G_0$  gör, ty då ändras rangordningen mellan intervallen. Således är en ändring av grafens lutning inte tillåten. Dessutom måste den vara sammanhängande, se FIG. 3.7. Följaktligen är grafen nödvändigtvis en rät linje och skalan en intervallskala.

Intervallskalan karakteriseras av att rangordningen mellan intervalllängderna bevaras. Som en konsekvens härav bibehålles kvoterna mellan intervalllängderna.

### 3.2 Numerisk bestämning av nyttofunktionalen $u$ när antalet attribut är ändligt

Indexmängden  $I$  numrerar attributen  $a_\alpha$  och har enligt rubriken ändligt många element. Vid exemplifieringarna i det första avsnittet (sid. 49) visades hur man i fallet med endast ett element i  $I$  kunde naturligt inbädda funktionsmängden  $X$  i  $R$ . Samma metod ger i fallet  $n$  element i  $I$  en naturlig inbäddning av  $X$  i  $R^n$ .

Nyttofunktionalen  $u$  kan på detta sätt representeras som en reellvärd funktion av  $n$  reella variabler, vilket genomgående blir fallet i detta avsnitt. Symbolerna  $x$  och  $y$  kommer här att beteckna reella variabler.

I den följande framställningen antages att nyttan mätes i intervallskala, dvs att mängden  $F$  endast består av sådana transformationer  $f$ , vars representationer  $\varphi$ .

$$\varphi : R \longrightarrow R$$

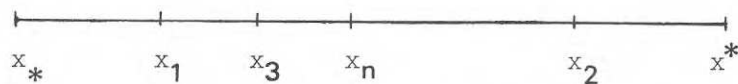
är positiva linjära avbildningar; dvs  $\varphi(z) = az + b$ , där  $a > 0$ . Om man numeriskt vill bestämma  $u(x)$  för en beslutsfattare i en viss situation är det praktiskt att utnyttja ett lotteriförfarande vid utfrågningen. Först bestäms de gränser  $x_*$  och  $x^*$  mellan vilka  $x$  antages variera i den aktuella situationen.

I de praktiska fall som behandlas i Keeney (1969) handlar det alltid om repetitiva situationer såsom avvägning mellan olika abonnenter i en telefonväxel eller daglig lagerbeställning till en blodbank.

Den subjektivistiska skolan efter Savage med Raiffa i spetsen hävdar, att även engångsbeslut kan göras på ett konsistent sätt med hjälp av lotteriförfarande. Om beslutsfattaren har obegränsad rätt att spela på lotteriet, borde han spela på ett medelvärdesriktigt sätt, såvida han inte njuter av spelsituationen.

Risk och bestämning av  $u(x)$ . (Keeney (1969) Ch. 5).

Beslutsfattaren ställs inför ett val mellan lotteriet  $x_*$  och  $x^*$  med sannolikhet  $\frac{1}{2}$  för vardera eller kortfattat  $\left[ x_* ; \frac{1}{2} ; x^* \right]$  - och det säkra utfallet  $x_1$ . Vi väntar oss att han föredrar lotteriet, vilket han också gör.



Därefter görs samma fråga men med det säkra utfallet  $x_2$ .

Antag att beslutsfattaren följer våra förväntningar och föredrar  $x_2$ . Lotteriet  $\left[ x_* ; \frac{1}{2} ; x^* \right]$  ställs mot det säkra utfallet  $x_3$ , där vi inte längre har en bestämd uppfattning om svaret. Efter  $n$  utfrågningar har följderna  $(x_i)$  uppnått punkten  $x_n$ , som upplevs som indifferent med lotteriet. Om  $x_n < \frac{1}{2}(x_* + x^*)$  säges beslutsfattaren ha riskaversion, vilket är liktydigt med att  $u$  är konkav.

Vill man bestämma formen på  $u$  närmare, kan det vara lämpligt att dela in sträckan  $\left[ x_* , x^* \right]$  i 10 lika stora intervall.



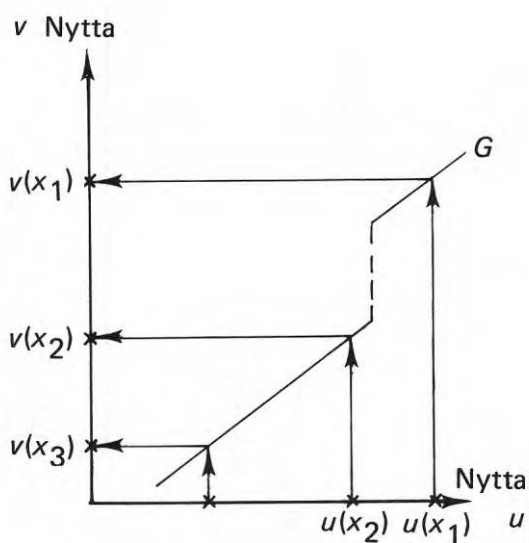


FIG. 3.7 Grafen  $G$  har konstant lutning. Diskontinuiteten medför att rangordning mellan intervallen ej bevaras.

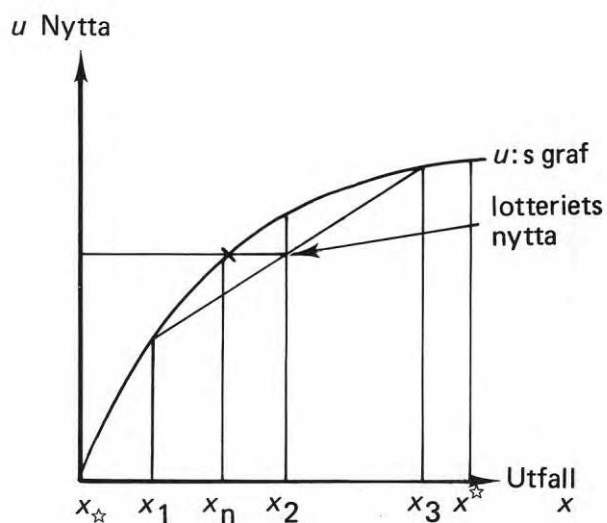


FIG. 3.8 Lotteri:  $[x_1; 1/2; x_3]$ . Mittpunkten i intervallet är  $x_2$ . Det säkra utfallet  $x_n$  har samma nytta som lotteriet. Riskaversion eftersom  $x_n < x_2$ .

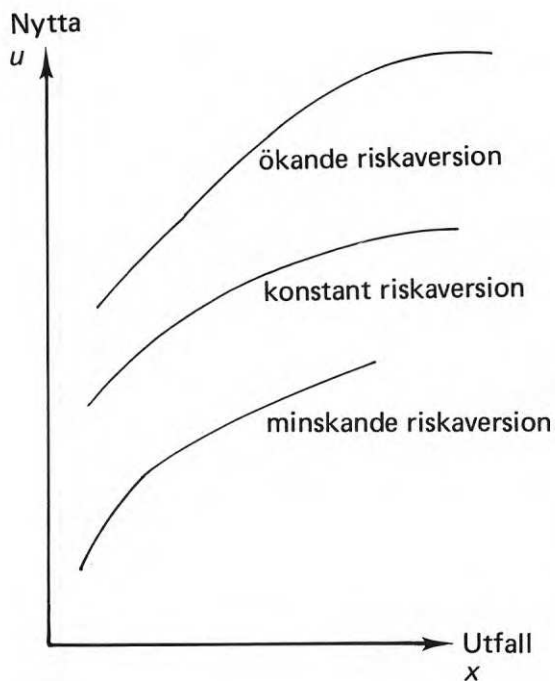


FIG. 3.9 Nyttofunktioner vid olika riskaversion.

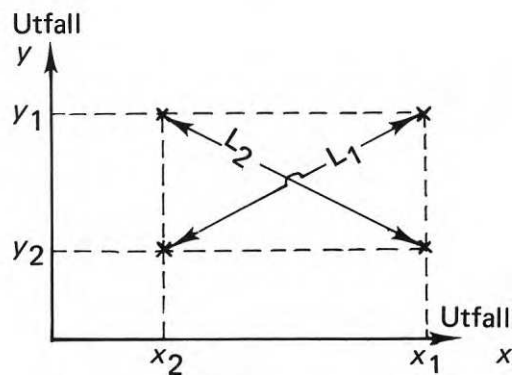


FIG. 3.10 Grafisk representation av två lotterier ( $L_1, L_2$ ) för undersökning av additivitet.

För varje par närliggande intervall testas om ett lotteri med sannolikheten  $\frac{1}{2}$  för de yttersta ändpunkterna i intervallparet föredras framför deras mittpunkt. Dessutom kan för varje intervallpar den ovan beskrivna tekniken ge den till lotteriet hörande indifferenspunkten. Om mittpunkten föredrages framför lotteriet för alla intervallpar, har beslutsfattaren visat riskaversion i hela intervallet, se FIG. 3.8.

Dessutom kan man bilda skillnaden mellan mittpunkten och indifferenspunkten för alla intervallpar. Om skillnaderna minskar, är konstanta respektive ökar med växande  $x$ , uppvisar beslutsfattaren minskande, konstant respektive ökande riskaversion.

Pratt's riskaversionfunktion  $r$  definieras som

$$r : x \longrightarrow r(x) = - \frac{\frac{d^2 u(x)}{dx^2}}{\frac{du(x)}{dx}}$$

och mäter graden av riskaversion i överensstämmelse med det ovan sagda, se Pratt (1964). Lutningen hos  $u$ 's graf i punkten  $x$  anges av  $\frac{du(x)}{dx}$ . Andra derivatan  $\frac{d^2 u(x)}{dx^2} = \frac{d}{dx} \left( \frac{du(x)}{dx} \right)$  visar lutningens förändring. Den relativa lutningsändringen uttrycks således som  $\frac{\frac{d^2 u(x)}{dx^2}}{\frac{du(x)}{dx}}$ . Riskaversionen  $r(x)$  anger följaktligen den relativa lutningsminskningen hos  $u$ 's graf i punkten  $x$ , se FIG. 3.9.

Efter denna förberedelse är vi mogna att gå över till det fall, där nyttan är en funktion av två variabler. Generaliseringen till  $n$  variabler är förhållandevis enkel.

#### Nyttooberoende (Keeney, 1969, Ch. 2)

Man säger att  $x$  är nyttooberoende av  $y$  om beslutsfattarens relativa preferenser för olika  $x$ , när  $y$  hålls konstant  $= y_0$ , är oberoende av  $y_0$ 's storlek. Nyttofunktionens form är  $m$  a o oberoende av värdet på  $y_0$ . Detta betyder att den är bestämd på en positiv linjär avbildning när, dvs att  $u(x,y) = a(y) + b(y)u(x,y_0)$ . Det räcker i detta fall att bestämma tre funktioner av en variabel.

Ett alternativt sätt att uttrycka nyttoberoende är följande:

Låt oss uppfatta variabeln  $y$  som en indexering av en familj (eller mängd) nyttofunktioner  $\{u_y\}$  av en variabel

$u_y : x \longrightarrow u_y(x) = u(x, y)$  sådana att för alla  $y$ ,  $u_y$  är en strikt växande funktion av  $x$ .

DEFINITION:

$x$  är nyttoberoende av  $y$  om alla medlemmarna i familjen  $\{u_y\}$  är ekvivalenta, mätt i intervallskala. Om det dessutom gäller att  $y$  är nyttoberoende av  $x$ , dvs att  $x$  och  $y$  är ömsesidigt nyttoberoende, kan man visa att  $u(x, y) = v(x) + w(y) + k \cdot v(x) \cdot w(y)$ . Man säger att  $u$  är kvasi-additiv.

Vad gäller om  $k$ ? Inför två lotterier  $L_1$  och  $L_2$  (se FIG. 3.10) sådana att

$$L_1 = \left[ (x_2, y_2) ; \frac{1}{2} ; (x_1, y_1) \right] \text{ och}$$

$$L_2 = \left[ (x_2, y_1) ; \frac{1}{2} ; (x_1, y_2) \right].$$

Antag att nyttorna växer med växande  $x$  respektive  $y$ . Man kan visa att  $L_1$  föredrages framför  $L_2$  omm  $k > 0$ ,  $L_1$  är indifferent med  $L_2$  omm  $k = 0$ , och  $L_2$  föredrages framför  $L_1$ , omm  $k < 0$ .

Antag att  $x$  är konsumtionen under en tidsperiod och  $y$  konsumtionen under en senare tidsperiod. Vad skulle normala tidspreferenser ge? Jämförelsen mellan lotterierna  $L_1$  och  $L_2$  kan delas upp i två steg:

i. Slumpdel.

Bestäm ett med  $L_1$  indifferent säkert utfall  $(x', y')$

och för  $L_2$  motsvarande utfall  $(x'', y'')$ .

Tekniskt har jämförelsen återförts på två säkra alternativ.

ii. Tidsdel.

Låt den nuvarande och framtida tidpunkten ingå i samma planeringsperiod. Eftersom konsumtionsnivåerna  $x$  och  $y$  är utifrån fastlagda enligt två alternativ, kommer valet av den ena - alternativet "båda lika bra" kan även förekomma - att avslöja tidspreferensen hos beslutsfattaren. I detta fall behövdes en ytterst begränsad information om nyttofunktionens utseende.

Ur gjorda intervjuundersökningar med konkreta exempel framgår att ingen person har hävdats att  $L_1$  är indifferent med  $L_2$ . Detta antyder att nyttor över tiden inte låter sig adderas.

Riskoberoende (Keeney (1969) Ch. 4)

Ett annat definitionssätt är:

$x$  är riskoberoende av  $y$  om

$$r_x(x,y) = - \frac{\frac{\partial^2 u(x,y)}{\partial x^2}}{\frac{\partial u(x,y)}{\partial x}}$$

är oberoende av  $y$ . Man kan visa, att begreppen nytto- och risk-oberoende är ekvivalenta för kontinuerliga nyttofunktioner.

Additivitet och homogenitet (Chakravarty (1969) p. 24)

Låt oss återvända till fallet att  $u$  är kvasi-additiv.

Om  $k = 0$  är  $u$  additiv:  $u(x,y) = v(x) + w(y)$ .

Om  $k \neq 0$  kan man addera  $1/k$  till uttrycket för  $u$ :

$$\begin{aligned} u(x,y) &= v(x) + w(y) + k \cdot v(x) \cdot w(y) + 1/k = \\ &= \frac{1}{k} (k \cdot v(x) + 1)(k \cdot w(y) + 1) = \underline{v(x) \cdot w(y)}. \end{aligned}$$

eftersom även  $v$  och  $w$  är bestämda på en positiv linjär avbildning när.

Man kan fråga sig under vilka avbildningar

$$f : u \longrightarrow f(u)$$

$u$  bevarar sin additiva form  $u(x,y) = v(x) + w(y)$ . Additiviteten

medför att  $\frac{\partial^2 u}{\partial x \cdot \partial y} = 0$ , vilket även krävs för

$$f : \frac{\partial^2 f}{\partial x \partial y} = 0. \text{ Omskrivning ger}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial y} \frac{df}{du} \right) = \frac{\partial^2 u}{\partial x \partial y} \frac{df}{du} + \frac{\partial u}{\partial y} \frac{\partial u}{\partial x} \frac{d}{du} \frac{df}{du} = \\ &= \frac{\partial^2 u}{\partial x \partial y} \frac{df}{du} + \frac{\partial u}{\partial y} \frac{\partial u}{\partial x} \frac{d^2 f}{du^2}. \end{aligned}$$

Eftersom båda blandade derivatorna = 0 och  $\frac{\partial u}{\partial x}$  och  $\frac{\partial u}{\partial y}$  antages  $> 0$  fås  $\frac{d^2 f}{du^2} = 0$ , vilket leder till  $f(u) = au + b$ , där  $a > 0$ ,

om samma rangordning mellan alternativen skall bibehållas. I det multiplikativa fallet ersättes  $u$  och  $f$  med  $\log u$  respektive  $\log f$  och resultatet blir  $f(u) = b \cdot u^a$ .

Nedanstående avsnitt om homogenitet finns i Chakravarty (1969) p. 26-27, men med tryckfel och en underlig användning av kedjeregeln.

Låt  $x$  och  $y$  vara konsumtionen vid två tidpunkter. Antag att  $u(x,y) = v(x) + w(y)$  är homogen av graden  $k$ , dvs att  $u(\lambda x, \lambda y) = \lambda^k u(x,y)$  för  $\lambda \geq 0$ . Vad medför detta antagande? Identiteten  $\frac{dv(x)}{dx} / \frac{dw(y)}{dy} = \frac{dv(\lambda x)}{dx} / \frac{dw(\lambda y)}{dy}$  gäller p g a homogeniteten. En omfördelning ger

$$\frac{\frac{dv(\lambda x)}{dx}}{\frac{dv(x)}{dx}} = \frac{\frac{dw(\lambda y)}{dy}}{\frac{dw(y)}{dy}}$$

Derivera båda sidorna m a p  $\lambda$ . Studera speciellt  $\frac{d}{dx} v(\lambda x)$  :

$$\begin{aligned} \frac{d}{d\lambda} \frac{d}{dx} v(\lambda x) &= \frac{d}{dx} \frac{d}{d\lambda} v(\lambda x) = \frac{d}{dx} \frac{d(\lambda x)}{d\lambda} \frac{dv(\lambda x)}{d(\lambda x)} = \\ &= \frac{d}{dx} \left[ x \cdot \frac{dv(\lambda x)}{d(\lambda x)} \right] = \frac{dv(\lambda x)}{d(\lambda x)} + x \frac{d(\lambda x)}{dx} \frac{d^2 v(\lambda x)}{d(\lambda x)^2} = \\ &= \frac{dv(\lambda x)}{d(\lambda x)} + \lambda x \frac{d^2 v(\lambda x)}{d(\lambda x)^2} . \end{aligned}$$

Sätt speciellt  $\lambda = 1$ , vilket ger

$$1 + x \frac{\frac{d^2 v}{dx^2}}{\frac{dv}{dx}} = 1 + y \frac{\frac{d^2 w}{dy^2}}{\frac{dw}{dy}} \quad \text{eller} \quad x \frac{d}{dx} \log \frac{dv}{dx} = y \frac{d}{dy} \log \frac{dw}{dy}$$

Jämför med definitionen av Pratt's riskaversionsfunktion! Likheten utsäger, att elasticiteten i marginalnyttan är densamma vid båda tidpunkterna. Således fås

$$x \frac{d}{dx} \log \frac{dv}{dx} = -q \quad \log \frac{dv}{dx} = -q \log x + \log a$$

$$\frac{dv}{dx} = a x^{-q} \quad v(x) = v(x) = \frac{a}{1-q} x^{1-q} + b \quad \text{eller}$$

$v(x) = \frac{1}{1-q} x^{1-q}$ . Ekvationen visas i FIG. 3.11,  $v$  är där bestämd på en positiv linjär transformation när.

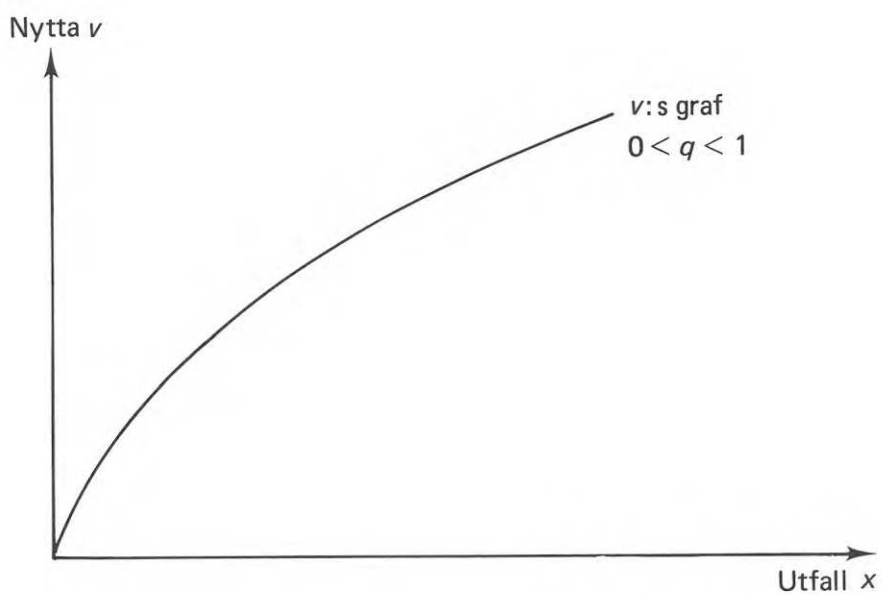


FIG. 3.11 Additiv och homogen nyttofunktion.

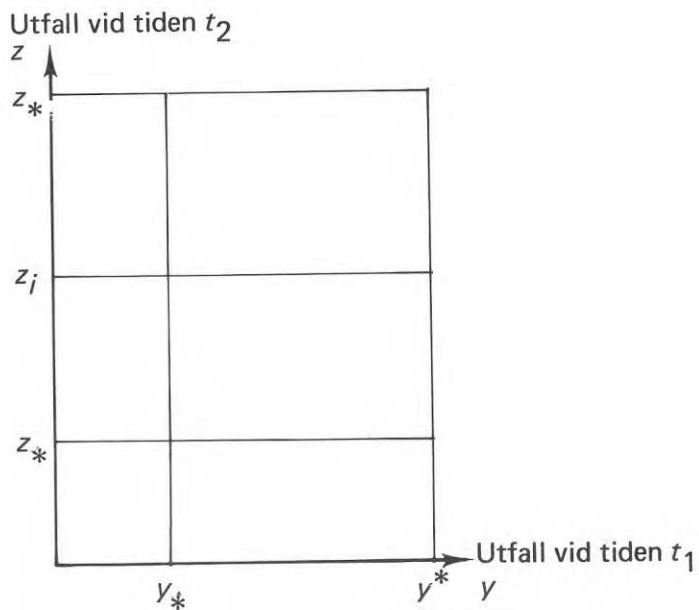


FIG. 3.12 Lotterier för bestämning av nyttofunktionalen, då alternativen beskriver utvecklingar över tiden.

### 3.3 Kvalitativ bestämning av nyttofunktionalen $u$ , när alternativen beskriver utvecklingar över tiden

Alternativ som beskriver tidsutvecklingar har kontinuerligt många attribut. Om endast en egenskaps tidsutveckling är intressant medför detta att indexmängden  $I$  då antingen är  $\mathbb{R}$  vid oändlig tidshorisont eller intervallet  $[0, T] \subset \mathbb{R}$  vid ändlig horisont. När egenskapen är uttryckt för ett lager eller bestånd som t ex konsumtion, är oändlig tidshorisont det naturliga valet. Emellertid kan utvecklingen under tiden  $(T, \infty)$  ersättas med ett randvillkor för tidpunkten  $T$ , angivande beståndet vid denna tidpunkt.

Antag att egenskapen ifråga är konsumtionen i ett land och att beslutsfattaren är dess finansminister. Välj indexmängden  $I$  till  $[0, T] \subset \mathbb{R}$  och  $x$  som konsumtionen

$$x : I \longrightarrow \mathbb{R}.$$

Vilken form bör funktionalen  $u$  anta för att på ett rimligt sätt spegla beslutsfattarens preferenser?

#### Möjlig form på funktionalen $u$

Låt oss återvända till fallet två attribut, där  $u$  är en reellvärd funktion av två variabler  $y$  och  $z$ , som anger konsumtionsintensiteten vid två tidpunkter  $t_1$  och  $t_2$ , där  $t_1 < t_2$ . En empirisk bestämning av  $u$  skulle kunna gå till på följande sätt:

För båda tidpunkterna  $t_1$  och  $t_2$  bestäms undre och övre gräns för konsumtionsintensiteten:  $y_*$  och  $y^*$  respektive  $z_*$  och  $z^*$ . Eftersom vi bara studerar mycket små intervall kring två tidpunkter i intervallet  $[0, T]$  behövs ingen budgetrestriktion i vårt fall; den lägger vi endast på hela intervallet. Vi bör alltså mäta upp hur beslutsfattarens nytta varierar i rektangeln i FIG. 3.12. För ett lämpligt antal nivåer  $z_*, \dots, z_i, \dots, z^*$  på  $z$  bestäms indifferenspunkten  $y_i^0$  till lotteriet  $\left[ y_*; \frac{1}{2}; y^* \right]$ . Om alla  $y_i^0$  sammanfaller, har vi visat att  $y$  är nyttoberoende av  $z$ . Man kan vänta sig att detta är fallet, eftersom nyttan av en framtida konsumtionsintensitet inte borde påverka den nuvarande. Som spekulation kan vi anta att även  $z$  är nyttoberoende av  $y$ , vilket borde testas i ett verkligt fall.



Ett kausalt argument mot det sista antagandet är, att vad vi nu gör påverkar förhållandena vid senare tidpunkter, medan sambandet orsak-verkan hypotetiskt endast går bakåt i tiden för vissa antipartiklar. Om ovan nämnda påverkans effekt är förhållandevis obetydlig, kan vi emellertid rättfärdiga spekulatjonen. Enligt vad som framkommit under avsnittet "additivitet och homogenitet" bör hopvägningen av nyttorna för tidpunkterna  $t_1$  och  $t_2$  ske multiplikativt, förutsatt att  $y$  och  $z$  är nyttooberoende, vilket innebär att  $u(y,z) = v(y) w(z)$ .

Antag att  $\log v(y)$  och  $\log w(y)$  är ekvivalenta enligt intervallskalan vilket innebär att dessa nyttofunktioner inte bör ändra form under planeringsperioden. Härav följer att  $v(y) = a w(y)^b$ , där  $a$  och  $b$  beror på tidsskillnaden  $t_2 - t_1$ . Ett ekvivalent skrivsätt är  $v(y) = (a w(y))^b$ .

I det kontinuerliga fallet med planeringsperioden  $[0, T]$  får således i ett litet intervall kring tidpunkten  $t$  en logaritmisk framställning formen (relationen definierar ett nytt  $v$ )

$$b(t) \cdot v(x(t), t) = b(t) \cdot \log (w(x(t)) \cdot a(t))$$

där  $x(t)$  är konsumtionsintensiteten vid tidpunkten  $t$ . Kan man möjligen sluta sig till hur  $b(t)$  borde se ut? Svaret ges av Strotz' fenomen, som beskrivs nedan.

Strotz' fenomen (Chakravarty (1969) p. 41-45)

Antag att vi har perfekt kunskap om hela planeringsperioden. Vid planering över en tidsperiod  $[0, T] = I$  kan man låta funktionalen  $u(x)$  även bero på den tidpunkt  $\tau$  vid vilken evalueringen sker. Låt diskonteringsfaktorn vara  $b(t)$ , normerad så att  $b(0) = 1$ . Definiera funktionalen

$$u_{\tau}(x) = \int_0^T b(t-\tau) v(x(t), t) dt \quad \text{där } x(t) \text{ är konsumtionsintensiteten vid tidpunkten } t.$$

Hur skall  $b$  få variera, så att planeraren inte känner ånger vid någon tidpunkt  $\tau$  för hur han har planerat hela tidsperioden från 0 till  $T$ ? Inför variabeln  $y(t) = \int_0^t x(s) ds$  med bivillkoret  $y(T) = K$ , där  $K$  är den totala konsumtions-

stocken, vilket tillgodoser fallet med ändlig tidshorisont. Den totala konsumtionen i intervallet  $[0, t]$  anges alltså av  $y(t)$ .

$$\text{Maximering av } u_{\tau}(x) = \int_0^T b(t-\tau) v(\dot{y}, t) dt \text{ ger:}$$

Första-variationen enligt Euler-Lagranges ekvationer:

$$\frac{\partial u_{\tau}}{\partial y} = \frac{d}{dt} \frac{u_{\tau}}{\partial \dot{y}} \text{ och}$$

$$0 = \frac{d}{dt} (b(t-\tau) \frac{\partial v}{\partial \dot{y}}) = \dot{b}(t-\tau) \frac{\partial v}{\partial \dot{y}} + b(t-\tau) \frac{\partial \dot{v}}{\partial \dot{y}}$$

$$\text{dvs } \frac{\dot{b}(t-\tau)}{b(t-\tau)} = - \frac{\frac{\partial \dot{v}}{\partial \dot{y}}}{\frac{\partial v}{\partial \dot{y}}} \quad \tau \leq t \leq T \quad (3.1)$$

Integrering ger att  $b(t-\tau) \frac{\partial v}{\partial \dot{y}} = \text{konst}$ , vilket utsäger att det diskonterade värdet av marginalnyttan av konsumtionen är konstant.

Andra-variationen kräver att  $\frac{\partial^2 v}{\partial \dot{y}^2} < 0$  för maximum, vilket gäller för avtagande marginalnytta. Antag att den optimala strategin  $x_0(t)$  gäller för  $\tau = 0$ . Vid en senare tidpunkt  $\tau'$  tänker planeraren igenom hela strategin igen och finner, att strategin  $x_{\tau'}(t)$  är bättre än  $x_0(t)$  från och med  $\tau'$ . Nu känner han ånger, eftersom han redan vid  $\tau = 0$  kunde ha planerat strategin  $x_0(t)$ ,  $0 \leq t \leq \tau'$  och  $x_{\tau'}(t)$ ,  $\tau' \leq t \leq T$ . Detta strider mot att  $x_0(t)$ ,  $0 \leq t \leq T$ , skulle vara optimal för  $u_0(x)$ .

Ekvation (3.1), som bestämmer strategin  $x_{\tau}(t) = x_0(t)$ , beror således ej på  $\tau$ :

$$\frac{\dot{b}(t-\tau)}{b(t-\tau)} = \frac{\dot{b}(t)}{b(t)} = - \frac{\frac{\partial \dot{v}}{\partial \dot{y}}}{\frac{\partial v}{\partial \dot{y}}} = k$$

eftersom  $x_0(t)$  är optimal hela tiden. Följaktligen gäller att  $b(t) = e^{kt}$ , dvs räntan  $k$  beror ej på tiden  $\tau$ .

Avslutningsvis kan vi således konstatera att en rimlig funktional för tidsförlopp kan ha formen

$$\log u(x) = \int_0^T e^{-kt} \log [w(x(t)) \cdot a(t)] dt$$

där  $w$  endast behöver bestämmas för en tidpunkt i intervallet  $[0, T]$ . Vi observerar att

$$\log u(x) = \int_0^T e^{-kt} \cdot \log w(x(t)) dt + u_0,$$

där  $u_0 = \int_0^T e^{-kt} \cdot \log a(t) dt$  är oberoende av konsumtionsinten-

siteten  $x(t)$ . Om vi använder intervallskala vid nyttobedömningen har  $u_0$  ingen betydelse. Vi kan således lika gärna använda formen

$$u(x) = e \int_0^T e^{-kt} \log w(x(t)) dt$$

3.4 Litteraturreferenser

Chakravarty, S, 1969, *Capital and Development Planning*. (M.I.T. Press). Cambridge, Massachusetts.

Godement, R, 1963, *Cours d'Algèbre*. (Hermann). Paris.

Keeney, R L, 1969, *Multidimensional Utility Functions: Theory, Assessment and Application*. Technical Report No 43, Operations Research Centre M.I.T. Press. Cambridge, Massachusetts.

Pratt, J W, 1964, *Risk Aversion in the Small and in the Large*. *Econometrica*, Vol. 25, p. 122-136.

Raiffa, H, 1968, *Decision Analysis. Introductory Lectures on Choices under Uncertainty*. (Addison-Wesley). Reading, Massachusetts.

#### 4.1 Inledning

I föregående artiklar har behandlats olika aspekter på beskrivande lokaliseringsmodeller (Folke Snickars) och på nyttobegreppet (Bertil Marksjö). Efter studiet av mera utpräglad statistiska metoder är det via diskussionen om måluppfyllelse (nytta) och därmed sammanhängande mätproblem naturligt att i denna framställning betrakta lokaliseringsproblemet som ett kontrollproblem. Jag kommer därvid att i stort ansluta mig till den terminologi för dynamiska kontrollsystem, som introducerades i den inledande artikeln (Anders Karlqvist).

Vi betraktar problemet att över tiden lokalisera verksamheter av olika slag, t ex boende och arbete, över givna ytor. Vårt system utgörs av den totala markyta (t ex en region), som är föremål för analys. I varje tidpunkt karaktäriseras systemets struktur (lokaliseringsmönstret) av tillståndet  $x(t)$ . Som funktion av tiden fattas beslut, som påverkar lokaliseringsmönstrets utveckling. Vi sammanfattar dessa beslut i styrvariabeln  $u(t)$ . Mängden tillåtna tillstånd och beslut begränsas vanligen av exogent givna restriktioner. Så kan t ex det över tiden föränderliga tillståndsrummet  $X(t)$  ange att mängden verksamhet per ytenhet måste vara positiv men mindre än maximal exploateringsgrad. Vidare kan beslutsmängden  $U(t)$  begränsas av tillgängliga resurser. Eventuellt är mängden tillåtna beslut beroende på tillståndet:  $U(t, x(t))$ . Dynamiken i systemet beskrivs av den förändring av tillståndet, som utgående från en viss struktur,  $x(t)$ , blir följden av det fattade beslutet  $u(t)$ .

Vi antar att tillståndsförändringen (från något starttillstånd,  $x(t_0)$ ) kan sammanfattas i följande differentialekvation:

$$\frac{dx(t)}{dt} = \varphi(t, x(t), u(t))$$

Vidare förutsätter vi att nyttan av lokaliseringsmönstrets utveckling över tiden kan avläsas i en målfunktion (egentligen en funktional)  $\eta(x(t), u(t))$ . Målfunktionen mäter uppfyllelsen av de målsättningar för regionens markanvändning, som uppställts exogent. Generellt står vi alltså inför följande kontrollproblem:

maximera  $\eta(x(t), u(t))$

då

$$\left\{ \begin{array}{l} \frac{dx(t)}{dt} = \varphi(t, x(t), u(t)) \\ x(t) \in X(t) \\ u(t) \in U(t, x(t)) \end{array} \right.$$

I fortsättningen skall vi göra viktiga begränsningar i denna allmänna problemformulering. Begränsningarna betingas väsentligen av att vi vill ägna huvuddelen av den fortsatta framställningen åt att diskutera tillämpningen av välkända metoder ur matematisk programmering på lösning av styrproblem.

#### 4.2 Begränsning av problemet

Vi skall i huvudsak göra tre begränsningar av det inledningsvis formulerade allmänna styrproblemet:

##### Deterministiskt problem

Övergångsfunktionen  $\varphi$  anger hur systemet förändras betingat av visst utgångsläge och vissa fattade beslut. Denna förändring innehåller i allmänhet stokastiska element av olika karaktär. Så kan t ex omgivningens inflytande på systemet göra att resultatet av en viss applicerad kontroll (beslut) påverkas av ej förutsedda ändringar av exogena variabler. En annan form av slumpmässighet består i en ofullständig information om systemets tillstånd. Den stokastiska styrteorin behandlar problem med bland annat dessa typer av störningar (se t ex Lindquist (1970)). Vi skall emellertid i fortsättningen inskränka oss till att studera det rent deterministiska problemet: Vi utgår ifrån att vi har fullständig information om systemets struktur (lokaliseringsmönstret) och att tillståndsförändringen på grund av ett fattat beslut beskrivs på ett från början känt sätt av övergångsfunktionen  $\varphi$ .

##### Ändlig tidshorisont

Av bland annat praktiska skäl (arbetsvolym, databrist) begränsas ofta planeringen till en ändlig tidsperiod. Vi inskränker oss till att betrakta detta fall och fixerar en tidshorisont  $T$ .

Tidsintervallet  $[0, T]$  utgör alltså vår planeringsperiod. Genom att ej betrakta systemet efter tiden  $T$  löper vi risken att gynna den kortsiktiga måluppfyllelsen (tiden  $[0, T]$ ) på bekostnad av den långsiktiga (tiden efter  $T$ ). På något sätt måste hänsyn tas till tiden efter planeringsperiodens slut. Två alternativa sätt att bygga in ett sådant hänsynstagande är (se exempel i Appelgren (1969)):

1. Restriktioner på sluttillståndet  $x(T)$ , t ex undre gränser för vissa komponenter. Detta förhindrar, att vi slutar med ett tillstånd, som leder till låg måluppfyllelse efter  $T$ .
2. Värdering av sluttillståndet, t ex genom att lösa grovplaneringsproblem för längre period.

Värderingen av nytta över tiden inom planeringsperioden har behandlats i föregående artikel (Bertil Marksjö).

#### Diskret tid

Beslut som påverkar regionens lokaliseringsmönster fattas kontinuerligt, och systemets utveckling bör därför i allmänhet bäst beskrivas av en modellformulering i kontinuerlig tid. För den fortsatta framställningen skall vi dock som bakgrund tänka oss en något annorlunda situation. Vi antar, att lokaliseringsmönstrets förändring bestäms av regionplaner, som fastställs för en tid av 5 år. I enlighet med antagandet om deterministisk övergångsfunktion förutsätter vi att dessa regionplaner är styrande för markanvändningen. Vi betraktar nu en myndighet, som skall göra upp en långtidsplan för regionens utveckling. Ur denna myndighets synpunkt kan tidpunkterna för fastställandet av de enskilda femårsplanerna uppfattas som de enda beslutssituationerna. Om vi dessutom antar, att nyttovärderingen av en given tidsutveckling av lokaliseringsmönstret approximativt kan anges som funktion av tillstånd och beslut i de olika beslutspunkterna, kan vi begränsa oss till att endast beskriva systemet vid dessa tidpunkter. Vårt kontrollproblem är därmed ett ändligt-dimensionellt problem i diskret tid. Det ovan förda resonemanget kan kritiseras i många avseenden och skall här endast tjäna som ett exempel på en situation, då approximation av ett kontinuerligt problem med en formulering i diskret tid kan vara motiverad.



Matematiskt innebär det tidskontinuerliga fallet ett oändligt-dimensionellt optimeringsproblem. Sådana problem behandlas i kontrollteorin med utgångspunkt från Pontyagins maximumprincip. Det kontrollteoretiska angreppssättet, som inte skall beröras närmare här, innebär att analysen hela tiden utförs i kontinuerlig tid, och att man först vid slutlig numerisk beräkning inför diskret tidssteg. Som kontrast till denna metodik kan man betrakta ett angreppssätt, som är mera naturligt för tillämpning av matematisk programmering. Man approximerar överföringsfunktionen enligt Eulers metod,  $\frac{dx(t)}{dt} = \frac{x(t+\Delta t) - x(t)}{\Delta t}$ , där  $\Delta t$  anger ett litet tillskott i  $t$ , och målfunktionen diskretiseras på motsvarande sätt. Därigenom har redan från början problemet reducerats till ett ändligt-dimensionellt problem, och hela klassen av metoder i matematisk programmering är därmed tillämpbar. Vi ansluter oss till detta senare angreppssätt. För en diskussion av samspillet mellan kontrollteori och matematisk programmering se Luenberger (1970).

#### Flerstegiga beslutsproblemet

Med ovan beskrivna begränsningar kan vi nu formulera vårt problem att över tiden optimalt lokalisera verksamheter till givna ytor i en region som ett deterministiskt flerstegigt beslutsproblem. Vi betraktar systemet i beslutstidpunkterna  $1, 2, \dots, T$ . Givet ett visst utgångstillstånd  $x(1)$  söker vi den beslutsplan eller strategi  $u(1), \dots, u(T-1)$  som med hänsyn till restriktionerna maximerar nyttan  $\eta(x(1), u(1), x(2), u(2), \dots, x(T))$ . Med hjälp av övergångsfunktionen  $x(t+1) = \varphi(t, x(t), u(t))$  kan målfunktionen förenklas till  $\eta(x(1), u(1), u(2), \dots, u(T-1))$ .

Sammanfattningsvis:

Sök den strategi  $u(1), u(2), \dots, u(T-1)$  som utgående från starttillståndet  $x(1)$

maximerar nyttofunktionen (minimerar kostnadsfunktionen)

$$\left\{ \eta(x(1), u(1), \dots, u(T-1)) \right\}$$

under restriktionerna

$$\begin{cases} x(t+1) = \varphi(t, x(t), u(t)) \\ x(t) \in X(t) \\ u(t) \in U(t, x(t)) \end{cases}$$

Den fortsatta framställningen koncentreras på tillämpningen av metoder i matematisk programmering för lösning av flerstegiga beslutsproblemet. Metoderna förutsättes i princip kända. Huvudvikten kommer att läggas vid de begränsningar, som vidlåder de olika metoderna. Ett enkelt exempel används för att belysa användningen av metodiken.

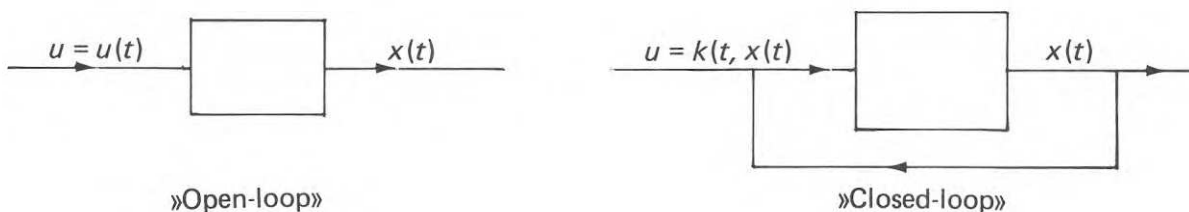
#### 4.3 Begreppen styrprogram och styrlag

Vid formuleringen av det flerstegiga beslutsproblemet ovan sökte vi en optimal strategi bestående av beslut vid tidpunkterna 1, 2, ... T-1. En sådan beslutsplan, då besluten endast beror på tiden ( $u = u(t)$ ), kallar vi ett styrprogram.

Inom både reglerteknik och kontrollteori utgör kontrollag eller styrlag viktiga begrepp. En styrlag är en beslutsplan som definierar optimal styrning i varje tillstånd i varje tidpunkt ( $u = k(t, x(t))$ ) (se Anders Karlqvists inledning). Vi skall ytterligare understryka skillnaden mellan ett styrprogram och en styrlag (se FIG. 4.1).

Med ett givet starttillstånd  $x_0$  definierar både styrlagen och styrprogrammet samma optimala trajektorier (bana i tillståndsrummet). Skillnaden uppstår om t ex systemet skulle utsättas för en exogen störning då det befinner sig i punkten  $x_1$  på optimala trajektorian. Tillståndet ändras då till  $x_2$ . Med styrlagen känner vi optimal styrning även från  $x_2$ , medan vi enligt styrprogrammet fortsätter med samma styrning som skulle varit optimal vid  $x_1$ . Ingenting säger, att denna skulle vara optimal även vid  $x_2$ .

Skillnaden mellan styrlag och styrprogram kan även åskådliggöras med begreppen 'closed-loop'-styrning respektive 'open-loop'-styrning från reglertekniken:



Orsaken till att vi lagt stor vikt vid dessa begrepp är att vi kommer att referera till dem senare.

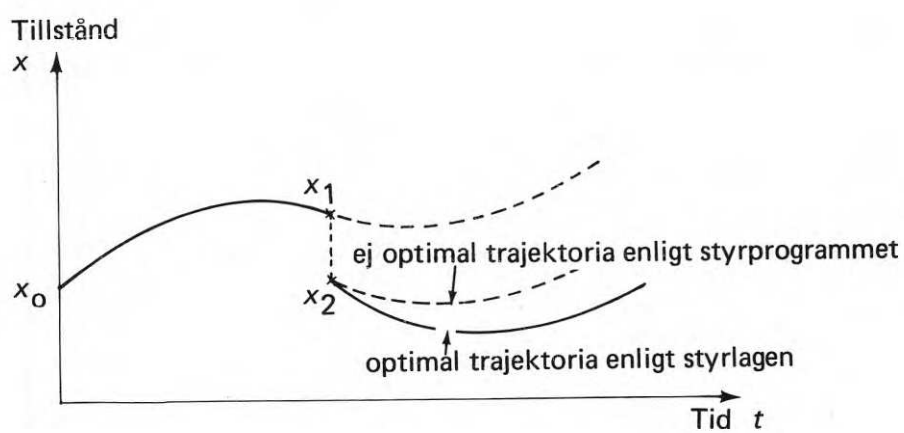


FIG. 4.1 Jämförelse mellan styrning enligt styrprogram respektive styrlag vid ett stört system.

#### 4.4 Linjära staden

Låt oss ta följande enkla konkreta exempel som utgångspunkt för en diskussion av några metoder för lösning av det deterministiska flerstegiga beslutsproblemet.

Vi betraktar problemet att över tiden lokalisera verksamheter till diskreta punkter längs en rät linje.



Dela upp den linjära staden i lika stora områden (intervall) och låt mittpunkten i varje intervall representera området. Antag att två aktiviteter med kända över tiden växande mängder skall lokaliseras till dessa områden (punkter). Med samma formalism, som användes vid formuleringen av det flerstegiga beslutsproblemet i avsnitt 2, kan vi allmänt formulera det dynamiska lokaliseringsproblemet i den linjära staden:

Tillstånd:

$x(t)$  (mängd av respektive verksamhet i respektive område)

$x(t) \in X(t)$  (givna gränser för exploateringsgraden i respektive område)

Beslutsvariabel:

$u^+(t), u^-(t)$  (nybyggnad, rivning under tidsintervallet  $(t, t+1)$ )

$u(t) = u^+(t) - u^-(t)$

$u(t) \in U(t)$  (given tillväxt över tiden av mängderna av respektive verksamhet)

Övergångsfunktion:

$x(t+1) = \varphi(t, x(t), u(t))$

Målfunktion:

$\eta(x(1), u(1), \dots, u(T-1))$

I det följande skall diskuteras metoder för att lösa detta problem under antaganden om olika restriktioner på tillåtna variabelmängder,  $X(t)$  och  $U(t)$ , och på ingående funktioner  $\varphi$  och  $\eta$ .

#### 4.5 Linjär programmering (LP)

Följande krav måste vara uppfyllda för att linjär programmering skall kunna tillämpas på flerstegiga beslutsproblemet:

- 1)  $\varphi$  och  $\eta$  linjära i ingående storheter.
- 2)  $X(t)$  och  $U(t)$  begränsade av linjära samband (polyedrar).

Fördel med LP: Effektiva algoritmer för stora problem (1000-tals variabler).

Nackdel med LP: Linjaritetskraven leder ofta till orealistiska approximationer av verkligheten (se t ex Baumol, Bushnell (1967)),

#### EXEMPEL 1

Sök en över tiden optimal lokalisering av två verksamheter i den linjära staden då målfunktionen anger dels kostnader för kontakt med stadens centrum (avståndskostnad), dels byggkostnader. Tillväxten av respektive verksamhet antages given i varje tidssteg. Om kostnaden för nybyggnad antas oberoende av i vilket område nybyggnaden sker, utgör netto-nybyggnadskostnaden en konstant. Endast eventuell kostnad för flyttning av verksamhet (rivning + nybyggnad) behöver då tagas in i målfunktionen. Denna kostnad är proportionell mot mängden av nedriven verksamhet. Vi får följande formulering:

$$X(t): \sum_a x_{a_i}(t) \leq h_i(t), x_{a_i}(t) \geq 0$$

$$U(t): u(t) = u^+(t) - u^-(t), u^+(t) \geq 0, u^-(t) \geq 0$$

$$\sum_i u_{a_i}(t) = n_a(t)$$

$$\varphi: x(t+1) = x(t) + u(t)$$

$$\eta: F = \sum_t \sum_a \sum_i (x_{a_i}(t) \cdot C_{a_i} \cdot w(t) + v \cdot w(t) \cdot u_{a_i}^-(t))$$

#### Beteckningar:

$x_{a_i}(t)$ : Mängd av verksamhet  $a$  i område  $i$  vid tiden  $t$ .  
Utgör en komponent av vektorn  $x(t)$ .

$h_i(t)$ : Maximal exploateringsgrad i område  $i$  vid tiden  $t$

$u_{a_i}^+(t)$ : Tillskott av verksamhet  $a$  i område  $i$  på grund av nybyggnad under tidsintervallet  $(t, t+1)$

- $u_{a,i}^-(t)$  : Minskning av verksamhet  $a$  i område  $i$  på grund av rivning under tidsintervallet  $(t, t+1)$   
 $u_{a,i}(t)$  : Nettotillskott av verksamhet  $a$  i område  $i$  under tidsintervallet  $(t, t+1)$   
 $n_a(t)$  : Total tillväxt av verksamhet  $a$  under tidsintervallet  $(t, t+1)$   
 $C_{a,i}$  : Olägenhet för verksamhet  $a$  att befinna sig i område  $i$  räknat som kostnad för kontakter med stadens centrum  
 $w(t)$  : Diskonteringsfaktor  
 $v$  : Viktsfaktor för 'flyttningkostnad' relativt 'avståndskostnad'.

Vårt problem  $\min \{ \eta \}$  under restriktionerna  $u(t)$

$$\begin{cases} x(t) \in X(t) \\ u(t) \in U(t) \\ x(t+1) = \varphi(t, x(t), u(t)) \end{cases}$$

uppfyller kraven 1 och 2 ovan och kan lösas med linjärprogrammering. Variablerna utgörs av  $u(1), \dots, u(T-1)$ . Utgångstillståndet  $x(1)$  antas innebära helt oexploaterade områden, dvs  $x(1) = 0$ .

Som illustration visas här några exempel på lösningar, FIG. 4.1, av detta enkla exempel. Lösningarna förutsätter följande:

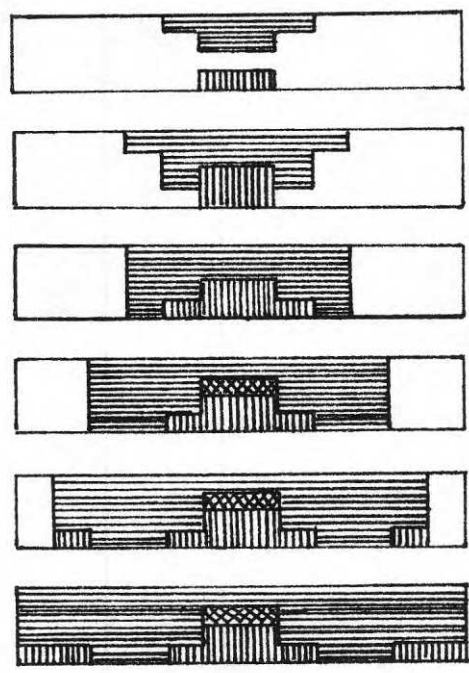
- Konstant exploateringsgrad:  $h_i(t) \equiv \text{konstant}$
- Linjär tillväxt:  $n_a(t) = N_a$  ( $N_a = \text{konstant som beror av } a$ )
- Ingen diskontering:  $w(t) \equiv \text{konstant}$
- Linjär avståndskostnad:  $C_{a,i} = k_a \cdot d(i)$ ,  $d(i) = \text{avstånd till centrum från område } i$ ,  $k_a = \text{relativt kontaktbehov till stadens centrum för verksamhet } a$
- Symmetri kring stadens centrum.

Viktsfaktorn  $v$  och 'kontaktbehoven' till stadens centrum  $k_a$  tillåts variera. Vid parvis jämförelse mellan lösningarna 1 - 4 som redovisas i FIG. 4.2 gäller:

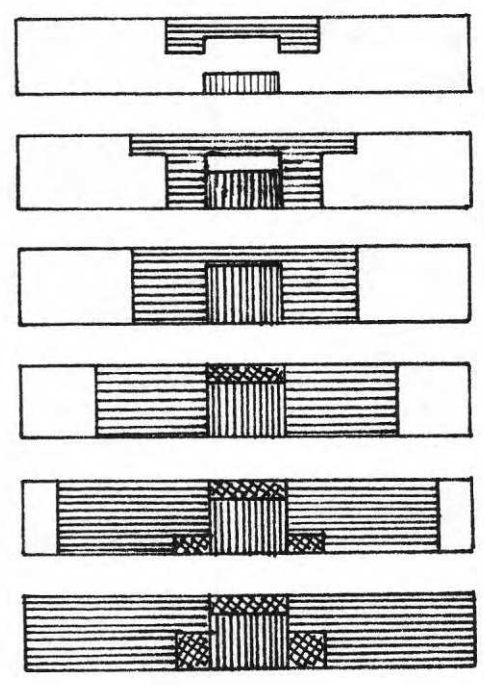
horisontell led:

kontaktbehovet arbete-centrum relativt boende-centrum är dubbelt så stort i den högra lösningen.

1.  $v = v_0, k_{\text{arbete}} = 1,5 \cdot k_{\text{boende}}$

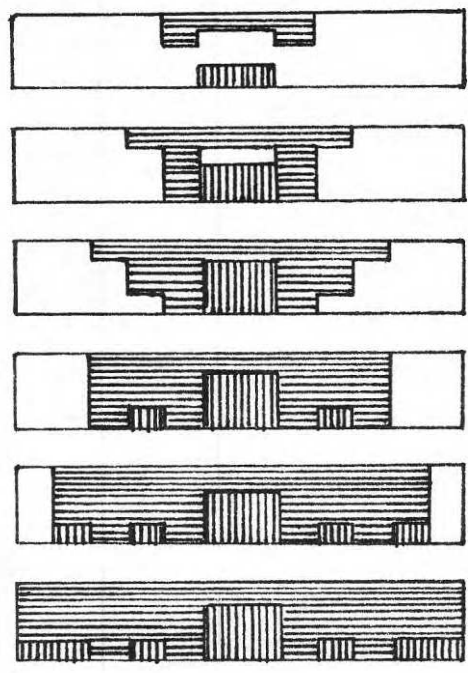


2.  $v = v_0, k_{\text{arbete}} = 3 \cdot k_{\text{boende}}$

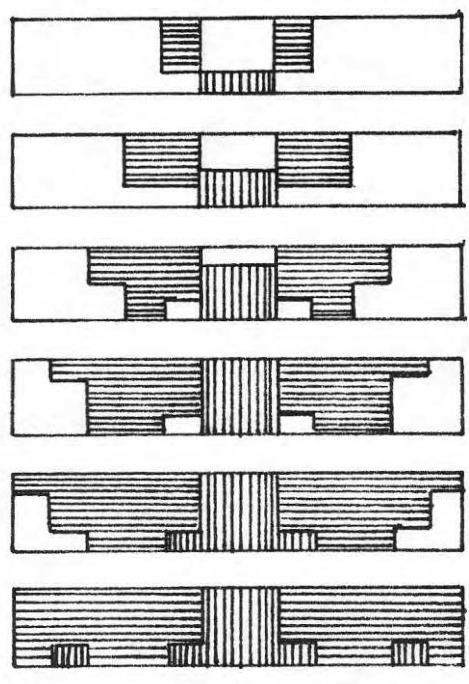


↓  
tid

3.  $v = 2 \cdot v_0, k_{\text{arbete}} = 1,5 \cdot k_{\text{boende}}$

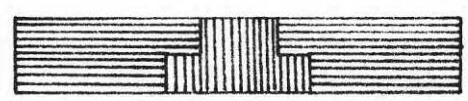


4.  $v = 2 \cdot v_0, k_{\text{arbete}} = 3 \cdot k_{\text{boende}}$



↓  
tid

statistiskt optimal lösning:



Beteckningar




-  : Arbetsplats
-  : Bostad
-  : Arbetsplats som ersatt rivna bostäder

FIG. 4.2 Dynamiska lösningar för den linjära staden då endast kontakten till centrum beaktas. Relativa kontaktbehovet ( $k_{\text{arbete}}/k_{\text{boende}}$ ) och tidsvärderingen ( $v$ ) varieras.



vertikal led:

relativa vikten av byggkostnaden i förhållande till avståndskostnaden är dubbelt så stor i den undre lösningen.

I samtliga exempel innebär den statistiskt optimala lösningen att arbetsplatserna koncentreras till stadens mitt ( $k_{\text{arbete}}$  förutsättes i dessa exempel större än  $k_{\text{boende}}$ ). Det faktum att områdena ligger längs en rät linje är enbart motiverat av pedagogiska skäl. Modellen är tillämpbar på godtyckliga centrumorienterade geografiska strukturer.

Vi noterar fem inte helt oväntade resultat:

1. Vid optimering över tiden fås i allmänhet strukturer som avviker från de i varje tidssteg optimala.
2. Koncentrationen av arbetsplatser till stadens centrum ökar då kontaktbehovet arbete-centrum växer relativt kontaktbehovet boende-centrum.
3. Ökad 'flyttningskostnad' bidrar till ökad spridning av arbetsplatserna.
4. Graden av anhopning av arbetsplatserna till stadens centrum avgörs av värderingen av 'kontaktbehov' i förhållande till byggkostnader.
5. Dessa resultat antyder att det för de olika parametrarna (värderingarna) existerar intressanta 'omslagspunkter', där lokaliseringens karaktär ändras (t ex spridningen av arbetsplatser). Andra omslagspunkter kan troligen hittas genom att variera tidsvärdering och avståndsvärdering.

#### 4.6 Dynamisk programmering (DP)

Dynamisk programmering är kanske den teknik som ligger närmast till hands av de metoder som står till buds för lösning av det flerstegiga beslutsproblemet. DP grundar sig på den så kallade optimalitetsprincipen.

##### Optimalitetsprincipen innebär

Givet är en optimal trajektoria från A till C, se FIG. 4.3. Den del (I) av denna trajektoria som förbinder B (godtycklig) med C är då optimal mellan B och C. Denna optimalitetsegenskap ställer följande villkor på målfunktionen  $\eta$ :

1. Rekursivitet:

Målfunktionens värde kan beräknas genom rekursion i varje tidssteg:  $\hat{V}(x(t)) = \min_{u(t)} V_t(x(t), u(t), \hat{V}(x(t+1)))$ , FIG. 4.4.

$\hat{V}(x(t))$  är, utgående från tillståndet  $x(t)$  vid tiden  $t$ , optimalt värde på målfunktionen för tiden  $t$  och framåt.

Sök min  $\eta = \hat{V}(x(1))$

2. Monotonitet:

$V_t$  är monotont växande i  $\hat{V}(x(t+1))$

Båda dessa villkor är uppfyllda i det praktiskt vanliga fallet då  $\eta = \sum_t \eta_0(x(t), u(t))$ .

Dynamisk programmering ger globalt optimum för målfunktionen oberoende av eventuell konvexitet hos denna. Betydelsen av denna egenskap kommer att framstå klart vid diskussionen av icke-linjär programmering.

Fördel med DP:

Ger globalt optimum utan allvarliga inskränkningar på funktionerna  $\eta$  och  $\varphi$ . Dessutom fås med DP en styrlag (definierad i kapitel 4.3) dvs optimala beslut som funktion av tillståndet i varje tidssteg.

Nackdel med DP:

Även problem med få variabler kan i praktiken visa sig olösbara. Vid dator-lösning av ett problem med DP måste speciellt två begränsningar beaktas: Krav på minnesutrymme och tidsåtgång. Kravet på minnesutrymme hänför sig dels till behovet att i kärnminnet lagra optimala funktionsvärden i senaste rekursionssteget, dels till behovet att lagra optimala beslut och funktionsvärden i samtliga rekursivt passerade tidssteg. Tidsåtgången hänför sig till minimeringen av funktionen  $V_t$  (se ovan) för varje tillstånd  $x(t)$  i varje rekursionssteg. Av dessa begränsningar är ofta kravet på utrymme i kärnminnet gränssättande. Som exempel kan vi betrakta ett problem med tre kontinuerliga variabler. Om dessa var och en diskretiseras till 100 nivåer leder detta till  $10^6$  tillstånd. Lagringsbehovet för optimala funktionsvärden i dessa tillstånd överskrider klart de flesta datamaskinens kärnminneskapacitet.

Ett försök att reducera kravet på kärnminneskapacitet utan att för den skull orimligt öka tidsåtgången är den sk 'state increment dynamic programming' som för vissa problem väsentligt minskar lagringsbehovet i snabbminnet, se Larson (1968). Metoden går ut på att det diskreta tillståndsrummet delas in i 'block' och att endast ett sådant block lagras i snabbminne. Övriga block lagras i långsamma minnet. Detta kombineras med en form av händelsestyrd rekursion, där en viss styrning appliceras tills någon tillståndsvariabel ändras med en nivåenhet, vilket innebär variabelt tidssteg.

#### EXEMPEL 2

Linjära staden, fem områden, två verksamheter. För att karakterisera strukturen krävs 10 variabler (en variabel för varje kombination område - verksamhet) :  $x(t)$ .

Vi antar att målfunktionen kan skrivas:  $\eta = \sum_t \eta_0(x(t), u(t))$ , där eventuellt  $\eta_0(x(t), u(t)) = \eta_1(x(t)) + \eta_2(u(t))$ , dvs målfunktionen sammansätts i varje tidssteg av en strukturkostnad och en styrkostnad. Strukturkostnaden kan t ex vara ett mått på kontaktolägenhet (invers till accessibilitet) medan styrkostnaden kan vara en funktion av byggkostnaden.

I övrigt gäller:

$$\begin{cases} x(t) \in X(t) \\ u(t) \in U(t) \\ x(t+1) = \varphi(t, x(t), u(t)) \end{cases}$$

Inför beteckningen  $\hat{V}(x(t))$  för optimala värdet av  $\sum_{s \geq t} [\eta_0(x(s), u(s))]$  i tillståndet  $x(t)$  och starta med  $\hat{V}(x(T)) = \eta_0(x(T), u(T))$ .

Vi får rekursionsekvationen:

$$\hat{V}(x(t)) = \min_{u(t)} \{ \eta_0(x(t), u(t)) + \hat{V}(\varphi(t, x(t), u(t))) \}$$

Trots att problemet bara omfattar 5 områden är det i praktiken lösbart med vanlig DP endast under förutsättning att varje variabel kvantifieras i högst fyra nivåer. Vi noterar att lösbarhet med DP i detta fall kräver en synnerligen grov indelning vid kvantifiering av tillstånden. Denna begränsning får vägas mot de fördelar, som vi beskrivit ovan.

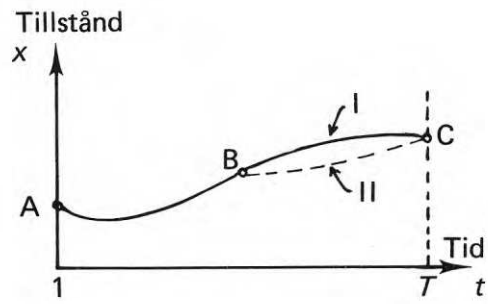


FIG. 4.3 Optimalitetsprincipens innebörd.

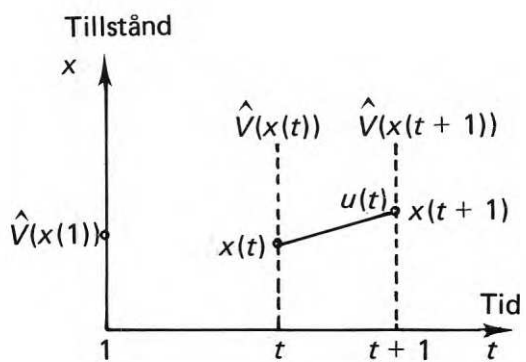


FIG. 4.4 Rekursivitet.

#### 4.7 Icke-linjär programmering (ILP)

Dynamisk programmering leder ofta till ett tillståndsrum av helt ohanterlig dimension. För de fall då ingående funktion är icke-linjär kan det tyckas naturligt att i stället försöka lösa det flerstegiga beslutsproblemet med icke-linjär programmering.

##### Fördelar med ILP:

Man kan lösa problem av relativt stor dimension (med vissa metoder och speciell problemstruktur: 500 - 1000 variabler). Inga inskränkningar i princip på funktionerna  $\eta$  och  $\varphi$  eller restriktionerna  $X(t)$  och  $U(t)$ .

##### Nackdelar med ILP:

De vanligaste algoritmerna för icke-linjär programmering leder i motsats till DP endast till stationära lösningar (uppfyller Kuhn-Tucker-villkoren för stationaritet). Sålunda krävs ytterligare konvexitetsvillkor på ingående funktioner för att säkra global optimalitet hos lösningen. Medan DP ger styrlag fås med ILP endast ett styrprogram, dvs styrningen som funktion av tiden (se diskussion i kapitel 4.3).

##### EXEMPEL 3

Vi skall illustrera användningen av ILP med ett problem av enkel struktur. Återigen väljer vi linjära staden som exempel. Staden delas upp i fem områden, och vi antar liksom i tidigare exempel två verksamheter: boende och arbete. Med i huvudsak samma beteckningar som i exempel 1 (kapitel 4.5) formulerar vi problemet enligt följande:

$$X(t): \sum_a x_{a_i}(t) \leq h_i(t), x_{a_i}(t) \geq 0$$

$$U(t): u(t) = u^+(t) - u^-(t)$$

$$\sum_i u_{a_i}(t) = n_a(t)$$

$$u^+(t) \geq 0, u^-(t) \geq 0$$

$$\varphi: x(t+1) = x(t) + u(t)$$

$$\eta: \eta = \sum_t \sum_{a,b} \sum_{i,j} (x_{a_i}(t) \cdot K_{ab} \cdot d_{ij} \cdot x_{b_j}(t) \cdot w(t)) +$$

$$+ \sum_t \sum_a \sum_i v \cdot w(t) \cdot u_{a_i}^-(t)$$

I stället för att som i exempel 1 endast betrakta kontakter med stadens centrum förutsätter vi här ömsesidiga kontakter mellan ingående verksamheter. Detta leder till att vi måste göra antaganden om 'kontaktbehov' eller kontaktfrekvenser mellan aktiviteterna  $a$  och  $b$ :  $K_{ab}$ . Vidare räcker det inte att ange avstånd till centrum, utan avstånd mellan varje par av områden måste bestämmas:  $d_{i,j}$ . Beteckningen  $\Sigma'$  står för en summa  $\Sigma$  där endast en av kombinationerna  $(a,i,b,j)$  och  $(b,j,a,i)$  förekommer.

$$\text{Problemet } \min_{x(1), u(t)} \{ \eta \} \quad \text{då} \quad \begin{cases} x(t) \in X(t) \\ u(t) \in U(t) \\ x(t+1) = \varphi(t, x(t), u(t)) \end{cases}$$

är ett kvadratisk programmeringsproblem. Problemet är i allmänhet av icke-konvex typ. Resultaten kan därför med den algoritm vi använt inte förväntas vara mer än stationära lösningar eller lokala optima. Ordet lösning skall i resten av detta avsnitt fattas i denna bemärkelse.

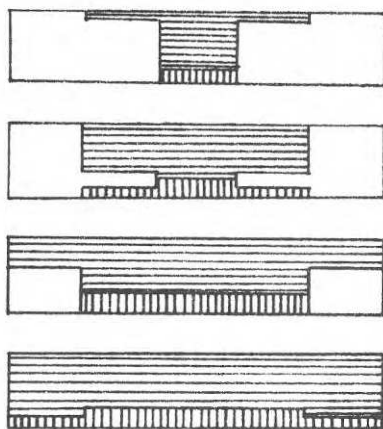
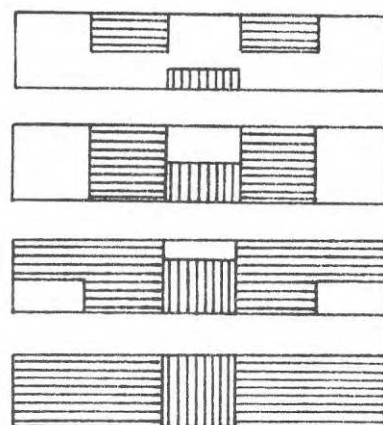
I likhet med exempel 1 (sid. 74) har ett antal förutsättningar gjorts:

- a) Konstant maximal exploateringsgrad:  $h_i(t) \equiv \text{konstant}$
- b) Linjär tillväxt:  $n_a(t) = N_a$
- c) Ingen diskontering:  $w(t) \equiv \text{konstant}$
- d) Linjär avståndskostnad:  $d_{i,j}$  = avståndet mellan område  $i$  och  $j$ .

Viktsfaktorn ('flyttningkostnad' relativt 'kontaktkostnad')  $v$  och 'kontaktbehoven'  $K_{ab}$  tillåts variera.

Vi skall i FIG. 4.6 illustrera med ett par exempel på 'lösningar'. Den algoritm som använts bygger på Zoutendijk's metod med 'feasible directions'. För att starta algoritmen används startlösningar enligt FIG. 4.5.

På grund av svårigheten att ange lösningarnas karaktär av stationära värden, lokala optima eller globala optima avstår vi från djupare analys av dessa resultat. Vi noterar blott att för lösning 3 och 4 överensstämmer sluttillstånden med de statistiskt optimala lösningarna. Spridningen av arbetsplatserna över område-

Startlösning  $x_0$ Startlösning  $x_1$ 

Startlösning vid best. av statistiskt optimal lösning:

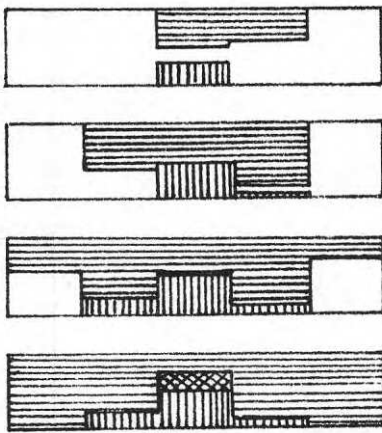


FIG. 4.5 Startlösningen utgör den utbyggnadsstrategi över tiden (dynamiska fallet) alternativt den aktivitetsfördelning (statiska fallet), som bestämmer utgångsläget för algoritmens första iteration (försök). Genom successiva iterationer förändras sedan lösningarna tills en situation uppstår då ingen ytterligare förbättring av målfunktionen kan åstadkommas genom små förändringar av variablerna. Den lösning som därigenom erhållits är vid icke-konvexa problem (problem som kan innehålla lokala optima och stationära lösningar) i allmänhet beroende av valet av startlösning.

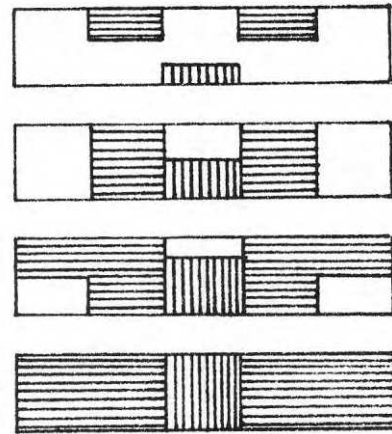
Beteckningar: Se FIG. 4.6.



1.  $K_{aa} = 3$   $v = v_1$   
 $K_{ab} = 2$  Startlösning:  $x_0$   
 $K_{bb} = 1$  (se FIG. 4.5)

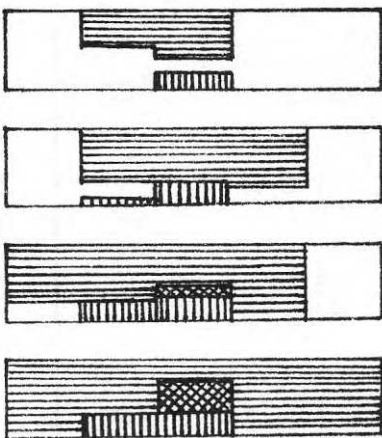


2.  $K_{aa} = 3$   $v = 2v_1$   
 $K_{ab} = 2$  Startlösning:  $x_1$   
 $K_{bb} = 1$  (se FIG. 4.5)

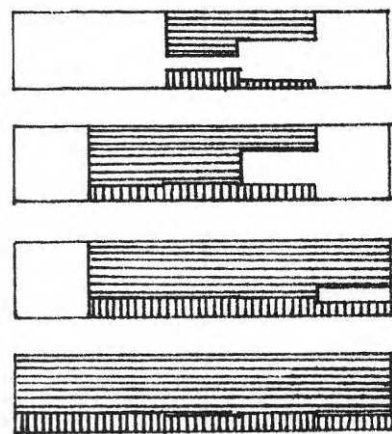


↓  
tid

3.  $K_{aa} = 3$   $v = v_1$   
 $K_{ab} = 2$  Startlösning:  $x_1$   
 $K_{bb} = 1$  (se FIG. 4.5)



4.  $K_{aa} = 1$   $v = v_1$   
 $K_{ab} = 5$  Startlösning:  $x_1$   
 $K_{bb} = 20$  (se FIG. 4.5)



↓  
tid

Statistiskt optimala »lösningar»:

1.  $K_{aa} = 3$   
 $K_{ab} = 2$   
 $K_{bb} = 1$



2.  $K_{aa} = 1$   
 $K_{ab} = 5$   
 $K_{bb} = 20$



Beteckningar



: Arbetsplats

$a =$  arbete

$K_{ab} =$  »kontaktbehov»  
mellan arbete  
och bostad



: Bostad

$b =$  bostad



: Arbetsplats som ersatt rivna bostäder

$v =$  viktsfaktor för  
byggkostnad  
relativt avstånds-  
kostnad

FIG. 4.6 Dynamiska lösningar för den linjära staden då ömsesidiga kontakter mellan arbete och bostad beaktas. De relativa kontaktbehoven ( $K_{aa}$ ,  $K_{ab}$ ,  $K_{bb}$ ) och tidsvärderingen ( $v$ ) varierar.

na bestäms dels av avvägningen mellan olika 'kontaktbehov' och dels av avvägningen mellan strukturkostnad (kontaktkostnad) och styrkostnad (byggkostnad).

I Abadie (1970) finns en uppsats skriven av J. Abadie som behandlar tillämpning av en metod för icke-linjär programmering, "Generalized reduced gradient", på kontrollproblem av både tidskontinuerlig och tidsdiskret typ. Enligt Abadie verkar resultaten hittills lovande.

#### 4.8 Heltalsprogrammering

Det är i beslutssituationer mycket vanligt att beslutet gäller ett antingen eller, dvs beslutet har en 0-1 lösning. Även det flerstegiga beslutsproblemet i linjära staden kan ges en sådan formulering. I många fall kan dessa problem lösas med dynamisk programmering. Ett alternativt sätt är att tillämpa heltalsprogrammering.

Vi skall formulera dels vår linjära och dels vår kvadratiska lokaliseringsmodell för den linjära staden som ett linjärt respektive kvadratiskt assignmentproblem. Dessa formuleringar skall endast ses som exempel på problem av heltalstyp, konsistenta med vår tidigare framställning. För att exemplen skall vara meningsfulla bör de aktiviteter, som är föremål för lokaliseringsanalysen, karakteriseras av odelbarheter. Det kan t ex vara fråga om produktionsenheter, som är beroende av ömsesidiga leveranser. För en utförligare diskussion se Koopmans, Beckmann (1957).

##### Linjära assignmentproblemet

Antag att staden är uppdelad i  $n$  st lika stora områden och att  $m$  st verksamheter skall lokaliseras till dessa på ett över tiden optimalt sätt. Låt oss betrakta  $T$  st tidpunkter. Vi förutsätter att en verksamhet, som en gång lokaliserats till ett visst område, stannar där under resten av planeringsperioden. Denna problemställning är alltså något förenklad jämfört med de problem som behandlats i tidigare exempel.

## BETECKNINGAR

$$x_{ij,t} = \begin{cases} 1 & \text{om aktivitet } i \text{ är lokaliserad till område } j \text{ vid} \\ & \text{tid } t \\ 0 & \text{annars} \end{cases}$$

$k_i$  = verksamhet  $i$ :s 'kontaktbehov' med stadens centrum

$d_j$  = avståndskostnad från område  $j$  till stadens centrum

$w_t$  = diskonteringsfaktor

$H_{i,t}$  = antal områden som enligt känd efterfrågan åtgår för verksamhet  $i$  vid tiden  $t$ .

Problem:

$$\text{Sök min } \sum_i \sum_j \sum_t x_{ij,t} \cdot d_j \cdot k_i \cdot w_t$$

$$\text{då } \sum_j x_{ij,t} = H_{i,t} \text{ (mängd av verksamhet } i \text{ vid tiden } t)$$

$$\sum_i x_{ij,t} \leq 1, \quad x_{ij,t} = 0 \text{ eller } 1$$

$$x_{ij(t+1)} \geq x_{ij,t}$$

Det finns algoritmer för heltalsprogrammering som löser problem av denna typ för 100-tals variabler.

Många problem av heltalstyp kan omformuleras till flödesproblem, som med speciella flödesalgoritmer är lösbara för betydligt flera variabler än vad som är möjligt med heltalsprogrammeringsmetoder.

Kvadratiska assignmentproblemet

Samma problemställning som i linjära fallet. Dock förutsätter vi här ömsesidiga kontakter mellan verksamheterna.

## BETECKNINGAR

$k_{il}$  = 'kontaktbehov' mellan verksamheterna  $i$  och  $l$  ('ett område' av varje).

$d_{jk}$  = avståndskostnad mellan område  $j$  och  $k$ .

Problem:

$$\text{Sök min } \sum_{\substack{i,j,t \\ l,k}} x_{ij,t} k_{il} d_{jk} x_{lkt} \cdot w_t$$

$$\text{då } \sum_j x_{ijt} = H_{it}$$

$$\sum_i x_{ijt} \leq 1$$

$$x_{ij(t+1)} \geq x_{ijt}$$

$$x_{ijt} = 0 \text{ eller } 1$$

I en artikel i Abadie (1970) diskuterar G.W. Graves och A.B. Whinston approximativa lösningsmetoder för det statistiska kvadratiske assignmentproblemet.

#### 4.9 Sammanfattning

Vi har formulerat uppgiften att planera markanvändningen över tiden som ett styrproblem. Denna generella modellformulering har genom tre viktiga begränsningar (ändlig, diskret tid; deterministiskt system) transformerats till ett flerstegigt beslutsproblem. Vi har koncentrerat vår framställning på tillämpning av metoder i matematisk programmering för lösning av detta problem.

Som illustration till denna diskussion har vi visat några enkla beräkningsmässiga resultat för planeringsuppgiften att över tiden lokalisera givna växande mängder av olika verksamheter till punkter längs en rät linje. Dessa exempel skall ses som renodlade illustrationer. Vi har t ex ej diskuterat relevansen av använda målfunktioner. Vi har inte heller försökt relatera våra resultat till det praktiska genomförandet, t ex möjligheten att genomföra optimal allokering av verksamheter via decentraliserat beslutsfattande. Denna senare aspekt behandlas i en följande artikel (Johan Lönnroth).

4.10 Litteraturreferenser

Abadie, J, 1970, *Integer and Nonlinear Programming*. (North Holland). Amsterdam.

Appelgren, L, 1969, *Matematiska modeller av långtidsplaneringsproblem* (Försvarets forskningsanstalt) FOA P Rapport A 8159-11. Stockholm.

Baumol, W J & Bushnell, R C, 1967, *Error Produced by Linearization in Mathematical Programming*. *Econometrica*, Vol. 35, p. 447-471.

Koopmans, T C, Beckmann, M J, 1957, *Assignment Problems and the Location of Economic Activities*. *Econometrica*, Vol. 25, p. 53-76.

Larson, R E, 1968, *State Increment Dynamic Programming*. (Elsevier). New York.

Lindquist, A, 1970, *Föreläsningar i linjär stokastisk styrteori* (Avd. för optimeringslära och systemteori, KTH) Rapport 40. Stockholm.

Luenberger, D G, 1970, *Mathematical Programming and Control Theory; Trends of Interplay* (Föredrag vid 7:de matematiska programmeringssymposiet). Stencil.

## 5 EN PLANERINGSMODELL FÖR KAPACITETSEXPANSION

Anders Karlqvist, Lars Lundqvist

5.1 Inledning

Det mesta av materialet i denna artikel är hämtat från en doktorsavhandling vid Stanford University 1969. Arbetet har utförts av D. Erlenkotter under ledning av A. Manne och har rubriken "Preinvestment Planning for Capacity Expansion: A Multi-location Dynamic Model", se Erlenkotter (1969).

Det allmänna problemet är följande: Bestäm planeringsstrategi och minimal kostnad för utbyggnad av kapaciteter i ett antal punkter i rummet för att täcka en över tiden växande efterfrågan i dessa olika punkter. En typisk egenskap för dylika problem är att investeringskostnaderna för kapacitetsutbyggnad karaktäriseras av skalekonomier. Kostnadsfunktionerna är således i allmänhet icke-konvexa och globala minima är svåra att finna.

Även för en lokaliseringpunkt är detta ett besvärande problem. Speciella resultat finns dock, t ex att om man har en oändlig yta med likformig efterfrågan och kostnaden består av en fast del och en linjär del så gäller att medelkostnaden/enhet efterfrågan kan ges konvex form, Bos (1965).

Mycket arbete har nedlagts på metodfrågor och algoritmer (se t ex Manne (1967)).

Problemet skall i korthet struktureras på följande sätt:

- a) Single-expansion timing  
Givet storlek och läge för en enda kapacitetsutbyggnad. När skall utbyggnad ske?
- b) Restriktioner så att beslut enligt a. kan fattas "närsynt" utan detaljerad kunskap om framtiden.
- c) Dekomponering av beslutsproblemet för  $n$  expansioner till  $n$  single-expansion problem.
- d) Sats: Optimal expansion från given initial kapacitet implicerar att optimal expansion från högre kapacitetsnivå inte görs tidigare än i det givna fallet.
- e) Sats d bildar grunden för en dynamisk programmeringsformulering av totalproblemet.

- f) Beräkningsproblem. Dimensionalitetbegränsningar.  
 g) Räkneexempel.  
 h) Motsvarande stationära planeringsproblem med oändlig tids-  
 horisont.  
 i) Tillämpningsexempel.

## 5.2 Single-expansion

Låt oss först införa ett antal beteckningar:

$1, 2, \dots, n$	lokaliseringspunkter,
$[0, T]$	tidsperiod
$\beta(t)$	tidsfaktor (teknologifaktor etc)
$\beta(t)p_{ij} \geq 0$	transportkostnad $i \rightarrow j$
$\beta(t)q_i \geq 0$	produktionskostnad $i$
$\theta_j(t)$	efterfrågan $i$ $j$ vid tidpunkten $t$
$\tau$	tidpunkt
$C(z, \Pi, \tau)$	kostnad för att från startkapaciteten $z$ vid tiden $\tau$ följa utbyggnadsstrategin $\Pi$ .
$f_{i(1)}(w_{i(1)}, \tau_1) > 0$	investeringskostnad för expansionen $w_{i(1)}$ $i$ $i(1)$ vid tiden $\tau_1$
$s_{ij}(t) \geq 0$	transporterad mängd slutprodukter från $i$ till $j$
$r$	diskonteringsfaktor
$w_{i(1)} > 0$	expansion $i$ $i(1)$ vid en enda tidpunkt $\tau_1$
$z_j^0 \geq 0$	startkapacitet $i$ $j$ vid tiden $t$
$z = (z_1, z_2, \dots, z_n)$	

Optimeringsproblemet kan nu formuleras:

Minimera kostnaden  $C$  med avseende på tidpunkten  $\tau_1$

där

$$C(z^0, (w_{i(1)}, \tau_1), 0) = \int_0^{\tau_1} \beta(t) \sum_i \sum_j (p_{ij} + q_i) s_{ij}(t) e^{-rt} dt +$$

$$+ e^{-r\tau_1} f_{i(1)}(w_{i(1)}, \tau_1) + \int_{\tau_1}^T \beta(t) \sum_i \sum_j (p_{ij} + q_i) s_{ij}(t) e^{-rt} dt \quad (5.1)$$

under bivillkoren

$$\sum_i s_{ij}(t) = \theta_j(t)$$

$$\begin{aligned} \sum_j s_{ij}(t) &\leq z_i^0 & 0 \leq t \leq \tau_1 \\ \sum_j s_{ij}(t) &\leq z_i^1 & \tau_1 < t \leq T \end{aligned} \quad z_i^1 = \begin{cases} z_i^0 & i \neq i(1) \\ z_i^0 + w_i & i = i(1) \end{cases}$$

I detta problem finns ett statistiskt transportproblem inbäddat, nämligen att i varje tidsögonblick minimera transportkostnaden

$$P(z,t) = \min_{s_{ij}} \sum_i \sum_j (p_{ij} + q_i) s_{ij}(t) \quad (5.2)$$

Under några tämligen milda restriktioner på transportkostnaderna kan man visa att  $s_{kk} = \min \{\theta_k(t), z_k\}$ , dvs man använder först allt man kan producera själv. Restriktionerna som behövs är av typen triangelolikheter. De kan dock bli ohållbara t ex om produktionskostnaderna beror på råvarukostnaderna och dessa varierar avsevärt i olika lokaliseringpunkter. Problemet reducerar sig under ovanstående restriktioner till att fördela överskottsproduktionen, vilket minskar dimensionen från  $n^2$  till  $n$  och övergår därmed till att bli en variant av transshipmentproblemet, se t ex Dantzig (1963).

Vidare gäller

$$P(z,t) - P(z + \Delta_1 z + \Delta_2 z, t) \leq \left[ P(z,t) - P(z + \Delta_1 z, t) \right] + \left[ P(z,t) - P(z + \Delta_2 z, t) \right],$$

där  $\Delta_1 z$  och  $\Delta_2 z$  är kapacitetstillskott.

Relationen implicerar att  $P$  är konvex i en dimension men inte nödvändigtvis i högre dimensioner. Detta transportproblem spelar en betydelsefull roll då praktiska beräkningar skall göras med modellen, och lösningen  $P(z,t)$  måste kunna uppskattas parametriskt.

Vi vill nu bestämma  $\tau_1$  så att  $C(z^0, (w_{i(1)}, \tau_1), 0)$  minimeras.

Detta är sannolikt den kritiska punkten i modellen och det fordras mycket speciella restriktioner för att garantera att man kan finna ett globalt minimum för  $C$ . Ett villkor är att  $C$  är negativt unimodal, dvs ett relativt minimum för  $C$  är också ett absolut minimum i hela intervallet  $[0, T]$ . I detta fall kan man enkelt bestämma minimum genom direkt derivering  $\frac{\partial C}{\partial \tau_1} = 0$ . Inget



är känt generellt om de villkor på  $f$ ,  $\beta$  och  $P$  som gör  $C$  unimodal. Erlenkotter ger två tillräckliga villkor.

1.  $f_{i(1)}(w_{i(1)}, \tau_1)$  är en konvex icke-växande och  $\beta(\tau_1)$  är en icke-avtagande funktion av  $\tau_1$  eller
2.  $\frac{\partial f}{\partial \tau_1} / f$  och  $\beta/f$  icke-avtagande som funktioner av  $\tau_1$

Dessa regler är som synes mycket speciella och man kan fråga sig om de är realistiska. I gengäld fordrar denna metod ingen detaljerad kunskap om framtida efterfrågan, transportkostnader eller investeringskostnader.

#### EXEMPEL

Låt  $\beta(t)$  och  $f_{i(1)}(t)$  vara konstanta, då erhålles minimum för  $C$  genom att derivera (5.1) sedan  $P(z,t)$  enligt (5.2) införts.

$$\frac{\partial C}{\partial \tau_1} = 0 = e^{-r\tau_1} \left[ P(z^0, \tau_1) - P(z^1, \tau_1) - r f_{i(1)}(w_{i(1)}, 0) \right] \quad (5.3)$$

Detta kan tolkas så, att kapacitetsutbyggnaden skall göras om omedelbar vinst i transport- och produktionskostnad är lika med omedelbar kapitalutgift i investeringar.

### 5.3 Flera utbyggnader

Den viktigaste byggbiten i modellen är nu avklarad. Givet en utbyggnad till storlek och läge kan en optimal tidpunkt för denna utbyggnad bestämmas. Nästa steg blir att se hur den informationen kan användas för att bestämma en optimal policy för flera utbyggnader

$$\Pi^{(m)} = \{(w_{i(1)}, \tau_1), \dots, (w_{i(m)}, \tau_m)\},$$

där  $w_{i(j)}$  är förutbestämd till storlek och läge.

Vi söker tidpunkterna  $\tau_1, \tau_2, \dots, \tau_m$  för dessa  $m$  utbyggnader så att  $\tau_{k-1} \leq \tau_k \leq \tau_{k+1}$ ,  $\tau_0 = 0$ ,  $\tau_{m+1} = T$ .

Grundidén är enkel. Antingen kan problemet direkt lösas som  $n$  stycken single-expansion problem eller, om någon restriktion

tion  $\tau_{k-1} \leq \tau_k$  är bindande, så erhåller man en lägre kostnad genom att permutera ordningen på utbyggnaderna.

Detta formuleras i två satser (för bevis hänvisas till Erlenkotters avhandling).

#### Sats 5.1

Om vi löst ovanstående problem och någon restriktion  $\tau_k = \tau_l$  är bindande så kan man få en lägre kostnad om man byter på  $k$  och  $l$ . Alltså de enda följder som är potentiellt optimala är de för vilka restriktionerna på  $\tau_k$  inte är bindande.

#### Sats 5.2

Om kostnaden för varje  $w_{i(k)}$  är negativt unimodal i  $\tau_k$  och  $\tau_1^*$ ,  $\tau_2^*$ , ...,  $\tau_m^*$  minimerar  $C$  utan bindande restriktioner då kan  $\tau_k^*$  bestämmas ur  $m$  stycken single-expansion problem

$$\min_k C(z^{k-1}, (w_{i(k)}, \tau_k), 0) = \int_0^{\tau_k} \beta(t) \cdot P(z^{k-1}, t) e^{-rt} dt +$$

$$+ e^{-r\tau_k} \cdot f_{i(k)}(w_{i(k)}, \tau_k) + \int_{\tau_k}^{T_k} \beta(t) P(z^k, t) e^{-rt} dt$$

$T_k$  är den tidpunkt när total efterfrågan överstiger tillgänglig kapacitet  $z^k$ , ( $\sum_j \theta_j(T_k) = z^k$ ).

#### 5.4 Formulering av dynamisk programmeringsmodell

De följder av utbyggnader som vi studerat i föregående avsnitt har inte givit någon total optimering av det ursprungliga problemet eftersom kapaciteterna  $w_{i(k)}$  har varit förutbestämda.

Det fullständiga problemet skall vi nu formulera med hjälp av dynamisk programmering. Vi vill påminna om en grundidé i dynamisk programmering: Befinner vi oss i en tidpunkt  $\tau_k$  och bestämmer optimal väg  $U^*$  för fortsättningen, så kommer vi, om vi står i tidpunkten  $\tau_l$  före  $\tau_k$  och fattar ett optimalt beslut då, att fr o m  $\tau_k$  följa vägen  $U^*$ . ( $U^*$  måste naturligtvis bestämmas för varje tänkbart tillstånd vi kan hamna i vid  $\tau_k$ ).

---

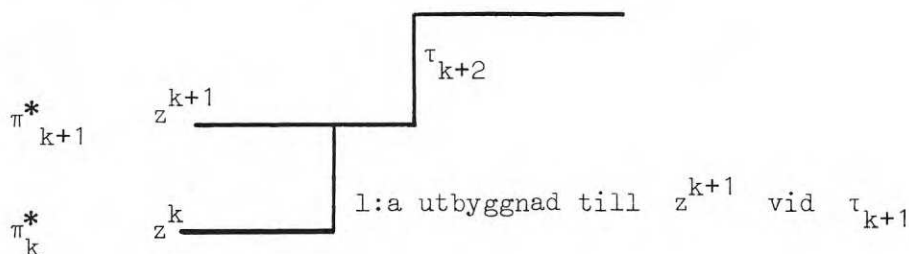
\* Stjärna betecknar här optimala storheter.

Denna princip skall vi utnyttja i vårt problem, men vi måste vara lite extra omsorgsfulla, eftersom tidpunkterna  $\tau_k$  inte är fixerade på förhand utan bestäms under optimeringens gång. För att undvika trassel med ordningen mellan tidsstegen formulerar vi följande sats (utan bevis).

### Sats 5.3

Låt  $\pi_k^*$  beteckna en initialt optimal policy då man startar med kapaciteten  $z^k$  vid tiden noll. Vid tidpunkten  $\tau_{k+1}$  görs nästa utbyggnad till  $z^{k+1}$ . Om man startar med  $z^{k+1}$  och bestämmer  $\pi_{k+1}^*$  görs nästa utbyggnad vid  $\tau_{k+2}$ . Det gäller att  $\tau_{k+2} \geq \tau_{k+1}$ .

Idén med denna sats är klar. En initialt optimal policy  $\pi_k^*$  kan sammansättas på följande sätt: Vi startar med ett första optimalt beslut om  $z^{k+1}$  vid  $\tau_{k+1}$ . Från  $\tau_{k+1}$  följer vi den optimala vägen  $\pi_{k+1}^*$ .



Satsen garanterar att "inget hänt med  $\pi_{k+1}^*$  innan vi hunnit stiga på vid  $\tau_{k+1}$ ".

I det funktions samband som behövs för att rekursivt bestämma en optimal strategi, skall vi definiera en bakåtrekursion i den meningen att vi går från en högre till en lägre kapacitetsnivå. Detta blir i själva verket p g a sats 5.3 liktydigt med att stega sig bakåt i tiden. Vi vill således knyta ihop  $C(z^h, \pi^*, 0)$  och  $C(z^k, \pi^*, 0)$ , där  $\pi^*$  betecknar optimal strategi från respektive initial nivå  $z^h$  och  $z^k$ ,  $h < k$ . Låt oss starta med  $z^h$  och göra övergången (expansion i en punkt) till  $z^k$ . Om vi betraktar en optimal följd så kan  $\tau_k^*$  bestämmas med sats 5.2.

$$C(z^h, \pi^*, 0) = \int_0^{\tau_k^*} \beta(t) P(z^h, t) e^{-rt} dt +$$

$$+ e^{-r\tau_k^*} f_{i(k)}(w_{i(k)}, \tau_k^*) + e^{-r\tau_k^*} C(z^k, \pi^*, \tau_k^*)$$

Från  $\tau_k^*$  följer vi initialt optimal policy  $\pi^*$  från  $z^k$ .

Men det gäller att

$$C(z^k, \pi^*, 0) = \int_0^{\tau_k^*} \beta(t) P(z^k, t) e^{-rt} dt + e^{-r\tau_k^*} C(z^k, \pi^*, \tau_k^*)$$

$$\therefore C(z^h, \pi^*, 0) = \int_0^{\tau_k^*} \beta(t) [P(z^h, t) - P(z^k, t)] e^{-rt} dt +$$

$$+ e^{-r\tau_k^*} f_{i(k)}(w_{i(k)}, \tau_k^*) + C(z^k, \pi^*, 0) =$$

$$= b_{hk} + C(z^k, \pi^*, 0). \quad (5.4)$$

Om vi definierar  $b_{hh}$  som kostnaden att permanent stanna i  $z^h$ ,

$$b_{hh} = \int_0^T \beta(t) P(z^h, t) e^{-rt} dt + e^{-rT} C(z^h, \pi^*, T), \quad (5.5)$$

så kan vi formulera rekursionsekvationen:

$$C(z^h, \pi^*, 0) = \min_{k \in B_h} \{b_{hk} + C(z^k, \pi^*, 0), b_{hh}\} \quad (5.6)$$

Mängden  $B_h$  definierar den mängd av möjliga kapacitetsnivåer  $z^k$  som kan nås från nivån  $z^h$  genom en tillåten ökning av en komponent i  $z^h$ . Vi kan begränsa oss till att studera ändringen i en komponent i taget på grund av att problemet är formulerat i kontinuerlig tid. Mängden av tillåtna kapacitetsnivåer ( $Z$ ) måste, på grund av kravet att ordningsföljden mellan beslut skall kunna permuteras (sats 5.1), definieras med viss omsorg. Permuteringskravet leder till att  $z$  måste besitta en sorts slutenhetsgenskap, som är uppfylld om den tillåtna nivån för varje komponent i kapacitetsvektorn utgöres av ett antal ekvidistanta nivåer. Avståndet mellan nivåerna kan variera mellan olika komponenter.  $Z$  utgör alltså ett gitter i form av en  $n$ -dimensionell hyperkub.

Egentligen borde  $B_h$  begränsas ytterligare av en restriktion  $\tau_k^* \leq \tau_{k+1}^*$  där  $\tau_k^*$  och  $\tau_{k+1}^*$  anger tidpunkten för initialt

optimala beslut från nivåerna  $z^h$  respektive  $z^k$ . En sådan restriktion är dock onödig, då vi vet (sats 3.1) att om  $\tau_k^* > \tau_{k+1}^*$  så får vi lägre kostnad genom att permutera expansionsbesluten. I den slutliga lösningen kommer sålunda villkoret  $\tau_k^* \leq \tau_{k+1}^*$  automatiskt att vara uppfyllt även om vi bortser från det vid definitionen av  $B_h$ . Betydelsen av detta är inte så stor i det icke-stationära fallet men egenskapen är viktig i det stationära fallet som behandlas nedan.

Vi har sålunda formulerat ett problem i termer av dynamisk programmering. Problemet kan uppfattas som ekvivalent med uppgiften att söka kortaste vägen (eller vägen med lägsta kostnad) mellan extrempunkter i ett gitter av kapacitetsnivåer utgörande en  $n$ -dimensionell hyperkub. Härav beteckningen 'shortest route' dynamisk programmering. I anslutning till problemformuleringen skall i nästa avsnitt några korta kommentarer göras avseende beräkningsmässiga aspekter.

### 5.5 Beräkningsmässiga aspekter

Värdering av sluttillståndet:  $C(z^h, \pi^*, T)$ .

I uttrycket för  $b_{hh}$  ingår en värdering av kostnaden för att vid tidshorisonten  $T$  kapacitetsvektorn utgöres av  $z^h$ . Uppskattning av denna kostnad kan göras på flera alternativa sätt:

- i. Beräkna den diskonterade transportkostnaden, som induceras av slutkapaciteten  $z^h$ :  $C(z^h, \pi^*, T) = \int_0^{\infty} \beta(t+T) P(z^h, T+t) e^{-rt} dt$
- ii. Koppla ihop problemet med ändlig tidshorisont med ett stationärt problem i oändlig tid.
- iii. Sätt låga värden på 'bra' sluttillstånd och höga kostnader på icke önskade slutkapaciteter:

$$C(z^h, \pi^*, T) = +\infty \quad \text{om } z^h \text{ förbjudet sluttillstånd}$$

$$C(z^h, \pi^*, T) = 0 \quad \text{om } z^h \text{ bra sluttillstånd i den meningen att } z^h \text{ förväntas ge låga framtida kostnader.}$$

### Krav på minneskapacitet

För att kunna lösa problem av typ dynamisk programmering med datamaskin måste vissa krav på minneskapacitet hos datamaskinen vara uppfyllda. Nivån på dessa krav bestäms av problemstorleken. Behovet av minnesutrymme är av två slag:

1. Behov av snabbminne (kärnminne) för att lagra kostnaden för redan genomlöpta kapacitetsnivåer:  $C(z^k, \pi^*, 0)$  för  $k \in B_h$  enligt rekursionsformeln.
2. Behov av långsamt minne för att lagra optimala expansionsbeslut i redan genomlöpta kapacitetsnivåer.

Av dessa är kravet på utrymme i snabbminnet begränsande. Utrymmesbehovet bestäms av tillståndsrummets storlek. Denna reduceras ej genom 'shortest route'-formuleringen av problemet i jämförelse med vanlig DP-formulering. Man har sålunda inte kommit ifrån den kanske allvarligaste praktiska begränsning som vidlåder dynamisk programmering som metod. Som exempel kan nämnas, att vid lösning med en 32 K IBM 7090 får tillståndsrummet högst innehålla 4000 tillstånd motsvarande 3 lokaliseringsorter med ca 15 kapacitetsnivåer eller 4 platser med ca 8 kapacitetsnivåer. Om vi endast tillåter kapacitetsexpansioner, som ligger inom vissa givna gränser, kan kravet på kärnminneskapacitet reduceras.

### Tidsåtgång, beräkningsmängd

Genom att välja 'shortest route'-formuleringen av dynamisk programmering och utföra rekursionen i initialt optimala beslut från de olika kapacitetsnivåerna kan beräkningsmängden kraftigt reduceras och därmed tidsåtgången. Man kan visa att antalet beräkningar av högerledet i rekursionsekvationen och jämförelser mellan sådana värden kraftigt minskar jämfört med vanlig dynamisk programmering.

### EXEMPEL

Vi skall jämföra summan av antalet beräkningar av högerled i rekursionsekvationen och antalet jämförelser mellan dessa för dels 'shortest route' dynamisk programmering med 10 kapacitetsnivåer per lokaliseringsort (kontinuerlig tid) och dels vanlig dynamisk programmering med 10 kapacitetsnivåer/lokaliseringsort och 10

tidssteg. Antalet lokaliseringssorter varierar från en till fem:

Antalet platser	'Shortest route'-DP	Vanligt DP
1	$5.5 \cdot 10^1$	$55.0 \cdot 10^1$
2	$10.0 \cdot 10^2$	$302.5 \cdot 10^2$
3	$14.5 \cdot 10^3$	$1663.8 \cdot 10^3$
4	$19.0 \cdot 10^4$	$9150.6 \cdot 10^4$
5	$23.5 \cdot 10^5$	$50328.4 \cdot 10^5$

Man noterar, att beräkningsmängden reduceras med en faktor 10 vid en lokaliseringsplats, och att denna reduktion ökar till en faktor 2000 vid fem platser. Reduktionen beror på att antalet kombinationer av expansionsbeslut för varje kapacitetsvektor är additivt beroende av antalet lokaliseringssorter i 'shortest route'-formuleringen, medan beroendet är multiplikativt i vanlig dynamisk programmering.

Den kraftiga reduktionen av beräkningsmängden vid rekursionen kan tyckas motverkas av att 'shortest route' kräver parametrisk lösning av transportproblemet. Sannolikt är det dock ekonomiskt att en gång för alla lösa transportproblemet parametriskt jämfört med att lösa det upprepade gånger period för period i vanlig DP.

#### 5.6 Enkelt exempel på icke-stationära problemet i ändlig tid

Som exempel på den hittills behandlade icke-stationära modellen i ändlig tid skall vi betrakta det nästan triviala problemet att bestämma kapacitetsutbyggnaden för fallet med en lokaliseringsort. En förutsättning är att inga kapacitetsunderskott får uppstå. En anledning att studera detta exempel är att jämföra resultatet med motsvarande resultat av en dynamisk programmeringsmodell i diskret tid och kontinuerlig expansionsstorlek ( $w$ ) utvecklad av Manne-Veinott, se Manne (1967).

Efterfråge- och investeringskostnadssamband enligt FIG. 5.1 och 5.2 antas gälla.

Kapacitetsnivåerna är 1.0, 2.0, ....., 6.0 ( $z^k$ ,  $k=1, \dots, 6$ ). Diskonteringsfaktor:  $e^{-r} = 0.8$ . Genom att inga kapacitetsunder-

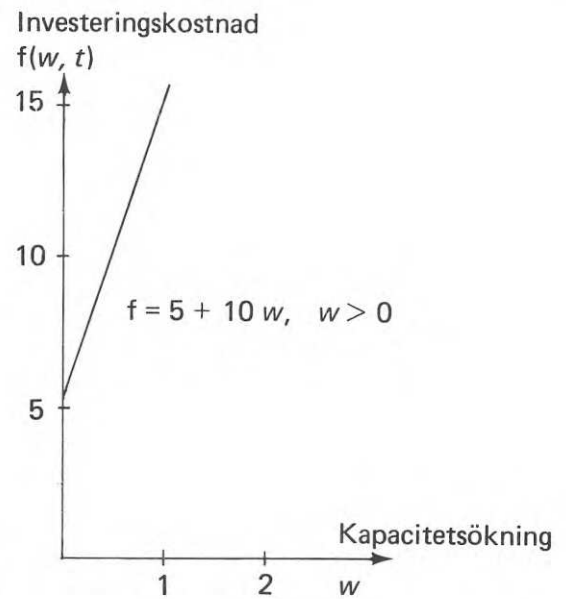
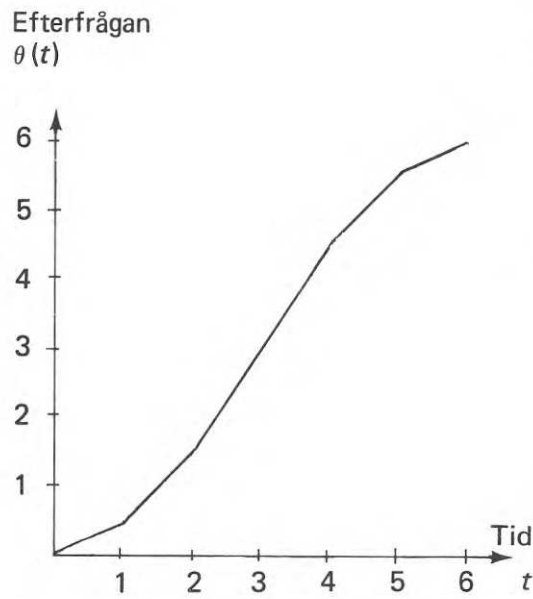


FIG. 5.1 Efterfrågesamband.

FIG. 5.2 Investeringskostnadssamband.

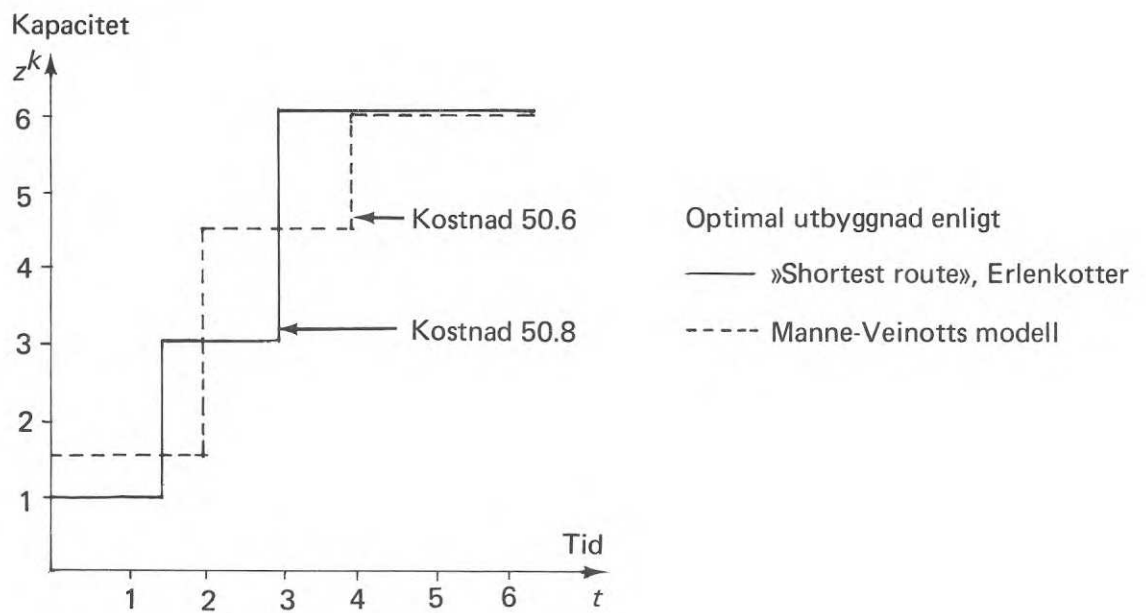


FIG. 5.3 Optimal utbyggnad enligt Erlenkotters och Manne-Veinotts modeller.



skott tillåts, underlättas beräkningen av tidpunkten för initialt optimala expansionsbeslut. Man får

$$b_{hk} = e^{-r\tau_k^*} f(w^k, \tau_k^*)$$

där  $\tau_k^*$  är den tidpunkt då efterfrågan balanserar existerande kapacitet  $z^h$  (se efterfrågesambandet i FIG. 5.1,  $z^h = \theta(t)$ ,  $\tau_k^* = t$ ):

$z^h$ :	1.0	2.0	3.0	4.0	5.0	6.0
$\tau_k^*$	1.50	2.33	3.00	3.67	4.50	6.00

Vidare definierar vi 
$$b_{hh} = \begin{cases} 0 & h \geq 6 \\ +\infty & h < 6 \end{cases}$$

Med hjälp av ovan definierade storheter kan rekursionsekvation (5.6) tillämpas.

$$C(z^6, \pi^*, 0) = b_{66} = 0$$

$$\begin{aligned} C(z^5, \pi^*, 0) &= \min_{5 < k \leq 6} \{ [b_{5k} + C(z^k, \pi^*, 0)] , b_{55} \} = \\ &= \min \{ [b_{56} + C(z^6, \pi^*, 0)] , +\infty \} = 5.5 + 0 = 5.5 \end{aligned}$$

osv

‘  
‘  
‘  
‘

Vi får den optimala policyn:

1. Enligt 'shortest route', Erlenkotter.  
 $\pi_1^* = \{(1.0, 0.0), (2.0, 1.5), (3.0, 3.0)\}$
2. Enligt Manne-Veinotts modell.  
 $\pi_2^* = \{(1.5, 0.0), (3.0, 2.0), (1.5, 4.0)\}$

Bägge modellerna åskådliggörs i FIG. 5.3. Minsta kostnad med de båda modellerna blir 50.8 respektive 50.6.

Med hänsyn till att 'shortest route' opererar med kontinuerlig tid i motsats till diskret tid i Manne-Veinotts modell och att kapacitetsnivåerna inte överensstämmer ger modellerna så

lika initialbeslut som överhuvudtaget är möjligt. Även kostnaderna är mycket lika. Skillnaden består i den optimala policyns utseende under senare delen av planeringsperioden ( $t \geq 2$ ).

### 5.7 Stationär planeringsmodell med oändlig tidshorisont

#### Teori

Hittills har framställningen helt ägnats åt det icke-stationära problemet att planera kapacitetsutbyggnader under en ändlig tidsperiod. I fortsättningen skall motsvarande stationära problem med oändlig tidshorisont något beröras. Huvudsakligen kommer därvid några exempel att behandlas men först skall vi mot bakgrund av den tidigare beskrivna icke-stationära modellen skissera något av teorin för den stationära motsvarigheten. Det stationära planeringsproblemet antages ha följande egenskaper:

- a) Investeringskostnader, transportkostnader och tillverkningskostnader är tidsberoende. Efterfrågefunktionen växer linjärt i tiden: Efterfrågeökning/år på plats  $j$  betecknas  $D_j$ .
- b) De kombinerade transport- och tillverkningskostnaderna måste uppfylla avståndsegenskaperna:
 
$$p_{ij} \geq 0, p_{kk} = 0, p_{ij} \leq p_{ik} + p_{kj}$$
- c) Ändlig övre gräns för skillnaden mellan överskottskapaciteter på olika platser.

Modellen för stationära fallet har som tillståndsvektor överskottskapaciteterna på de olika platserna (jämför icke-stationära modellen, där tillstånden utgjordes av absoluta kapaciteter). Tidsdimensionen undertrycks på samma sätt som tidigare genom att lösa timing-problemet för varje expansion och normera kostnaderna till en initialtidpunkt. De beräkningsmässiga fördelarna med den icke-stationära modellen bevaras i vår utvidgning till det stationära fallet. Ytterligare en fördel tillkommer: tillståndsrummet kan i den stationära modellen för  $n$  st lokaliseringssorter reduceras så att det motsvarar ett icke-stationärt problem med  $n-1$  orter. Detta på grund av att tillståndsrummet endast behöver betraktas vid tidpunkter då minst en av överskottskapaciteterna är noll.

Vi skall visa hur rekursionsekvationen för den stationära modellen kan härledas genom att utgå från den icke-stationära modellen, överföra dennas formulering på problemet med oändlig tidshorisont och omformulera rekursionsekvationen i termer av överskottskapaciteter. Vi kan med små ändringar överföra vår tidigare rekursionsekvation för det ändliga icke-stationära fallet till det stationära planeringsproblemet i oändlig tid:

$$C(z^h, \pi^*, 0) = \min_{k \in B_h} \{b_{hk} + C(z^k, \pi^*, 0), b_{hh}\}$$

där

$$b_{hk} = \int_0^{\tau_k^*} [P(z^h, t) - P(z^k, t)] e^{-rt} dt + e^{-r\tau_k^*} f_{i(k)}^k(w_{i(k)}^k)$$

$$b_{hh} = \int_0^{\infty} P(z^h, t) e^{-rt} dt$$

$B_h$  = mängden (möjligen oändlig) av tillstånd  $z^k$  som kan nå från  $z^h$  via en tillåten expansion i en komponent. Observera att frånvaron av restriktioner  $\tau_k^* \leq \tau_{k+1}^*$  här är betydelsefull, då sådana restriktioner skulle kräva kunskap om framtida expansionstidpunkter. Jämför diskussion kring  $B_h$  i icke-stationära fallet.

Den enda ändringen i jämförelse med icke-stationära modellen är att terminalkostnaderna vid beräkning av  $b_{hh}$  utgått. Denna rekursionsekvation kan inte lösas eftersom ett oändligt antal kapacitetsnivåer är möjliga (framåtrekursion) eller eftersom initialkostnader inte kan anges (bakåtrekursion).

Vi vill nu omformulera ovanstående rekursionsekvation med överskottskapaciteter som tillstånd. Definiera överskottskapaciteten  $z_i(t)$  på plats i då kapaciteten där är  $z_i$ :

$z_i(t) = z_i - \theta_i(t)$ . Dessa överskott kan normeras med den årliga efterfrågeökningen på plats i. Samma normering införes för expansionsstorlekar:

$$y_i(t) = \frac{z_i(t)}{D_i}, \quad x_i = \frac{w_i}{D_i} \quad (y_i(t) \text{ och } x_i \text{ får genom normering- en dimensionen tid}).$$

Härigenom kan vi införa den (tidsberoende) övre gränsen för skillnader mellan överskottskapaciteter:

$$|y_i(t) - y_j(t)| \leq M \text{ för alla } i, j$$

Om vid tiden  $t$  tillståndet beskrives av överskotten

$$y(t) = (y_1(t), y_2(t), \dots, y_n(t))$$

där  $y(t) > 0$  så är det ej optimalt att göra någon ytterligare expansion. Detta inses ur ekvationen

$$r f_{i(k)}(w_{i(k)}^k) = P(z^h, \tau_k^*) - P(z^k, \tau_k^*) \quad (\text{jämför ekv. (5.3)})$$

som för stationära fallet anger optimal tidpunkt för expansion från  $z^h$  till  $z^k$ . Om båda representerar överskottskapaciteter på varje plats blir högerledet noll och  $w_{i(k)}^k = 0$ . Detta leder till att vi begränsar oss till att söka optimal strategi utgående från överskottskapaciteter  $y^*$  med minst en komponent lika med noll. Låt  $y^h$  beteckna överskottskapaciteten  $y^h$  med minst en komponent lika med noll. Om vi startar med  $y^h$  får vi ekvationen

$$C(y^h, \pi^*, 0) = \min_{k \in B_h} \{b_{hk} + C(y^k(0), \pi^*, 0), b_{hh}\}$$

Från den resulterande tillståndsvektorn  $y^k(0)$  kommer ingen ytterligare expansion att göras under tiden  $\tau_{hk} = \min_i \{y_i^k(0)\}$ . Vi får

$$y^k(*) = (y_1^k(0) - \tau_{hk}, y_2^k(0) - \tau_{hk}, \dots, y_n^k(0) - \tau_{hk}) \text{ och}$$

$$C(y^k(0), \pi^*, 0) = e^{-r\tau_{hk}} C(y^k(*), \pi^*, \tau_{hk})$$

P g a stationariteten gäller också:

$$C(y^k(*), \pi^*, \tau_{hk}) = C(y^k(*), \pi^*, 0). \text{ Genom att kombinera våra resultat får vi den slutliga rekursionsekvationen:}$$

$$C(y^h, \pi^*, 0) = \min_{k \in B_h} \{b_{hk} + e^{-r\tau_{hk}} C(y^k(*), \pi^*, 0), b_{hh}\}$$

En lösning till denna rekursionsekvation kan erhållas med sk policyiteration, se Howard (1960). Metoden består av två delar: värdebestämning och förbättring av strategin. En given start-

strategi förutsättes. Utgående från denna löses kostnaderna (värdena)  $C(y^h(*), \pi_h, 0)$  ur ett ekvationssystem. Med hjälp av dessa värden bestäms sedan ny strategi med lägre kostnad än utgångsstrategin. Denna iterationsprocess kan visas konvergera mot en optimal lösning.

### 5.8 Tillämpningar av den stationära modellen i oändlig tid

Vi skall som exempel på tillämpning av den stationära modellen betrakta följande tre uppgifter:

1. Undersökning av kostnadsbesparing till följd av regional samordning av expansionsbeslut.
2. Analys av utseendet hos optimal utbyggnadsstrategi.
3. Bestämning av den förlust som uppstår på grund av icke-optimala initiala beslut.

Dessa analyser kommer att göras med uppkonstruerade data. För att åskådliggöra vad resultaten innebär i ett praktiskt fall kommer i tillämpliga delar att hänvisas till en case-study av Indiens gödselindustri, som ingår i Erlenkotters avhandling.

#### Kostnadsbesparing genom samordning av utbyggnadsbeslut

Orsaker till att vi väljer den stationära modellen som instrument för analysen är att

- den oändliga tidshorisonten eliminerar svårkontrollerbara effekter av en godtycklig ändlig tidshorisont
- det enkla mönstret för efterfrågetillväxten gör att vi slipper välja mellan ett stort antal möjliga icke-stationära efterfrågefunktioner
- problemets dimensionalitet kan reduceras i stationära fallet med motsvarande en lokaliseringsort.

Trots det reducerade tillståndsrummet måste med den datorutrustning som använts analysen begränsas till högst tre orter. Vi förutsätter för analysen att följande idealiserade förhållanden råder:

En oändlig, plan markyta täcks fullständigt av liksidiga trianglar (FIG. 5.4). I varje triangelhorn uppträder en efter-

frågan, som årligen växer likformigt med en enhet. Kostnaderna för att transportera en enhet mellan närbelägna punkter är lika över hela ytan ( $p_{ij}$ ). Vi vill jämföra kostnaden per lokaliseringsort för följande tre fall:

- a) Planering av optimal utbyggnad i varje plats för sig utan samordning.
- b) Koordinering av planeringen för par av platser.
- c) Koordinering av planeringen i grupper om tre platser.

Investeringskostnaden anges i termer av årligt efterfrågetillskott ( $D$ ).

$$f(y \cdot D) = k \cdot (y \cdot D)^a$$

där  $y$  anger antal årsefterfrågetillskott.

I våra kalkyler är  $k = 1.0$  och  $a = 0.70$  (se FIG. 5.5). Kalkylräntan antages vara 10%.

I FIG. 5.6 anges kostnaden per lokaliseringsort för lösningen till stationära modellen i de tre fall som beskrivs ovan. Som parameter varieras transportkostnaden  $p_{ij}$ . För jämförelse har medtagits extremfallen då transportkostnaden antingen är oändlig eller noll. Förra fallet innebär att alla transporter är omöjliga och att varje orts utbyggnad måste planeras var för sig. Transportkostnaden noll innebär att hela efterfrågan kan samlas till en enda plats inom varje grupp.

Man observerar, att en väsentlig del av den potentiella kostnadsreduktionen faller på övergången från en till två platser per grupp. Kostnadsminskningen är betydligt mindre vid utvidgningen av en grupp från två till tre platser. I en case-study kring kapacitetsutbyggnad av Indiens gödselindustri har motsvarande beräkning utförts för kostnad/plats som funktion av gruppstorleken. Utan att gå in på det bakomliggande datamaterialet kan nämnas, att kostnadsreduktionen vid övergång från grupper om en till grupper om två platser var ca 9% medan motsvarande reduktion vid övergång från grupper om två platser till grupper om tre platser var ca 6%.

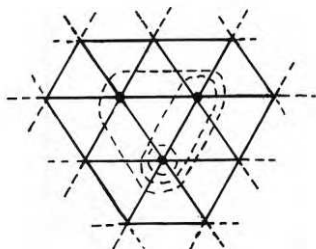


FIG. 5.4 Geografisk struktur.

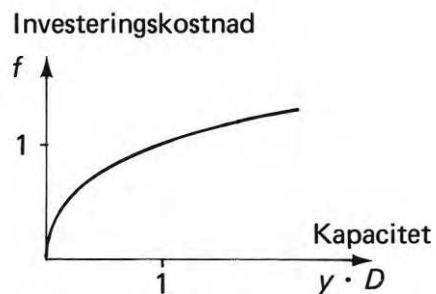


FIG. 5.5 Investeringskostnads-funktion.

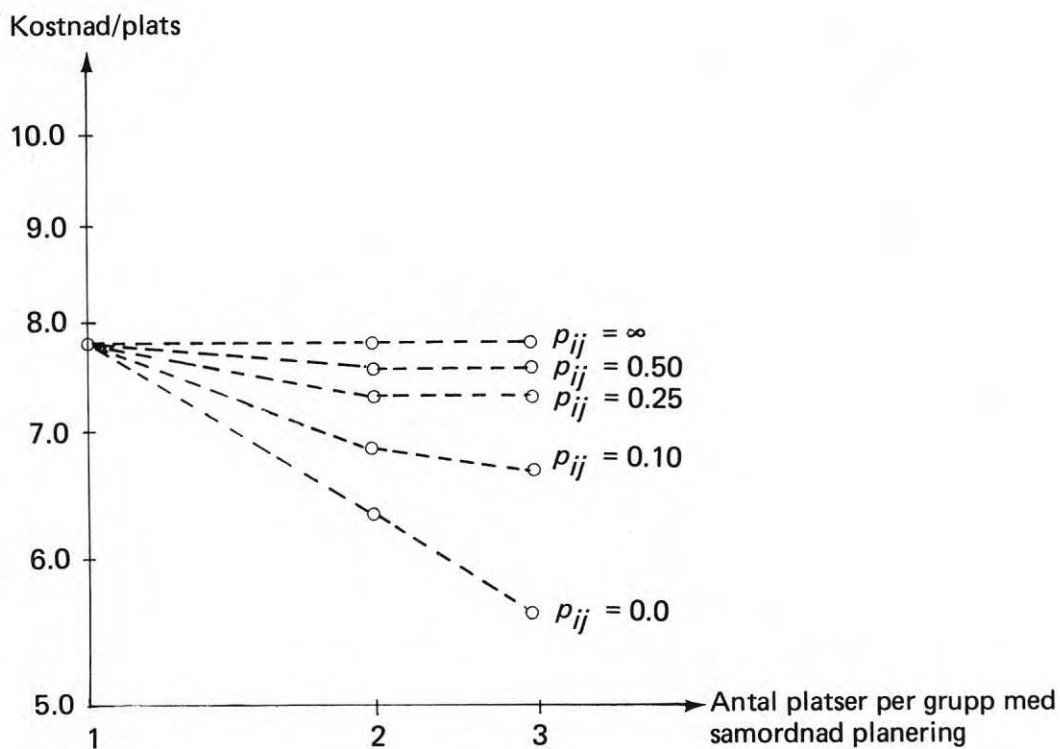


FIG. 5.6 Total kostnad/lokaliseringssort vid olika grad av samplanering.

### Analys av utseendet hos optimal utbyggnadsstrategi

Vi skall i detta avsnitt ge exempel på och något kommentera utseendet hos optimal utbyggnadsstrategi för några olika kombinationer av parametrar. Framställningen begränsas till fallet med två lokaliseringsorter per grupp. Efterfrågeökningen per år är störst i ort 1 där den är 1.0 enheter/år ( $D_1$ ). Motsvarande ökning för ort 2 tillåts variera ( $D_2$ ). Transportkostnaden är lika så variabel medan övriga förutsättningar är samma som i föregående avsnitt.

Vi åskådliggör strategiernas utseende i fyra tillståndsdia-gram, FIG. 5.7. Betrakta övre vänstra diagrammet i FIG. 5.7. Från ett utgångsläge med överskottskapaciteterna  $(y_1, y_2) = (0,0)$  sker omedelbar utbyggnad enligt streckade pilen till  $(6,0)$ . Sedan följer överskottskapaciteterna den heldragna 45-graderslinjen tills linjen  $L$  ( $y_1 D_1 + y_2 D_2 = 0$ ) skärs. Då sker ny omedelbar utbyggnad enligt streckad pil till  $(3,8)$ . Därifrån följer överskottskapaciteterna under en tidsperiod den heldragna 45-graderslinjen (tills linjen  $L$  nås). Sedan sker ytterligare utbyggnad (utgör tio-årsefterfrågetillskottet på plats 1) till  $(7.5, 2.5)$ . Den stationära cykeln med växelvis utbyggnad av respektive tio-årsefterfrågetillskott på plats 1  $(-2.5, 2.5) \rightarrow (7.5, 2.5)$  och plats 2  $(2.5, -2.5) \rightarrow (2.5, 7.5)$  har nu nåtts. Hela tiden måste  $(y_1, y_2)$  befinna sig i tillåtet område, dvs kapacitetsunderskott i ena orten får högst vara av samma storlek som överskott i andra orten ( $y_1 D_1 + y_2 D_2 \geq 0$ ). På varje streckad pil anges utbyggnadsstorleken i antal årsefterfrågetillskott på den plats där utbyggnad sker.

Med anledning av dessa diagram skall några korta kommentarer göras:

1. Efter en relativt kort transientfas konvergerar strategierna mot respektive stationära cykel.
2. I de flesta fall är den slutliga stationära cykeln av samma längd för båda lokaliseringsorterna. Om totala efterfrågeökningen är mycket skevt fördelad och transportkostnaderna är låga (fallet  $p_{ij} = 0.10, D_2 = 0.25$ ) kan dock cykellängderna bli olika långa. I det nedre vänstra diagrammet ser vi att den stationära cykeln består av två (små) utbyggnader på ort 1 (8 enheter vardera) under samma tid som det sker en (stor) ut-



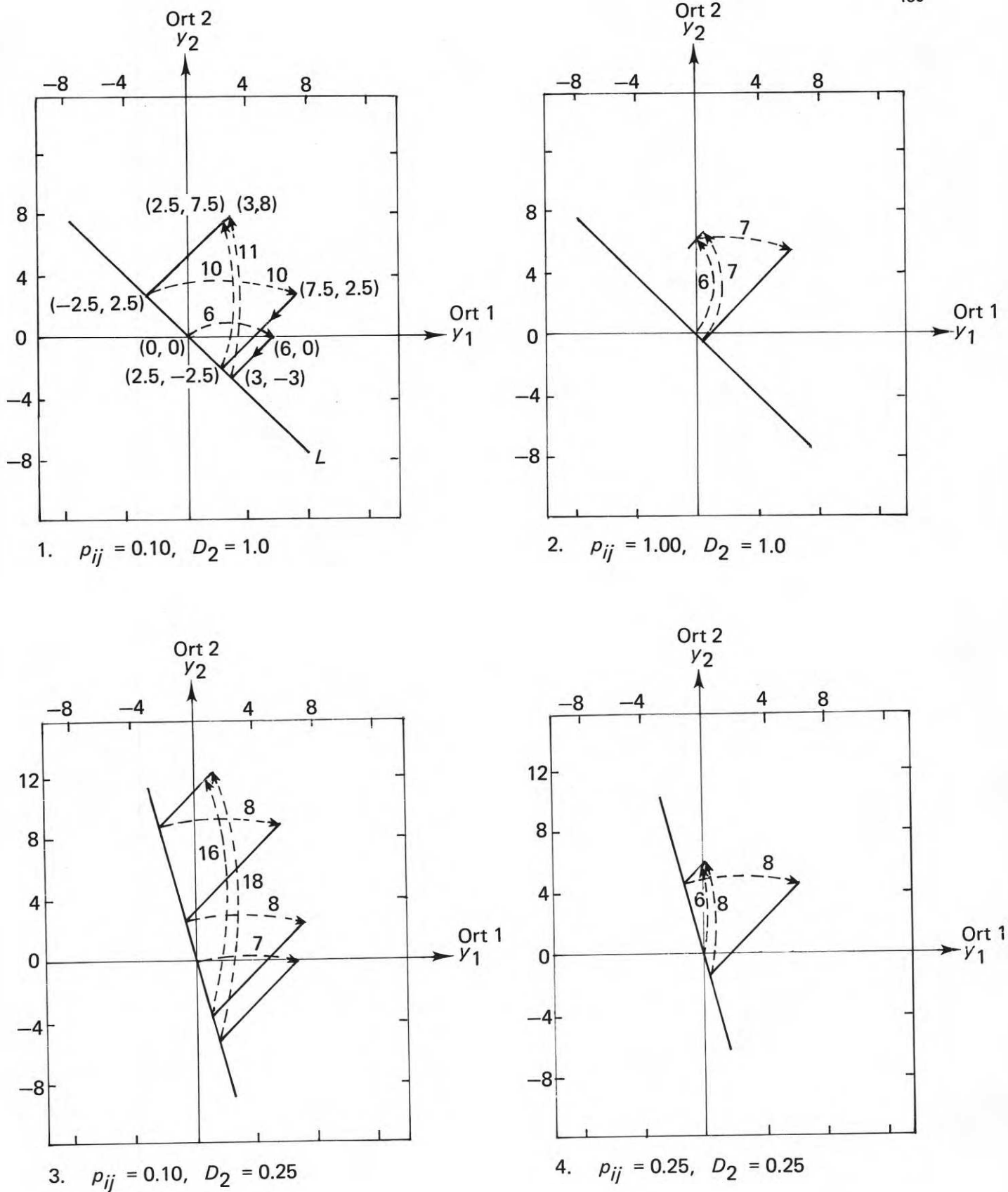


FIG. 5.7 Optimala utbyggnadsstrategier från utgångstillståndet (0,0) för olika kombinationer av transportkostnad och efterfrågestruktur. Koordinaterna anger överskottskapaciteter i respektive ort. På de streckade pilarna anges utbyggnadens storlek.

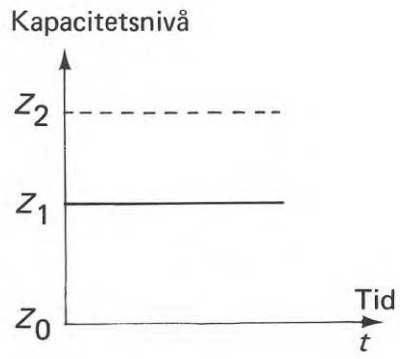


FIG. 5.8 Optimalt ( $z_1$ ) respektive initialt icke-optimalt ( $z_2$ ) beslut.

- byggnad på ort 2 (16 enheter).
3. Observera att lokaliseringen av första utbyggnaden växlar beroende på transportkostnadernas storlek.

Graden av samordning av utbyggnadsbeslut i olika orter påverkar storleken av optimal utbyggnad. Större koordinering tenderar att öka utbyggnadsstorleken, men ökningen är mindre uttalad vid höga transportkostnader. Som exempel på längden av transientfas respektive slutlig cykel betraktar vi åter planeringen av optimal utbyggnad av Indiens gödselindustri. Följande tabell visar längderna (i antal år) för tre grupper om vardera tre orter för två olika övre gränser på kapacitetsutbyggnaden.

	Grupp 1	Grupp 2	Grupp 3
Övre gräns 250.000 ton			
Transientfas (antal år)	37,2	0,0	20,1
Cykellängd (antal år)	135,0	36,0	24,0
Övre gräns 400.000 ton			
Transientfas (antal år)	30,3	61,6	33,7
Cykellängd (antal år)	56,0	30,0	25,0

För fallet med övre gränsen 250.000 ton fås en mycket jämn fördelning av utbyggnadsstorlek mellan grupperna. Produktionsenheterna blir större och av mera enhetlig storlek då grupperna växer från en till tre orter. Motsvarande slutsats gäller då övre gränsen är 400.000 ton även om spridningen på olika storlekar här är något större.

#### Bestämning av förlust på grund av icke-optimala beslut

Med dynamisk programmering erhåller vi optimal kostnad för varje initial kapacitetsnivå. Denna information kan vi utnyttja för att beräkna extra kostnad av ett icke-optimalt beslut ifrån utgångsläget.

- $z_0$  : utgångsläge
- $z_0 \rightarrow z_1$  : optimal expansion från utgångsläget  $z_0$

$z_0 \rightarrow z_2$  : icke-optimal expansionsbeslut från utgångsläget  
 $C(z_2, \pi^*, 0)$ : kostnad för optimal strategi utgående från  $z_2$   
 $C(z_0, \pi^*, 0)$ : kostnad för optimal strategi utgående från  $z_1$   
 $f(z_0, z_2)$  : kostnad för utbyggnad från  $z_0$  till  $z_2$ .

Extrakostnaden för att i utgångsläget fatta det icke-optimala utbyggnadsbeslutet  $z_2$  kan nu definieras:

$$\frac{f(z_0, z_2) - (C(z_0, \pi^*, 0) - C(z_2, \pi^*, 0))}{C(z_0, \pi^*, 0) - C(z_2, \pi^*, 0)}$$

I FIG. 5.9 och 5.10 har detta estimat bestämts för ett par olika parameterkombinationer i fallet med två orter per grupp. Vi ser att om fel ort väljes för första utbyggnaden fås en extrakostnad på åtminstone 3-4%. Kring optimalt beslut kan en variation på ca 20% ske utan att kostnaden ökar med mera än 1%. (Observera att det initialt optimala beslutet för motsvarande problem i FIG. 5.7, diagram 3 och 4, anges i antal årsefterfrågetillskott.)

### 5.9 Sammanfattning

I denna översikt av innehållet i Erlenkotters avhandling har vi behandlat problemet att finna en optimal utbyggnadsstrategi för en en-produktindustri. Metoden är dock användbar för en klass av tillämpningar. Det viktiga är att kostnadsfunktionen  $P(z, t)$  uppfyller vissa kritiska egenskaper (avsnitt 2). Exempel på problem där dessa egenskaper är uppfyllda kan t ex vara vissa lagerproblem eller planering av utbyggnad av broar över ett vattendrag. Allmänt gäller att metoden är tillämpbar på problem med både skalfördelar vid investering och beroende mellan de olika projektens ekonomiska utfall. En avgjord begränsning hos Erlenkotters metod är att den ej kan behandla problem med external economies.

Extra kostnad för icke-optimalt beslut

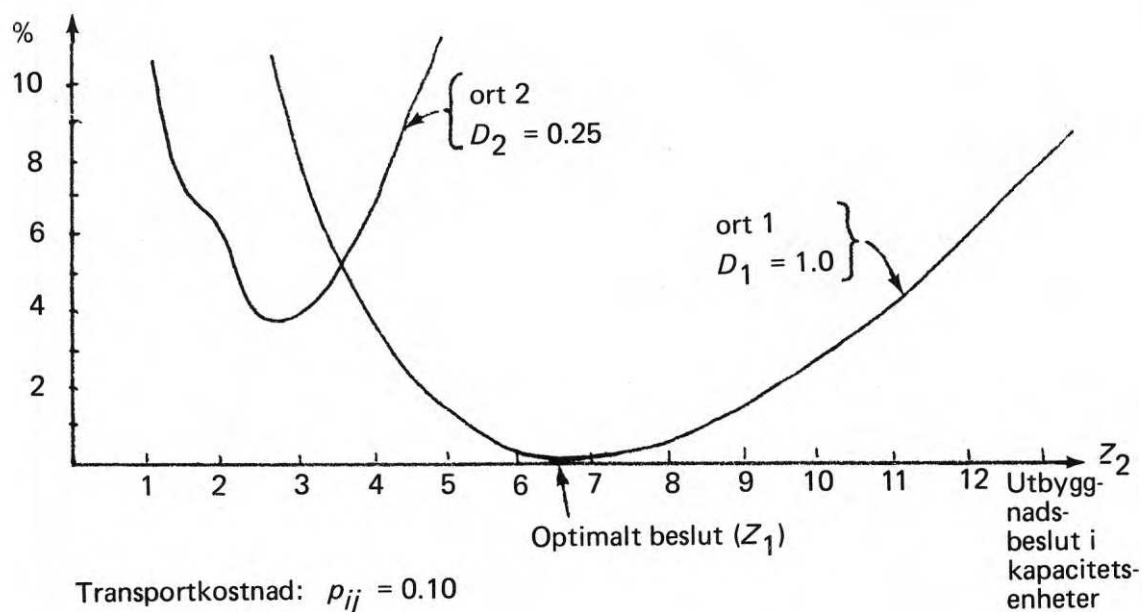


FIG. 5.9 Extrakostnad vid initialt icke-optimalt beslut (jämför FIG. 5.7 diagram 3).

Extra kostnad för icke-optimalt beslut

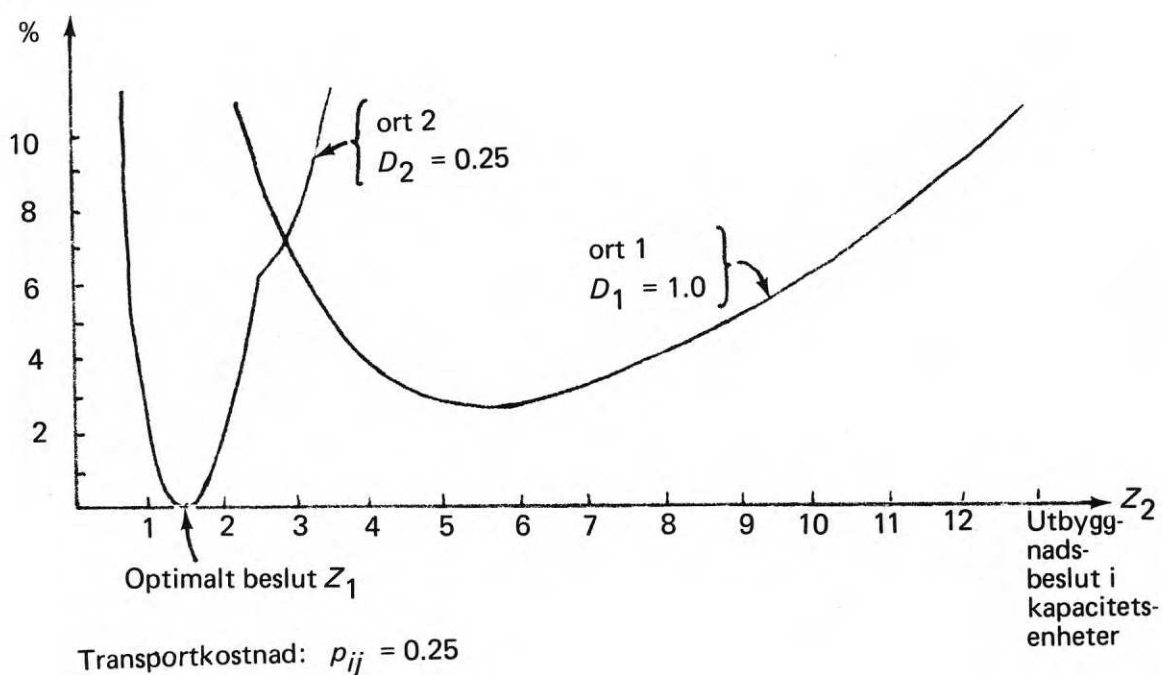


FIG. 5.10 Extrakostnad vid initialt icke-optimalt beslut (jämför FIG. 5.7 diagram 4).

5.10 Litteraturreferenser

Bos, H C, 1965, *Spatial dispersion of economic activity*. (Rotterdam University Press) Rotterdam.

Dantzig, G B, 1963, *Linear Programming and Extensions*. (Princeton University Press) Princeton.

Erlenkotter, D, 1969, *Preinvestment Planning for Capacity Expansion: A Multi-Location Dynamic Model*. (Stanford University).

Howard, R, 1960, *Dynamic Programming and Markov Processes*. (John Wiley and Sons Inc.) New York.

Manne, A S, 1967, *Investments for Capacity Expansion: Size, Location and Time-Phasing*. (George Allen and Unwin Ltd) London.

6 HELTALSPROGRAMMERINGSMODELLER FÖR DYNAMISK LOKALISERING  
Anders Holvid

6.1 Problemformulering

I modeller, som avser att belysa de kombinatoriska aspekterna av ett lokaliseringsproblem, är heltalsprogrammering ett lämpligt hjälpmedel.

Ett välkänt problem är det s k simple plant location problem. Ett antal potentiella produktionsenheter, karakteriserade av en fast anläggningskostnad och en rörlig produktionskostnad, är givna. Dessa skall försörja ett antal konsumtionsenheter, vilka har en specificerad transportkostnad från de olika produktionsenheterna. Problemet är att söka den kombination av produktionsenheter, som ger minsta totala kostnad (dvs minsta summa av fasta anläggningskostnader, produktionskostnader och transportkostnader). Matematiskt formuleras problemet:

Minimera

$$z = \sum_{ij} c_{ij} x_{ij} + \sum_i f_i y_i$$

$$\sum_i x_{ij} = 1 \quad (j = 1, \dots, n)$$

$$x_{ij} \leq y_i \quad (i = 1, \dots, m, j = 1, \dots, n)$$

$$x_{ij} \geq 0$$

$$y_i = 0 \text{ eller } 1$$

Här är

$m$  = antal produktionsenheter

$n$  = antal konsumtionsenheter

$c_{ij}$  =  $t_{ij} \cdot D_j$

$D_j$  = behov hos konsumtionsenhet  $j$

$t_{ij}$  = generaliserad transportkostnad från produktionsenhet  $i$  till konsumtionsenhet  $j$ , vilken är lika med verklig transportkostnad plus rörlig produktionskostnad i produktionsenhet  $i$ .

- $x_{ij}$  = andel av konsumtionsenhet  $j$ 's behov som tas från produktionsenhet  $i$ .  
 $f_i$  = fast kostnad i produktionsenhet  $i$ .  
 $y_i$  = logisk variabel lika med 1 om produktionsenhet  $i$  byggs, lika med 0 annars.

För ytterligare beskrivning, se Efroymsen (1966). Denna modell kan kritiseras på många sätt. Produktionsenheterna tillåts exempelvis att bli hur stora som helst, dvs det finns ingen övre gräns på

$$\sum_j x_{ij} \cdot D_j \quad i = 1, 2, \dots, m$$

Den viktigaste inskränkningen är dock den statiska situationen. I allmänhet är ju förutsättningarna av följande typ.

Vid nu-tidpunkten är en viss produktionsstruktur given. Den är avsedd att möta den rådande konsumtionsstrukturen. Man har dock prognoser för hur konsumtionsstrukturen kommer att ändras i framtiden. Problemet är att på bästa sätt förändra och anpassa produktionsstrukturen med utgångspunkt från dessa prognoser. Det kan här vara fråga om såväl nybyggnad av produktionsenheter som utbyggnad.

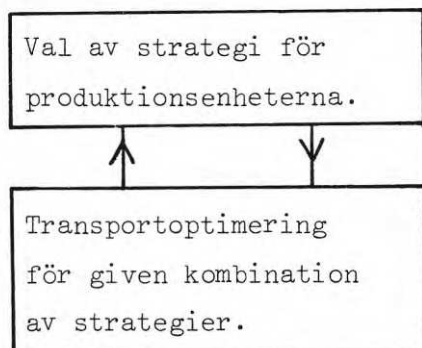
## 6.2 Modellförslag

Det ligger här nära till hands att söka utvidga modellen i "simple plant location problem" till flera perioder. Man råkar då i stora svårigheter då det gäller att uppskatta kostnaderna för produktionsenheterna. En enkel och flexibel modell fås om problemet i stället formuleras som en strategivalmodell. Detta innebär en omformulering av det flerperiodiga "simple plant location problem", vilken är av samma typ som den som göres av det ursprungliga linjärprogrammeringsproblemet vid Dantzig-Wolfedekomposition. I modellen är strategibegreppet viktigt:

En strategi för en existerande eller potentiell (den behöver inte vara byggd vid nu-tidpunkten) produktionsenhet är en fullständig beskrivning av framtida åtgärder i form av utbyggnader, nedläggning etc. Varje strategi karakteriseras av en viss kostnad (för en given strategi är den möjlig att uppskatta) samt krav och



utbud av olika resurser. Problemet är att söka den kombination av strategier för samtliga produktionsenheter som minimerar eller maximerar någon kriterievariabel. Detta är ett blandat heltalsproblem vilket består av två delar, kopplade till varandra enligt figuren nedan.



#### Strategivalgssteget

Detta blir ett heltalsprogrammeringsproblem med nedanstående systemmatris (se FIG. 6.1).

Variablerna är logiska variabler där värdet 1 svarar mot att denna strategi väljs och värdet 0 mot att den inte väljs. Resultatvariablerna sammanfattar de respektive strategiernas resurskrav och resursutbud. Vissa av resultatvariablerna är kopplade till transportoptimeringssteget, medan andra skall uppfylla restriktioner av olika typer. Det enklaste fallet är att  $r_i$  skall ligga mellan två gränser:

$$r_i' \leq r_i \leq r_i'' \quad i = 1, 2, \dots, p$$

Resultatvariablerna kan alltså vara kostnader under någon tidsperiod, totala diskonterade kostnader, utbud av produktionskapacitet under olika perioder etc.

#### Transportoptimeringssteget

Detta utgår alltså från en given produktionsstruktur, där det gäller att tillfredsställa konsumtionsenheternas behov på bästa sätt. Denna del kan utformas på flera sätt. Enklast som "enkelt" transportproblem", mer detaljerat som flödesproblem.

Hittills har inget nämnts om generering av strategier. Modellen kan härvid tänkas använd med två olika ambitionsnivåer.

Strategi 1  
Produktions-  
enhet 1

$a_{11}^1$ $a_{21}^1$	$a_{12}^1$ .....	$a_{11}^2$ $a_{12}^2$ .....	.....	$a_{11}^m$ $a_{12}^m$ ...	$= r_1$ $= r_2$ $\vdots$ $= r_p$	resultat- variabler
	$a_{22}^1$ .....	$a_{21}^2$ $a_{22}^2$ .....	.....	$a_{21}^m$		
1	1 ..... 1				= 1	grupprest- riktioner
		1    1 ..... 1			= 1	
				1    1 .... 1	= 1	
Prod.enhet 1		Prod.enhet 2		Prod.enhet m		

FIG. 6.1 Systemmatris strategivalssteget.

Den första innebär utvärdering av ett antal givna utbyggnadsalternativ. Exempelvis kan produktionsenheterna vara massa-industrier tillhörande olika företag. För varje fabrik diskuteras ett antal olika möjliga utbyggnadsalternativ. Ett överordnat planeringsorgan ställs här inför problemet att välja kombination av utbyggnadsalternativ med hänsyn till råvaran - skogens utveckling. Den andra högre ambitionsnivån innebär att söka optimala strategier. Det blir här fråga om någon form av systematisk generering av dessa. Detta kan tänkas ske enligt någon heuristisk metod eller med en bakomliggande teoribyggning såsom i det sk trimproblemet.

### Lösningsmetoder

Generellt är problemet ett blandat heltalsprogrammeringsproblem av nedanstående typ:

minimera

$$z = c^T x + d^T y$$

då

$$A_1 x + A_2 y = b$$

$$x_i = 0 \text{ eller } 1, \quad i = 1, \dots, n_1$$

$$y_j \geq 0 \quad j = 1, \dots, n_2$$

De lösningsmetoder, som behandlar generella strukturer av ovanstående typ kräver för ett problem med t ex ett par hundra variabler relativt långa beräkningstider i dator. Här är dock en utvidgning av Geoffrions metod med "surrogate constraints", Geoffrion (1967) säkert värd att testa. Det kan även löna sig att undersöka huruvida specialmetoder, t ex byggda på någon form av dekomposition är användbara. Här kan nämnas exempelvis Benders' metod. I denna har man möjlighet att utnyttja "transportstrukturen" i den kontinuerliga delen av problemet.

6.3 Litteraturreferenser

Efroymsen, M A & Ray, T L, 1966, *A Branch-bound Algorithm for Plant Location*. Op. res. 14, p 361-368, 1966.

Geoffrion, A M, 1967, *Implicit Enumeration Using an Embedded Linear Program*. (U.S.A.F.). Memorandum RM-5406-Pr.

7 LOKALISERINGSANALYS FÖR STOCKHOLMSREGIONEN<sup>1</sup>

Åke E. Andersson

7.1 Lokaliseringsteorin som en del av teorin för allmän ekonomisk jämvikt

Ekonomer har nästan alltid en teori för allmän jämvikt som bakgrund även vid praktiskt orienterade analyser. Förhoppningsvis präglar detta också den uppläggning av lokaliseringsanalyser som här skall presenteras. Men även om en teori för allmän jämvikt bildar bakgrund är det inte säkert att en planeringsorienterad lokaliseringsanalys kan genomföras med ständigt beaktande av alla ömsesidiga beroenden i det sociala eller ens ekonomiska systemet.

Lefebvre var den som först formulerade en helt igenom konsistent allmän jämviktsteori för konsumtions-, produktions-, lokaliserings- och transportsystemet (se Lefebvre (1968)). I Lefebvres modell diskuteras dock inte ömsesidiga beroenden mellan företag. I Andersson (1970) återges i slutavsnittet Lefebvres allmänna modell. Lefebvres ansats skall här bara antydans som inledning till beskrivningen av en modell som praktiskt kan användas för analys av pendlingsrörelser i en storstadsregion.

Det finns två lokaliseringpunkter, läge 1 och läge 2. Det finns tre produkter, produkt 1 som lokaliserats till läge 1, produkt 2 som lokaliserats till läge 2, och produkt 3, transporttjänster, som sker mellan läge 1 och läge 2. Produkterna betecknas:

$x^1$  = produkt 1, som produceras i läge 1

$x^2$  = produkt 2, som produceras i läge 2

$x^3$  = transporttjänster

Det finns två produktionsfaktorer. Produktionsfaktorerna kan vara lågutbildad och högutbildad arbetskraft. Faktortjänsterna från dessa produktionsfaktorer antas delbara och är de enda fak-

---

<sup>1</sup> Artikelns har sammanställts under medverkan av Magnus Holm.

torer från vilka faktorinputs kan levereras till valfri produkt och till valfri lokalisering.

$V_{ij}^k$  = leverans av faktortjänster från faktor  $i$  bosatt i läge  $j$  som används för produkt  $k$  i läge  $k$ .

Produktionen i läge 1 beror på hur stor insats som sker av faktortjänsterna från faktor 1 respektive faktor 2, vilka kan hämtas från läge 1 respektive läge 2:

$$x^1 = \varphi^1(V_{11}^1 + V_{12}^1, V_{21}^1 + V_{22}^1) \quad (7.1)$$

På motsvarande sätt finns det en produktionsfunktion för sektor 2:s produktion i läge 2.

$$x^2 = \varphi^2(V_{11}^2 + V_{12}^2, V_{21}^2 + V_{22}^2) \quad (7.2)$$

Transportsektorns produktionsresultat bestäms också av insatsen av produktionsfaktortjänsterna 1 och 2.

$$x^3 = \varphi^3(V_{11}^3 + V_{12}^3, V_{21}^3 + V_{22}^3) \quad (7.3)$$

Efterfrågan på transporttjänster antas bero på de två produktionsfaktorernas pendlingsarbete som är proportionellt mot insatsen av faktortjänster hämtade från annat läge.

$$x^3 = \tau(V_{11}^2, V_{12}^1, V_{21}^2, V_{22}^1) \quad (7.4)$$

För systemet gäller ett villkor för jämvikt, nämligen att efterfrågan på transporttjänster skall svara mot produktionen av transporttjänster.

$$\tau(V_{11}^2, V_{12}^1, V_{21}^2, V_{22}^1) = \varphi^3(V_{11}^3 + V_{12}^3, V_{21}^3 + V_{22}^3) \quad (7.5)$$

(7.6)-(7.9) Bosättnings beroende av produktionsfaktorerna antas fixerad, varför ekvationerna (7.6) till (7.9) kan formuleras som restriktioner på systemet.

Produktionsfunktionerna ekv. (7.1) till (7.3) antas vara konkava. Lefeber är inte i första hand ute efter att undersöka

villkoren för välfärdsmaximering. Det gäller snarare att undersöka hur stor produktion av den ena produkten som maximalt kan uppnås när produktionsresultatet fixeras till en given nivå för den andra produkten och när jämviktsvillkoret 5 gäller. Rörligheten hos produktionsfaktorerna kan vara begränsad, exempelvis så att följande villkor gäller.

$$\begin{aligned} V_{12}^1 &= 0 \\ V_{12}^3 &= 0 \\ V_{22}^1 &= 0 \\ V_{22}^3 &= 0 \end{aligned} \tag{7.10}$$

När dessa rörlighetshinder formulerats är det möjligt att maximera produktionsresultatet i läge 1 samtidigt som produktionsresultatet i läge 2 hålls på en viss nivå och samtidigt som samband (7.5) till och med (7.10) beaktas. Det nödvändiga villkoret för maximering av produktionsresultatet i läge 1 kan sammanfattas på följande sätt:

$$\frac{\frac{\partial \varphi^1}{\partial V_1}}{\frac{\partial \varphi^1}{\partial V_2}} \left[ 1 + \frac{\frac{\partial \tau}{\partial V_{11}}}{\frac{\partial \varphi^3}{\partial V_1}} \right] = \frac{\frac{\partial \varphi^2}{\partial V_1}}{\frac{\partial \varphi^2}{\partial V_2}} \left[ 1 + \frac{\frac{\partial \tau}{\partial V_{21}}}{\frac{\partial \varphi^3}{\partial V_2}} \right] \tag{7.11}$$

( $V_i$  innebär produktionsfaktor  $i$ ,  
 $V_{ij}$  innebär produktionsfaktor  $i$  bosatt i läge  $j$ ).

Om det inte finns någon friktion i systemet säger villkor (7.11) att utbytesrelationen mellan de två produktionsfaktorerna skall vara densamma i båda lokaliseringarna. Om det däremot finns friktioner i systemet, kommer derivatorna från transportsektorns jämviktsvillkor in och korrigerar villkoret för maximum. Begränsningarna på rörligheten i samband (7.10) kan ges många olika utseenden. För vart och ett av dessa går det att härleda nya nödvändiga villkor för maximum och man kan på det sättet successivt få en kurva som anger den maximala produktionen av produkt 1 för alternativa värden på produkt 2. Denna kurva brukar kallas transformationsfunktionen.

Den här återgivna modellen har tagits som utgångspunkt för en undersökning av lönsamheten av ett vidgat pendlingsarbete. Produktionsfunktionerna har antagits vara linjära i sina logaritmer, dvs av Cobb-Douglas-typ. Produktionsfaktorerna är arbetskraft på olika utbildningsnivå. Transportsektorns verksamhet har inte behandlats lika väl som i Lefebers modell. Det gäller i stället att avgöra vilken pendlingsvolym och vilket produktionsresultat som skulle uppstå om det inte fanns några restriktioner på arbetskraftens geografiska rörlighet i regionen.<sup>1)</sup>

## 7.2 Beräkning av överflyttningsvinster vid ökad rörlighet

Avsikten med den analys som presenteras nedan är att med en optimeringsmodell beräkna de rörlighetsvinster man kan göra såväl sektoriellt som regionalt. Modellen kan i huvudsak beskrivas på följande sätt. Man maximerar regionens produktion i de nationella sektorerna under restriktioner på produktionen i de lokala sektorerna och tillgången på olika slags arbetskraft. Med nationella sektorer avses här sådana sektorer vars produktion huvudsakligen avsättes utanför regionen. En lokal sektor avsetter sin produktion inom regionen eller eventuellt t o m inom delar av regionen.

Restriktioner på produktionen i de lokala sektorerna införes för att säkerställa att dessa inte tömmer på arbetskraft vid maximeringsförfarandet. Att en sådan restriktion är nödvändig inses enklast om de lokala sektorerna är servicesektorer som detaljhandeln. Praktiskt säkras produktionen i de lokala sektorerna genom att dessa och den däri ingående arbetskraften aldrig införes i själva maximeringsproblemet. Rörlighetsvinster - sektoriellt och regionalt - belyses genom att man inför olika restriktioner på arbetskraften.

---

<sup>1</sup> Modellen har vidareutvecklats ur en modell som utvecklats och anpassats för empirisk optimeringsanalys av Olle Ohlsson och Arne Granholm. Den sistnämnda modellen har använts för utbildningspolitiska analyser. Magnus Holm och Ulf Strömquist har tillsammans med Olle Ohlsson arbetat med den lokaliseringspolitiska formuleringen. Ett enkelt exempel på modellens användning i transportpolitisk analys ges i Andersson (1969).



Formulering av maximeringsproblemet

Inför först följande beteckningar:

$Q_{ij}$  = produktionen i den  $i$ :te sektorn i den  $j$ :te regiondelen

$i = 1, \dots, n, n + 1, \dots, N$ ;  $j = 1, \dots, J$

där  $1 \leq i \leq n$  betecknar nationella sektorer

och  $n + 1 \leq i \leq N$  betecknar lokala sektorer<sup>1)</sup>

$Q = \sum_i^n \sum_j^J Q_{ij}$  = samlad regional produktion i nationella sektorer

$L_{ijk}$  = arbetskraft av den  $k$ :te kategorin använd i den  $i$ :te sektorn i den  $j$ :te regiondelen.

Inför följande produktionsfunktion av Cobb-Douglastyp:

$$Q_{ij} = A_{ij} \prod_{k=1}^K L_{ijk}^{\alpha_{ijk}}$$

där  $\Pi$  anger produkten.

Maximera:

$$Q = \sum_{i=1}^n \sum_{j=1}^J Q_{ij} \quad (7.12)$$

där

$$Q_{ij} = A_{ij} \prod_{k=1}^K L_{ijk}^{\alpha_{ijk}} \quad \begin{array}{l} i = 1, \dots, n \\ j = 1, \dots, J \end{array} \quad (7.13)$$

Detta innebär att för varje regiondel specificeras de ingående sektorernas respektive produktionsfunktion.

$$\sum_{i=1}^n \sum_{j=1}^J L_{ijk} = \bar{L}_{..k} \quad k = 1, \dots, K \quad (7.14)$$

Restriktionen innebär att man tillåter rörlighet mellan såväl sektorer som regiondelar för varje arbetskraftskategori.

<sup>1</sup> De lokala sektorerna är här endast med för fullständighetens skull. Som redan tidigare nämnts behandlas de icke vid själva maximeringsförfarandet.

Vill man i stället undersöka effekten av ökad rörlighet mellan olika delar av regionen, samtidigt som man utesluter all rörlighet mellan sektorer, kan detta ske genom att man i stället inför en restriktion av följande typ.

$$\sum_j L_{ijk} = \bar{L}_{i \cdot k} \quad \begin{array}{l} i = 1, \dots, n \\ k = 1, \dots, K \end{array} \quad (7.15)$$

Ökad sektorrörlighet inom varje regiondel för de skilda arbetskraftskategorierna ger en produktionsförändring som kan studeras i modellen, om denna i stället tillföres restriktionen:

$$\sum_{i=1}^n L_{ijk} = \bar{L}_{\cdot jk} \quad \begin{array}{l} j = 1, \dots, J \\ k = 1, \dots, K \end{array} \quad (7.16)$$

### Skattningsproblem i och datakrav på modellen

#### Skattning av produktionsfunktionen

Här skall kort beröras den allmänna tekniken vid skattningar av produktionsfunktioner av Cobb-Douglas-typ.

Utgå från följande produktionsfunktion:

$$Q_{ij} = a_{ij} \cdot \prod_m C_{ijm}^{\beta_{ijm}} \cdot \prod_k L_{ijk}^{\alpha_{ijk}} \quad (7.17)$$

$C_{ijm}$  betyder användning av kapital av  $m$ :te kategorin i den  $i$ :te sektorn och  $j$ :te regiondelen.

$Q_{ij}$ ,  $L_{ijk}$  har samma betydelse som tidigare.

Med marginalproduktivtetsanpassad avkastning på såväl arbetskraft som kapital gäller då:

$$W_{ijk} = \frac{\partial Q_{ij}}{\partial L_{ijk}} = \alpha_{ijk} \frac{Q_{ij}}{L_{ijk}} \quad (7.18)$$

$$r_{ijm} = \frac{\partial Q_{ij}}{\partial C_{ijm}} = \beta_{ijm} \frac{Q_{ij}}{C_{ijm}} \quad (7.19)$$

Här är

$W_{ijk}$  = lönen för den k:te arbetskraftskategorin i den i:te sektorn och j:te regiondelen.

$r_{ijm}$  = räntan, eller avkastningen per kapitalenhet för den m:te kapitalkategorin i den i:te sektorn och j:te regiondelen.

Ur ekv. (7.18) och (7.19) löses  $\alpha$  och  $\beta$ .

$$\alpha_{ijk} = \frac{W_{ijk} \cdot L_{ijk}}{Q_{ij}} \quad (7.20)$$

$$\beta_{ijm} = \frac{r_{ijm} \cdot C_{ijm}}{Q_{ij}} \quad (7.21)$$

Härur framgår att för varje arbetskraftskategori, sektor och regiondel kan  $\alpha$  skattas som lönesummans andel av produktionen, dvs vad som normalt kallas produktionens löneandel. På motsvarande sätt kan  $\beta$  bestämmas som produktionens kapitalandel, dvs kapitalavkastningens totala värde i förhållande till produktionen.

#### Modellens datakrav

Det är nu möjligt att formulera följande slutsats.

Produktionsfunktioner av Cobb-Douglas-typ, för varje sektor och regiondel, kan skattas om man känner produktionsresultatet och dess fördelning på löner och kapitalavkastning. Normalt innebär detta att man bör känna till de förut definierade storheterna:

$Q_{ij}$  = produktionsresultat per tidsperiod

$L_{ijk}$  = arbetskraftsanvändning per tidsperiod

$W_{ijk}$  = arbetslönen

$C_{ijm}$  = kapitalanvändning per tidsperiod

$r_{ijm}$  = räntan (kapitalavkastning per använd kapitalenhet).

#### Tillgängliga data

För bearbetning av modellen finns följande data till hands:

$Q_i$  = Produktionen fördelad på sektorer för hela regionen.

$L_{ijk}$  = Arbetskraftens fördelning på sektorer, regiondelar och arbetskraftskategorier.

$W_{ijk}$  = Genomsnittslönens fördelning på sektorer, regiondelar och arbetskraftskategorier.

Data har erhållits genom intervjuer med ett urval ur befolkningen över 14 år om ca 10.000 personer.

#### Skattningsproblem orsakade av brist på data

Jämför man tillgången på data med de datakrav som modellen ställer inses omedelbart att datatillgången är otillräcklig för en fullständig behandling av denna.

För att kunna genomföra en analys med modellen krävs att denna skattas under vissa speciella antaganden. Det är då uppenbart att innebörden och konsekvensen av dessa antaganden måste studeras ingående.

Bristen på data medför att följande två problem måste lösas:

1. Produktionen är, för varje sektor, endast känd för hela regionen och måste således, på något sätt, fördelas på regiondelar.
2. Det saknas data angående kapitalets roll i produktionen. Man måste således kringgå denna informationsbrist och försöka skatta produktionsfunktionen utan explicita kapitalargument.

Problemet under punkt 1 angripes på det sättet att den totala sektorproduktionen ( $Q_i$ ) fördelas på regiondelarna i proportion till dessas andel av sektorns samlade lönesumma. Detta kan tyckas vara en godtycklig fördelningsprincip, varför det är angeläget att något närmare granska dess innebörd.

I själva verket gör man följande påstående:

$$Q_{ij} = \frac{\sum_k W_{ijk} L_{ijk}}{\sum_j \sum_k W_{ijk} L_{ijk}} \cdot Q_i \quad (7.22)$$

där

$$Q_i = \sum_j Q_{ij} \quad (7.23)$$

Under förutsättning att produktionsfunktionerna är homogena av 1:a grad, dvs att

$$\sum_m \beta_{ijm} + \sum_k \alpha_{ijk} = 1$$

gäller det att

$$Q_{ij} = \sum_m \frac{\partial Q_{ij}}{\partial C_{ijm}} C_{ijm} + \sum_k \frac{\partial Q_{ij}}{\partial L_{ijk}} \cdot L_{ijk} \quad (7.24)$$

Med marginalproduktivtetsanpassad avkastning på såväl arbete som kapital erhålles då

$$Q_{ij} = \sum_m r_{ijm} C_{ijm} + \sum_k W_{ijk} \cdot L_{ijk} \quad (7.25)$$

Ekv. (7.25) innebär att produktionsresultatet kan ses som det samlade värdet av avkastningen på använt kapital och den utbetalade lönesumman. Annorlunda uttryckt innebär detta att produktionsresultatet alltid kan fördelas på en andel som tillfaller arbetskraften (löneandelen) och en andel som tillfaller kapitalet (kapitalandelen). Med hjälp av ekv. (7.25) och (7.23) kan påståendet ekv. (7.22) efter förenkling skrivas om som

$$Q_{ij} = \frac{\sum_k W_{ijk} L_{ijk}}{\sum_j \sum_k W_{ijk} L_{ijk}} \cdot \sum_j \left[ \sum_m r_{ijm} C_{ijm} + \sum_k W_{ijk} L_{ijk} \right] \quad (7.26)$$

För att påståendet då skall gälla, dvs för att ekv. (7.25) och (7.26) skall vara lika, krävs att

$$\frac{\sum_k W_{ijk} L_{ijk}}{\sum_j \sum_k W_{ijk} L_{ijk}} \cdot \sum_j \sum_m r_{ijm} C_{ijm} = \sum_m r_{ijm} C_{ijm} \quad (7.27)$$

Detta uttryck kan enkelt skrivas om som

$$\frac{\sum_k W_{ijk} L_{ijk}}{\sum_j \sum_k W_{ijk} L_{ijk}} = \frac{\sum_m r_{ijm} C_{ijm}}{\sum_j \sum_m r_{ijm} C_{ijm}} \quad (7.28)$$

Då man fördelar sektorprodukten för hela regionen på regiondelarna i proportion till dessas andel av lönesumman i sektorn, så förutsätter man således att varje enskild regiondels andel av sektorns lönesumma är densamma som andelen av sektorns kapitalavkastning. Annorlunda uttryckt innebär detta att för en given sektor skall för två godtyckliga regiondelar förhållandet mellan dessas produktion, lönesummor och kapitalavkastning vara detsamma.<sup>1)</sup> Är räntan lika i alla regiondelar innebär den ovan relaterade fördelningsprincipen för sektorprodukten, att man antar att produktionen i de olika regiondelarna förhåller sig till varandra på samma sätt som regiondelarnas kapitalvolym och, liksom tidigare, deras lönesummor.

Huruvida de redovisade antagandena är realistiska eller ej är naturligtvis en rent empirisk fråga. Det enda man kan säga på nuvarande stadium av arbetet är att modellen gäller under antaganden som intuitivt förefaller rätt rimliga.

### 7.3 Icke-konvexa egenskaper hos inomregionala lokaliseringsproblem, förekomst av stordriftsfördelar

I Lefebbers modell och i den presenterade pendlingsmodellen för Stockholmsregionen antas produktionsfunktionerna vara strikt konkava. Härigenom vinner man fördelen att problemet blir lösbart med jämförelsevis enkla optimeringsmetoder. Dessutom får variablerna i dualproblemen en entydig betydelse vid mera politiskt orienterade tolkningar.

Det finns emellertid också stora nackdelar med dessa konkava produktionsfunktioner. Vid mera långsiktig analys tycks de flesta produktionsteoretiska och empiriska undersökningar tala för att det finns markerade odelbarheter eller stordriftsfördelar

<sup>1</sup> Formellt kan detta skrivas för två godtyckliga regiondelar  $j$  och  $j'$

$$\frac{Q_{ij}}{Q_{ij'}} = \frac{\sum_k W_{ijk} L_{ijk}}{\sum_k W_{ij'k} L_{ij'k}} = \frac{\sum_m r_{ijm} C_{ijm}}{\sum_m r_{ij'm} C_{ij'm}}$$

i produktionen. I en Cobb-Dauglas-funktion tar det senare sig uttryck i att när kapitalet också blir variabelt så kommer summan av exponenter på arbetskraften plus exponenten på kapitalet att tillsammans överstiga värdet 1. Det innebär att de problem som har ett entydigt optimum vid pendlingsanalysen kommer att sakna denna entydighet vid en lokaliseringspolitisk analys där också kapitalet tillåts variera i lokalisering. Det är kanske lämpligt att något belysa graden av stordriftsfördelar enligt ett par empiriska undersökningar som använt relativt olika undersökningsteknik.

Den ena undersökningen är utförd på ett amerikanskt tvärsnittsmaterial och syftar till att estimeras produktionsfunktioner av Cobb-Douglas-typ, Hildebrand, Liu (1965). I undersökningen indelas arbetskraften i två utbildningsnivåer, medan kapitalet ses som en enhetlig storhet. Beräkningarna sker med hjälp av tvärsnittsanalys för USA:s olika stater med separat estimering för var och en av industrisektorerna. TAB. 7.1 återger de uppmätta stordriftsfördelarna enligt Hildebrand och Lius undersökning. Stordriftsfördelarna kan där avläsas som summan av faktorernas output-elasticiteter.

Industrigren	Rang efter stordriftsfördel	$\alpha$	$\alpha_1$	$\alpha_2$	$\beta$	$\sum_i \alpha_i + \beta$
Livsmedel	1		0,31	0,40	0,53	1,24
Kemisk	2		0,34	0,57	0,27	1,18
Metall	3	0,96			0,16	1,12
Instrument	4	0,67			0,44	1,11
Trä	5	0,79			0,31	1,10
Papper	6		0,55	0,27	0,28	1,10
Jord & sten	7		0,66	0,30	0,13	1,09
Gummi	8	0,85			0,23	1,08
Maskin	9		0,47	0,27	0,33	1,07
Metallvaror	10		0,53	0,34	0,15	1,02
Transportmedel	11		0,41	0,28	0,32	1,01
Petroleum	12		0,27	0,50	0,23	1,00
Elektroindustri	13		0,41	0,24	0,30	0,95
Läderindustri	14	0,85			0,07	0,92

TAB. 7.1. Mätning av stordriftsfördelar under konstant teknologi.

De rubriker som återges i tabellhuvudena är koefficienterna i följande två produktionsfunktioner:

$$Q = A K^\beta L^\alpha \quad (7.29)$$

där

$Q$  = förädlingsvärde i en sektor per år

$K$  = användning av kapital i en sektor per år

$L$  = användning av arbete i en sektor per år

$A$  = teknologisk nivåparameter

$$Q = A K^\beta L_1^{\alpha_1} \cdot L_2^{\alpha_2} \quad (7.30)$$

där

$L_1$  = arbetares arbetsinsats per år

$L_2$  = tjänstemäns arbetsinsats per år

Av tablån framgår att huvuddelen av sektorerna skulle ha stor-driftsfördelar, trots att den amerikanska marknadens storlek tillåtit sektorerna att koncentrera produktionen till stora anläggningar i mera uttalad utsträckning än i länder som Sverige med dess trots utrikeshandeln mera begränsade marknader. De flesta ekonomer antar att summan av elasticiteter bara kan sättas lika med en konstant inom ett begränsat produktionsintervall. Det finns därför skäl att anta, att det problem med icke-konvexitet i allokeringsanalysen, som iakttagits i den refererade undersökningen skärps när det gäller lokalisering inom Stockholmsregionen med dess i många sektorer mycket begränsade marknader.

Produktionsfunktioner är i allmänhet att föredra som analytiska hjälpmedel, eftersom de innehåller information om möjligheten att optimalt kombinera de knappa produktionsfaktorerna. I praktiska undersökningar visar det sig emellertid ofta vara lämpligt att arbeta med kostnadsfunktioner som visar den samlade faktorkostnadens variation med produktionsskalan (se FIG. 7.1). I många lokaliseringsanalyser har stordriftsfördelar behandlats med denna mera förenklade teknik.

SOU 1970:30 lämnar viss information om kostnadskurvornas förlopp över växande produktionsvolym. Med hjälp av sådana kostnadsfunktioner kan man kanske fastslå produktionsvolymen hos de nya an-



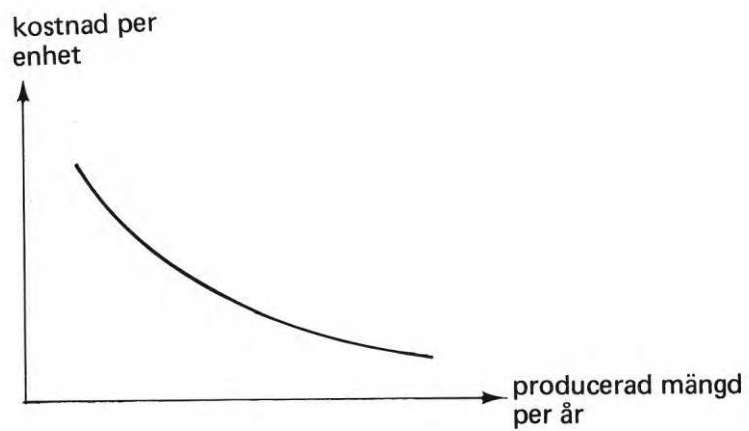


FIG. 7.1 Styckekostnadens variation med produktionskostnaden.

läggningar som kommer till under en planperiod. TAB. 7.2 kan något belysa vilka storlekar på nya anläggningar som kan bli aktuella inom industrin under de närmaste tio åren.

	Optimalt antal anläggningar	Antal anlägg- ningar 1970
Bryggerier	4	56
Mejerier	färre än 50	ca 230
Charkuterier	ca 12	ca 200
Bagerier	ca 10	ca 2000
Kvarnar	4	85% av prod. i 11 kvarnar
Oljeraffinaderier	1	3
Cementfabriker	3-4	7

TAB. 7.2 Jämförelser mellan optimal och existerande anläggningsstruktur, några branscher.

Det ligger en väsentlig nackdel i koncentrationsutredningens bedömningar, eftersom dessa endast avser tillverkningsindustrin. Produktions- och konsumtionservice av privat och offentlig art kommer att bli de mest expansiva sektorerna under överskådlig tid. Det är därför nödvändigt att få motsvarande bedömningar genomförda för dessa sektorer om informationen om odelbarheter i anläggningar skall ha något väsentligt intresse i lokaliseringsanalyser för Stockholmsregionen.

Det är ofta svårt att välja mellan en produktions- och kostnadsteoretisk ansats vid studiet av odelbarheter och stordriftsfördelar i industrin. Svårigheterna blir emellertid än större om det gäller att fastslå odelbarhetsgränser i serviceverksamheter, eftersom dessa sektorer producerar tjänster med svårsmätbara kvalitetsegenskaper.

I Lefebers modell är det - på grund av antaganden om delbarhet hos såväl input som output och antagandet om konkava produktionsfunktioner - alltid möjligt att finna enkla villkor för maximal produktion av den ena varan givet den andra varans produktionsmål skall vara uppfyllt. Genom att ansätta en prisrelation mellan va-

ra 1 och vara 2 (prisrelationen  $P_0 - P_0$  i FIG. 7.2) bringas företagen vid vinstmaximering att producera mängderna  $x_0^1, x_0^2$  som utgör en "paretoeffektiv" kombination, eftersom den ligger på randen  $T_0 - T_0$  av transformationsmängden. Om de individuella nyttorna hos konsumenterna också kunde aggregeras upp till en välfärdsfunktion  $w = w(x^1, x^2)$  - som uppvisar samma konvexitet så kan systemet bringas i allmän jämvikt genom prisrelationen  $P_0 - P_0$ , såvida hushållen nyttomaximerar.

Om det i stället råder odelbarheter i produkterna  $x^1$  och  $x^2$  kan transformationsmängden få ett utseende som i FIG. 7.3 (O, A, B, C, D, E) som är icke-konvex, därigenom att endast punkterna på och innanför randen är tillåtna. Låt oss anta punkterna  $\alpha, \beta$  och  $\gamma$  är indifferentpunkter för hushållen, varvid kombinationen  $\alpha$  är den punkt bland de kombinationer som tillhör mängden som ger högsta välfärd. I detta fall av odelbarheter - stordriftsfördelar - är det omöjligt att ansätta en prisrelation som förmår att föra både hushåll och företag till den optimala kombinationen  $\alpha$ . I många fall med odelbarheter är det sålunda svårt att åstadkomma ett optimalt decentraliserat beslutsfattande med enkel prisinformation som huvudsakligt beslutsunderlag för företag och hushåll. Det finns naturligtvis möjligheter att med mera komplicerad prisstruktur åstadkomma ett decentraliserat beslutsfattande, som leder till välfärdsmaximering.

#### 7.4 Transportmodeller som hjälpmedel för inomregional lokaliseringsanalys vid odelbarheter

Odelbarheter är ett ofrånkomligt inslag i lokaliseringsanalyser. Varje nytt lokaliseringsobjekt får förutsättas ha en relativt fixerad storlek i första hand med hänsyn till produktionsteknikens villkor. På samma sätt får man förutsätta att de ytor som kan komma ifråga för lokaliseringsändamål i allmänhet av produktionstekniska skäl bör dimensioneras så att de svarar mot tekniskt rimliga storlekar eller multiplar av sådana.

Optimering under odelbarheter är i själva verket inte alls ovanligt inom lokaliseringsteorin. Transportmodellerna har fördelen att tåla odelbarheter under vissa begränsande förutsättningar.

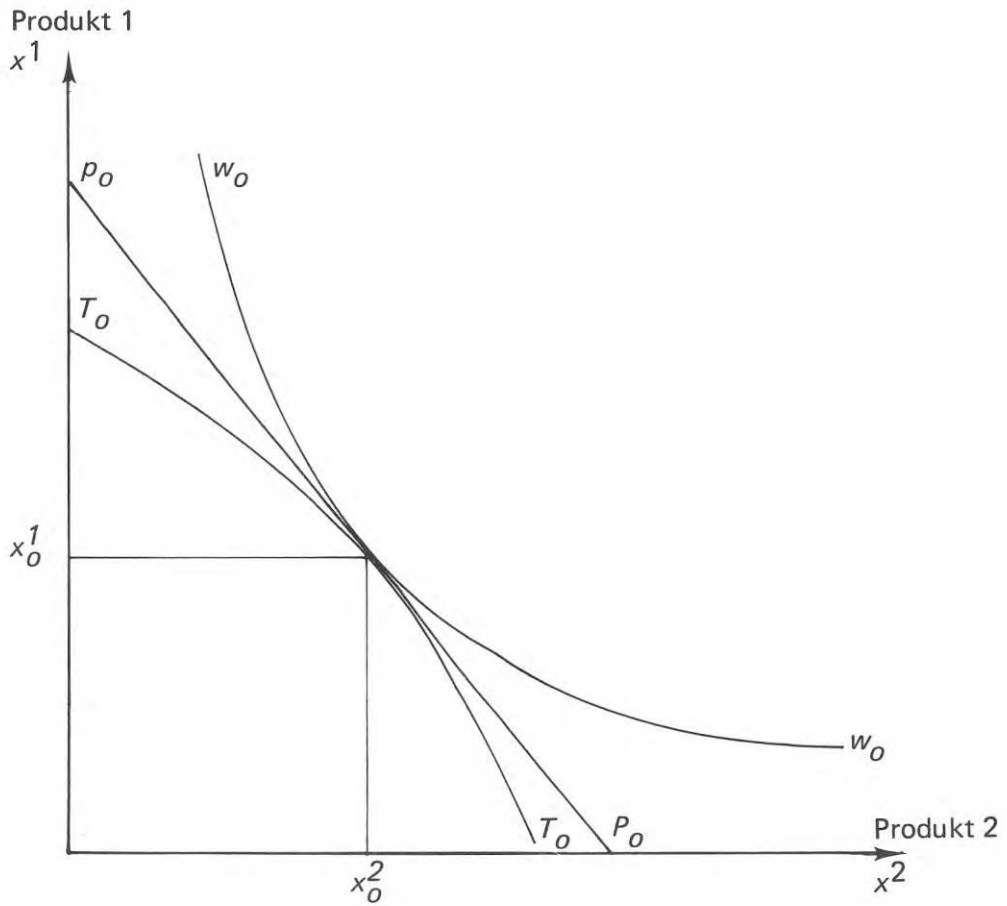


FIG. 7.2 Exempel på transformationsmängd ( $T_0$ ), välfärdsfunktion  $w_0$  och prisrelation  $P_0$  vid fullständig delbarhet.

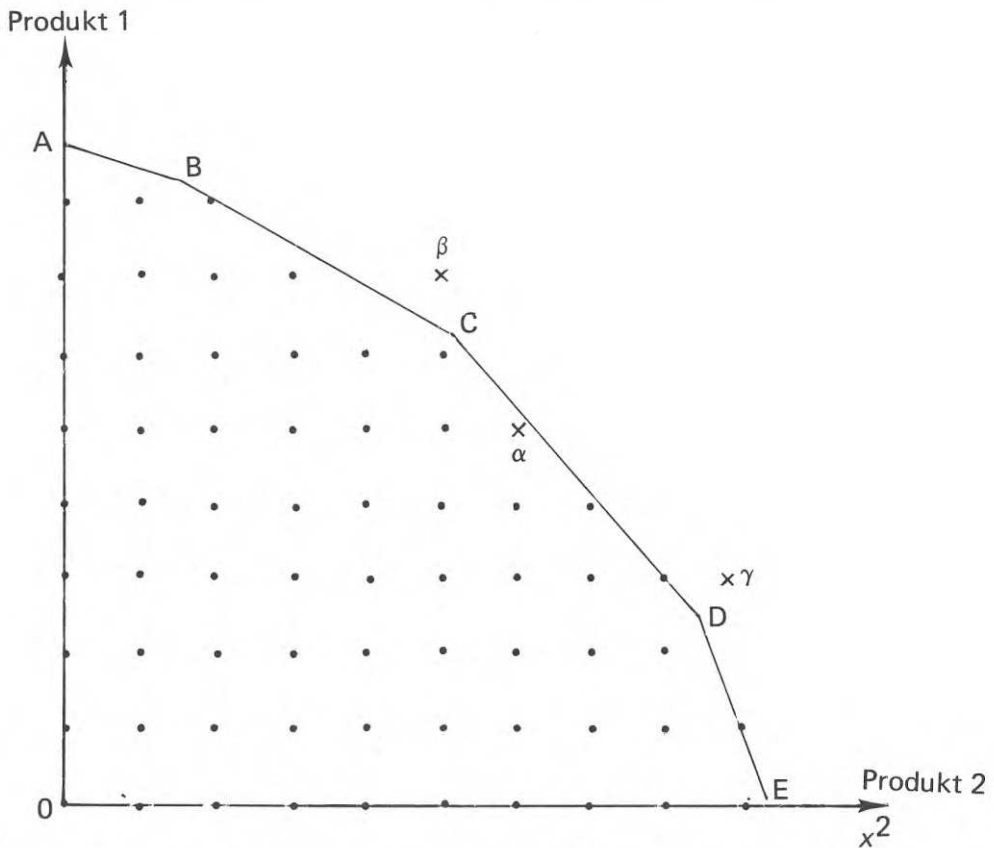


FIG. 7.3 Exempel på en transformationsmängd vid odelbarhet.

PrimalproblemetMaximera:

$$\sum_i \sum_j a_{ij} x_{ij}$$

Restriktioner:

$$\sum_i x_{ij} = \bar{x}_j$$

$$\sum_j x_{ij} = \bar{x}_i$$

$$x_{ij} \text{ heltal} \leq \min(\bar{x}_i, \bar{x}_j)$$

DualproblemetMinimera:

$$\sum_i u_i \bar{x}_i + \sum_j w_j \bar{x}_j$$

Restriktioner:

$$u_i + w_j \geq a_{ij}$$

$a_{ij}$  = lönsamhet av anläggning i i läge j

$x_{ij}$  = antal anläggningar av typ i i läge j

$u_i$  = dualpris för användning av anläggning i

$w_j$  = dualpris för användning av läge j (dvs markränta)

$x_{ij}$  antar heltalsvärden under förutsättning att restriktionerna  $\bar{x}_i$  och  $\bar{x}_j$  antar heltalsvärden. Den egenskapen går förlorad, om något av  $\bar{x}_i$  eller  $\bar{x}_j$  antar fraktionella värden.

Man kan visa, se Gale (1960), att  $u_i$  och  $w_j$  kan väljas  $\geq 0$ .

Det ekonomiskt komplicerade i modellen ligger i bestämningen av a-koefficienterna. Den samhällsekonomiska lönsamhetsanalysen är ett ofrånkomligt hjälpmedel för att beräkna dessa koefficienter. Enligt den samhällsekonomiska kostnads-intäktsanalysen kan a-koefficienten ses som den samhälleliga lönsamheten av att lägga en viss verksamhet i ett visst läge. I ett statistiskt sammanhang kan a-koefficienten ses som skillnaden mellan intäkter och kost-

nader, där intäkter och kostnader är beräknade utifrån samhällets värderingar. Gäller det en längre planperiod måste fördelen utvärderas över hela tidsförloppet.  $a_{ij}$  blir då bestämda som summan av alla perioders avkastning diskonterad till nuvärde i enlighet med samhällets tidspreferens. Om nettoavkastningen inte beräknas växa över tiden blir  $a_{ij}$  bestämda som i ekvationerna (7.30), (7.30 a) och (7.30 b).

$$a_{ij} = \int_0^{\infty} (I_{ij} - K_{ij}) e^{-rt} dt \quad (7.30)$$

$$a_{ij} = (I_{ij} - K_{ij}) \int_0^{\infty} e^{-rt} dt \quad (7.30 a)$$

$$a_{ij} = V_{ij} \frac{1}{r} \quad \text{för } r > 0. \quad (7.30 b)$$

I den samhälleliga utvärderingen gäller det sålunda att fastslå både nettoavkastningen och samhällets tidspreferens på ett enligt samhällets värderingar korrekt sätt.

Konsumentöverskottet är oundgängligt vid beräkningen av intäkterna av odelbara projekt <sup>1)</sup>. I samhällsekonomisk lönsamhetskalkyl gäller det att beräkna den samhälleliga intäkten av ett investeringsprojekt som summan av konsumentöverskott. Konsumentöverskottet för en individ kan något förenklat definieras som skillnaden mellan det pris som konsumenten vore beredd att betala hellre än att avstå från konsumtionstillskotten och det pris som han faktiskt får betala. Alternativt kan konsumentöverskottet definieras som den summa pengar som konsumenten skulle kräva som subsidie för att avstå från konsumtionstillskottet. Dessa två grunddefinitioner är ekvivalenta under vissa betingelser. Konsumentöverskottet kan illustreras på följande sätt (FIG. 7.4).

Efterfrågekurvan E-E uttrycker vilka priser konsumenten vore beredd att betala hellre än att avstå från konsumtionstillskott. Vid mängden  $q_0$  är konsumenten beredd att betala priset  $p_0$ .

<sup>1</sup> Arvidsson, G, Niklasson, H, 1968, On consumer surplus and related concepts.

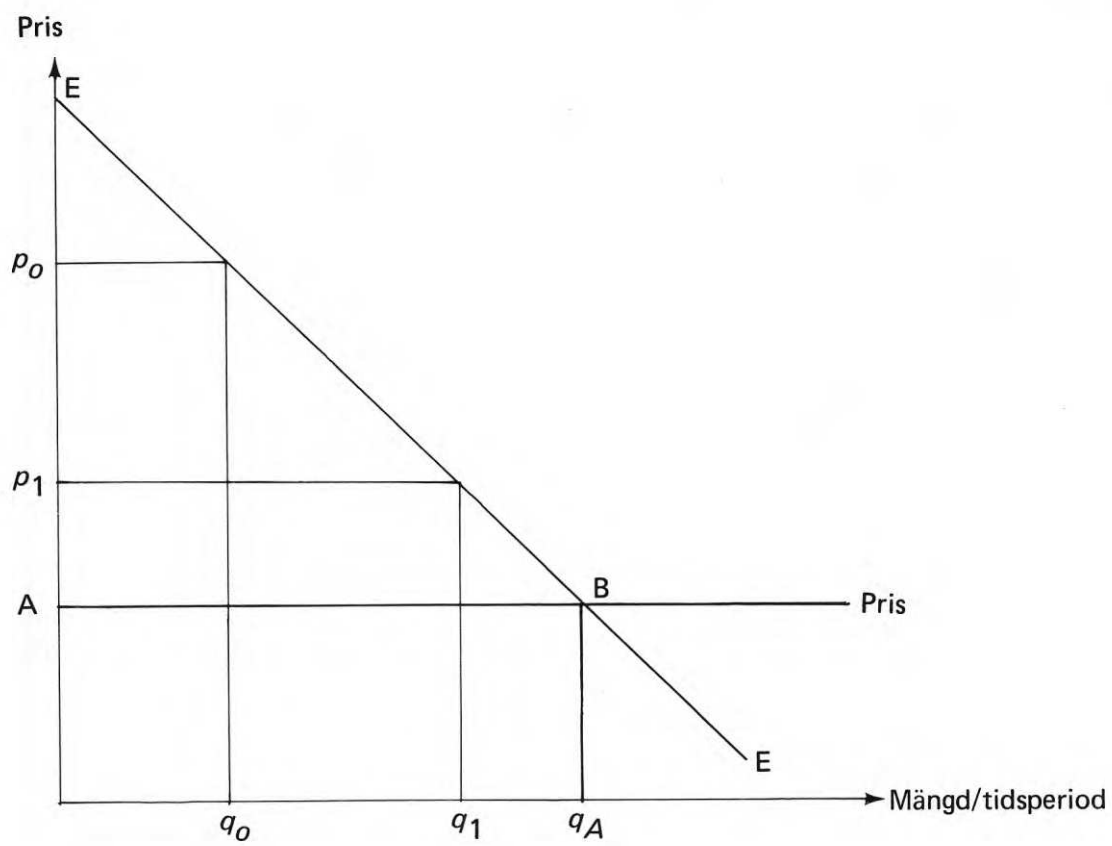


FIG. 7.4 Konsumtionsöverskottets definition.

hellre än att avstå från ett konsumtionstillskott i närheten av  $q_0$ . Vid tilldelningen  $q_1$  är han beredd att betala priset  $p_1$  hellre än att avstå från tillskott i närheten av  $q_1$ . Den totala konsumenträntan är sålunda vid tilldelningen  $q_A$  triangeln A-E-B. Den mängd som efterfrågas vid ett givet pris varierar självfallet från individ till individ, men för var och en av individerna efterfrågas en viss mängd per tidsperiod, när ett visst pris gäller. Det är därför möjligt att aggregera fram en efterfrågefunktion för alla hushåll tillsammans, för vilken det gäller att ytan under efterfrågekurvan över prislinjen svarar mot summan av de individuella konsumenternas konsumenträntor. Den korrekta efterfrågekurvan för dessa beräkningar av konsumenträntan är inte så helt enkel att bestämma. De estimeringsproblem som uppstår har emellertid grundligt utretts i Hicks (1956).

Ytan under efterfrågekurvan torde kunna härledas ganska väl i lokaliseringssammanhang, eftersom vi där inom vissa intervall har perfekt prisdiskriminering som uppstår på så sätt, att hushållets pris är summan av ett enhetligt marknadspris och en med distansen till produktionsenheten varierande reskostnad (inklusive värdet av tidsförlusten). Det pris som användes vid konsumentränteberäkningen skall vara sådant, att det uttrycker den samhällseliga marginalkostnaden. Lönsamhetstalet blir följaktligen skillnaden mellan konsumenträntan och den fasta årskostnaden integrerat från nuläget mot oändligheten eller till planhorisonten. Sådana värden skall beräknas sektorvis för alla möjliga lokaliseringpunkter. I den här återgivna modellen existerar alltid en lösning förutsatt att det inte råder starka ömsesidiga beroenden i lönsamheten hos olika sektorers lokalisering. I så fall är det omöjligt att hitta någon bästa lösning annat än med fullständig genomgång av alla tänkbara alternativ, vilket är en omöjlig procedur om antalet sektorer och antalet lägen är stort.

#### 7.5 Ömsesidiga beroenden som komplicerande faktor vid odelbarheter

Det är välkänt att ömsesidiga beroenden förekommer i stor utsträckning i det ekonomiska systemet. De ömsesidiga beroendena är emellertid av mycket olikartad natur. Vissa ömsesidiga beroenden är av mera teknisk natur, dvs produktionen av en vara eller tjänst



förutsätter insatser av producerade varor och tjänster som inputs i produktionsprocessen. Varje vara och tjänst ingår dessutom i någon utsträckning som input till andra produktionsprocesser. Ömsesidiga beroenden av sådan teknisk natur är så omfattande att det finns anledning att försöka lösa några av dem redan före en egentlig analys av det optimala lokaliseringsmönstret. Den kanske besvärligaste typen av ömsesidiga beroenden ligger väl i samspelet mellan företagens lokaliseringsmönster, transportsystemets struktur och hushållens bosättningsmönster. Om det skall vara möjligt att utreda det optimala samspelet mellan dessa tre mönster torde det vara nödvändigt att arbeta med en modell, där olika delar av produktionssystemet (utom transporter och bosättning) visar låg grad av ömsesidigt beroende.

#### 7.6 En väg att lösa de tekniskt bestämda ömsesidiga beroendernas problem

Input-outputtabblån är ett hjälpmedel att beskriva olika produktionssektorers leveranser av varor och tjänster till varandra. Ett produktionssystem av input-outputtyp har en form som ges i ekv. (7.31).

$$x = Ax + f \quad (7.31)$$

där

$x$  är en kolonnvektor med  $n$  sektorer

$A$  är en koefficientmatris med  $n$  kolonner och  $n$  rader

$I$  är en enhetsmatris

$f$  är en kolonnvektor med  $n$  rader som anger användning av olika sektorers varor och tjänster vid slutlig förbrukning, dvs för konsumtion, investeringar och export.

Om de teknologiska sambanden är stabila kan modellen användas för beräkningar av olika sektorers produktionsnivåer vid en viss tänkt sammansättning av slutförbrukningen ( $f$ ), ekv. (7.32).

$$x = (I - A)^{-1} f \quad (7.32)$$

Om alla sektorer är ömsesidigt beroende av varandra skulle  $A$ -matrisen ha positiva värden i alla element. I realiteten är vissa

element lika med 0 och vissa av elementen är mycket små. Det finns därför möjlighet att arrangera A-matrisen i sådan form att beräkningar underlättas. Det finns exempelvis ofta möjlighet att åstadkomma en i det närmaste triangulär A-matris. Sådana trianguleringar har genomförts för många länders input-output-matriser och då visat sig ha mycket god stabilitet vid länderjämförelser. Det har exempelvis visat sig, att trianguleringsmönster som uppnåtts när en given trianguleringssteknik applicerats på så skilda ekonomier som Turkiets, Israels, Norges och USA:s input-outputmatriser givit i stort sett samma trianguleringsmönster. (Rangkorrelationstest ger värden om 0.7 - 0.9 vid parvisa jämförelser av den triangulerade sektorordningen). Om omsorteringar av input-outputmatriser genomföres på ett ur lokaliseringspolitisk synvinkel rimligt sätt kan det alltså väntas, att dessa mönster har en strukturell stabilitet även vid drastiska förändringar i institutionella betingelser och förändringar i pris-systemet.

Input-outputtablåer av mera fullständig art innehåller ofta över 100 sektorer. Det säger sig självt, att många av sektorerna inte har några ömsesidiga kontakter. Konfektyrindustrin köper exempelvis inga varor från varvsindustrin och varvsindustrin köper heller inte varor och tjänster från konfektyrindustrin. Det är uppenbart att två sådana sektorer bara indirekt kan ha nytta av samlokalisering via kontakter med någon annan sektor. Såväl de element som uttrycker leveransbehov från konfektyrsektorn som i de element i A-matrisen, som uttrycker leveransbehov i motsatta riktningen är koefficienterna nära noll. Summan av dessa koefficienter är följaktligen också noll. Varvsindustrin har däremot stort inputbehov från plåtmanufaktursektorn, medan den sistnämnda sektorn knappast köper något från varvssektorn. Det samlade samspelet mellan dessa sektorer kan därför uttryckas som summan av en stor input-output-koefficient och 0. På detta sätt kan sektorkombination för sektorkombination ordnas så, att element  $a_{ki}$  adderas till element  $a_{ik}$ . Dessa summerade element som uttrycker intersektoriell association kan därefter sättas in som värden både i läge  $ik$  och i läge  $ki$ , medan den ledande diagonalen självklart visar maximal association, dvs värdet 1. På detta sätt erhålles en symmetrisk matris som uttrycker styrkan i de tekniska sambanden mellan sektorerna.

Sektor	1	2	3	...	n
1	1	$(a_{12}+a_{21})$	$(a_{13}+a_{31})$	...	$(a_{1n}+a_{n1})$
2	$(a_{12}+a_{21})$	1	$(a_{23}+a_{32})$	...	$(a_{2n}+a_{n2})$
3	$(a_{13}+a_{31})$	$(a_{23}+a_{32})$	1	...	$(a_{3n}+a_{n3})$
...	...	...	...	...	...
n	$(a_{1n}+a_{n1})$	$(a_{2n}+a_{n2})$	$(a_{3n}+a_{n3})$	...	1

FIG. 7.5 Symmetriserad och korrigerad input-outputmatris.  
 $a_{ij}$  = leverans av produkter från sektor i till sektor j per produktenhet j.

Efter en korrigerad av den ursprungliga input-output-matrisen kommer det att finnas en stor mängd koefficienter med värdet 0. Många koefficienter har mycket låga värden medan andra blir höga. Eftersom styrkan i sektorernas ömsesidiga beroende varierar starkt, finns det möjlighet att omforma matrisen till kvasidiagonalitet, där matrisen segmenteras i ett antal approximativt kvadratiske delmatriser utefter huvuddiagonalen. Dessa delmatriser kan ses som lokaliseringpolitiska klusterbildningar med mera markerade odelbarheter än vid den noggrannare sektoriseringen i ursprungsmatrisen. Dessa kluster har emellertid ett väsentligt reducerat ömsesidigt beroende gentemot övriga delar av ekonomin och kan därför hanteras i en modell för allokering under odelbarheter utan ömsesidigt beroende (dvs assignment-modellen eller transportmodellen).

De rena tekniska koefficienterna har ett begränsat värde vid en blockdiagonalisering för lokaliseringsändamål. Lagrings-tiden för olika sektorprodukter varierar mycket kraftigt och eftersom input-output-koefficienter estimeras över en ettårsperiod måste en korrigerings ske med hänsyn till olika outputs lagringstid.

Det tycks inte finnas någon metod för att finna en globalt sett optimal klusterbildning i en matris av denna art. Den modell som redovisas av Lundqvist och Karlqvist (se Lundqvist (1970)) kan möjligen användas för klustringsanalysen även om dimensionen på meningsfulla input-output-matriser kan vålla besvär. Den input-output-matris som utarbetats för Sverige är indelad i över ett hundra sektorer, vilket innebär att input-output-matrisen innehåller mer än 10.000 element.

Faktoranalys kan möjligen vara en metod att undersöka förekomsten av lokaliseringskluster. Vi tänker oss ett  $n$ -dimensionellt euklidiskt rum, där varje sektor  $i$  representeras av en enhetsvektor  $x_i$ . Elementen i den symmetriserade I-O-matrisen anger graden av association mellan sektorerna. I det  $n$ -dimensionella rummet mätes två sektorer association som vinkeln mellan två motsvarande vektorer. Koefficienten för en sektors egen-association ( $c_{ii}$ ) sättes godtyckligt lika med 1, och summan av övriga element i motsvarande rad (kolumn) sättes mindre än 1, vilket innebär att alla element utanför huvuddiagonalen multipliceras med ett lämpligt positivt tal. Denna åtgärd innebär att en sektors egenassociation är starkare än associationen till alla andra sektorer.

Faktoranalysen kan formuleras som ett minimeringsproblem. En restriktion säger att vektorn  $x_i$  skall vara en linjär kombination av  $p$  faktorer  $f_j$  och en restterm  $\epsilon_i$ ;  $p$  är förutbestämt och mindre än  $n$  och anger dimensionaliteten hos det under- rum som approximativt innefattar alla  $x_i$ . En annan restriktion säger att den ursprungliga C-matrisen skall kunna återskapas genom användning av den första restriktionen och villkoret att de  $p$  faktorerna är ömsesidigt ortogonala och av längden 1. Sammanfattningsvis kan problemet formuleras på följande sätt:

Minimera:

$$\sum_{i=1}^n \epsilon_i^2$$

Restriktioner:

$$1. \quad x_i = \sum_{j=1}^p d_{ij} \cdot f_j + \epsilon_i ; \quad (i = 1, \dots, n)$$

med egenskapen hos  $f$  att  $\left\{ \begin{matrix} f_i^T \\ f_j \end{matrix} \right\} = I$ ;  $I$  är en  $n \times n$ -enhetsmatris

$$2. \quad \left\{ \begin{matrix} x_i^T \\ x_j \end{matrix} \right\} = C$$

$\{d_{ij}\}$ -matrisen anger vinkeln mellan enhetsvektorn  $x_i$  och faktorn  $f_j$  och kallas ofta för  $x_i$ 's faktorladdning på  $f_j$ . Beteckningen  $T$  anger här transponering.

Längden  $h_i$  (kommunaliteten) av  $x_i$ 's projektion på basen  $f_j$  fås ur sambandet  $h_i^2 = \sum_{j=1}^p d_{ij}^2$ . Eftersom  $x_i$  är en enhetsvektor gäller att  $h_i^2 + \epsilon_i^2 = 1$ . Minimering av  $\sum \epsilon_i^2$  innebär således, att det underrum som spänns upp av basen  $f_j$  så fullständigt som möjligt innehåller alla vektorerna  $x_i$  (se FIG. 7.6).

Faktoranalysen bör kunna användas på en symmetriserad  $I$ - $O$ -matris om den är diagonaldominant såsom tidigare förutsatts och följaktligen positivt definit. Om alla egenvärden dessutom är olika, vilket i allmänhet gäller, är det möjligt att uppfatta matrisen  $C$  som  $G^T G$  där  $G$  är en  $n \times n$ -matris.  $C$  får då samma egenskaper som en korrelationsmatris, och det är då tillåtet att uppfatta matriselementen som vinklar mellan par av vektorer. Faktoranalysens metod ger naturligtvis bara ett bland många möjliga hopgrupperingsmönster. Om klustringsanalysen, som den utföres med hjälp av faktoranalys, skall accepteras som utgångspunkt för den fortsatta lokaliseringsanalysen blir en bedömningsfråga. Skulle de faktoranalytiska kalkylerna ge för dåligt resultat blir det nödvändigt att försöka anpassa Lundqvist-Karlqvist-modellen för detta ändamål.

$x_j$ : s vinkel-  
räta komponent  
mot  $f$ -basen

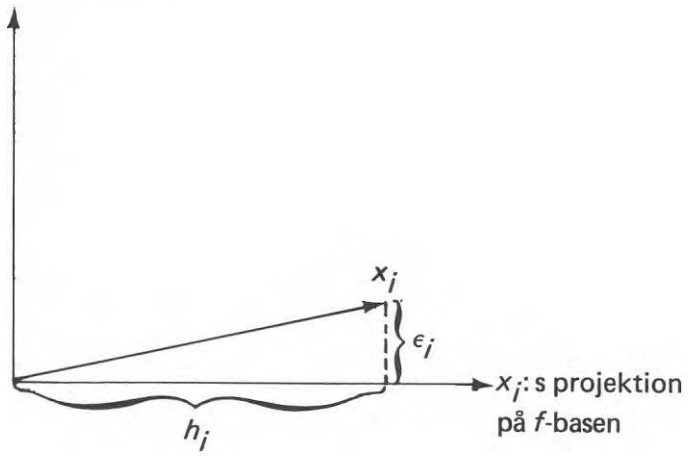


FIG. 7.6 Illustration av  $x_j$ :s projektion på basen  $f_j$ .

Det kan också vara meningsfullt att pröva flera av de faktoranalysen närstående metoderna för klustringsanalys.

### 7.7 Värdering med en dynamisk transportmodell

Om klustringsanalysen kan genomföras på ett framgångsrikt sätt - med faktoranalys eller med en modifierad variant av Lundqvists och Karlqvists modell - finns det möjlighet att betrakta dessa kluster som kärnor i en flerkärnig region. Dessa kärnor kan sedan placeras ut i sådana lägen som med hänsyn till odelbarheterna tål framväxten av fullständiga kluster. Hur dessa kluster skall placeras ut i rummet beror på sådana interdependenser som inte behandlats i den tidigare beskrivna klustringsanalysen. Det gäller här att uppnå en optimal placering med hänsyn till under planperioden oföränderliga lokaliseringar av företag, bosättning och transportsystem samt med hänsyn tagen till på särskilt sätt planerade förändringar i bosättningsmönstret. Det torde här vara en nödvändig förutsättning att analysera nattbefolkningens bosättning med en särskild, bostadspolitiskt orienterad teknik. Den bostadspolitiska allokeringssmodellen för perioden måste inriktas på en prioritering av det sociala integrationskravet. Detta innebär att generalplanen skall byggas upp så, att de nuvarande tendenserna till segregation kommer att brytas. Det kommer i sin tur att innebära att tendenser till klustring av högutbildade personer i vissa områden och lågutbildade personer i andra områden bryts genom en kombination av bostadslokalisering och nykonstruerade system för hyresbidrag. Härigenom nedbringas fördelarna av att lokalisera verksamheter med mycket sned utbildningsprofil i anslutning till segregerade bostadsområden.

Utbildningsstrukturen i olika kluster kommer fortfarande att ha en avgörande betydelse för vart företagsklustren bör lokaliseras - i termer av nylokalisering respektive förstärkning eller försvagning av redan uppnådda klusterbildningar - på så sätt, att de kluster, som efterfrågar många och relativt sett små yrkeskategorier, bör välja sådana lokaliseringar som har god tillgänglighet mot hela regionen med såväl det kollektiva som det privata transportsystemet. Sådana kluster, som däremot efterfrågar större volymer av utbildningsmässigt vanliga eller adaptiva yrkesgrupper, kan söka sig mot sådana kärnor som inte har bästa accessibilitet



gentemot hela regionen. Men det är inte bara användningen av arbetskraft som får avgöra valet av lokalisering för klustren. Vissa kluster kan tänkas vara inriktade på export i större utsträckning än övriga. I så fall måste närheten till större terminaler och liknande kommunikationsanläggningar inom flyg-, järnvägs- och vägsystem få stor betydelse vid utvärderingen. Användning av vissa fast lokaliserade naturresurser som vatten liksom utsläpp av med produktionen komplementära föroreningar måste med nödvändighet ingå som element vid utvärderingen.

En fråga av särskilt stor betydelse är den grad av anpassning till förändringar i efterfrågeutvecklingen inom klustren som skall säkras genom reservation av markutrymmen för expansion i vart och ett av lägena. Det är möjligt att den av Markowitz förordade modellen för optimal portföljsammansättning under osäkerhet och risk kan få en användning vid beräkningen av lämpliga markreserver i de olika klustrens respektive lokaliseringsområden. En avgörande restriktion vid sådana kalkyler måste vara det finansiella utrymme, som kan skapas för en aktiv markpolitik.

Det gäller här att formulera en modell som används vid allokering av nya anläggningar till olika avgränsade ytor under olika tidsperioder. Sålunda gäller att liksom i den tidigare återgivna transportmodellen maximera den samhällseliga lönsamheten av de lokaliseringar som genomförs under tidsperioden. Den samhällseliga lönsamheten måste beräknas med utgångspunkt från att anläggningen kan tillkomma under någon godtycklig tidsperiod mellan, säg 1975-1980 och år 2000-2005. I varje sektor skall antalet anläggningar som tillkommer under en viss tidsperiod räknat över alla lokaliseringpunkter, hålla sig innanför en med annan teknik beräknad tillkomst av nytt kapital i sektorn under tidsperioden. Dessa investeringar, som skall ske i odelbara anläggningar, måste placeras in på de olika avgränsade ytor, som kan iordningställas under en given tidsperiod. Det kan förutsättas, att ytanvändningen är komplementär med kapitalmängden för en given sektor och en given regiondel i en given tidsperiod. Sambandet mellan expansion av kapaciteten och behovet av ytor kan därför troligen i allmänhet bestämmas genom användning av fasta koefficienter. Tillkomsten av nytt anläggningskapital för också med sig behov av mera arbetskraft. Det torde därför vara nödvändigt att upp-



märksamma de totala arbetskraftstillgångarnas tillväxt genom en särskild restriktion. Om de här nämnda tämligen enkla kraven uppmärksammas, skulle en utvärderingsmodell när det gäller lokalisering av anläggningar utanför transport- och bostadssektorn få följande utseende:

Maximera

$$\sum_i \sum_j \sum_t a_{ijt} X_{ijt}$$

Restriktioner

$$\sum_j X_{ijt} \leq \bar{X}_{it}$$

$$\sum_i b_{ijt} X_{ijt} \leq \bar{Y}_{jt}$$

$$\sum_i \sum_j c_{ijt} X_{ijt} \leq \bar{L}_t$$

$X_{ijt}$  är heltal  $\geq 0$

där

$X_{ijt}$  = under tidsperioden  $t$  tillkomna nya anläggningar i sektor  $i$  och läge  $j$

$a_{ijt}$  = den samhällseliga lönsamheten för all framtid av en under tidsperiod  $t$  tillkommen ny anläggning i sektor  $i$  och läge  $j$

$b_{ijt}$  = relationen mellan markyta och kapitalmängd i en under period  $t$  tillkommen anläggning i sektor  $i$  vid förläggning till läge  $j$

$c_{ijt}$  = bemanningskrav för en under period  $t$  tillkommen anläggning i sektor  $i$  vid förläggning till läge  $j$ . Denna koefficient kan vara densamma för alla lägen.

De problem med bestämningen av en samhällselig räntesats som behandlas i Marksjöes uppsats innebär särskilda problem i denna mo-

dell, där det gäller att fördela tillkomsten av ny anläggningskapacitet över olika tidsperioder och där diskonteringen skall ske från olika startår. Modellen är av heltalstyp och innebär därför stora lösningsproblem även i den här givna formen. Antalet variabler måste därför hållas nere. Redan om man arbetar med tre tidsperioder (säg 10 år vardera), sju kluster och sju områden, uppstår troligen lösningsproblem med de algoritmer och maskinella utrustningar som finns till hands.<sup>1)</sup> Det finns därför stor anledning att diskutera enkla modeller, som tål stor driftsfördelar, och som är utformade för kontinuerlig analys.

#### 7.8 Kan prognoser över lokaliseringsmönstret upprättas med rekursiv programmering?

De modellansatser som här redovisats kan ge god information om lämpliga lokaliseringsmönster, förutsatt att samspelet mellan investeringar i transportsystemet och investeringar i lokaliseringsmönstret kan klaras på ett logiskt tillfredsställande sätt. Om man i modellarbetet undviker att föra in beteendesamband av olika slag uppstår det problem när ett lämpligt genomförande av förändringarna i lokaliseringsmönstret skall övervägas. I realiteten kan det finnas så starka bindningar till det gamla lokaliserings- och bosättningsmönstret, att besluten om lokaliseringsförändringar i de enskilda företagen bildar beslutskedjor, som har en tendens att bevara och kanske förstärka den i föregående period uppnådda rumsliga strukturen. Dessa problem har behandlats av Serck-Hanssen i en nyutkommen avhandling (Serck-Hanssen (1970)).

En första teknik att komma åt skillnaderna mellan de samhälleligt optimala och de mikroekonomiskt optimala lokaliseringsmönstren består i att de samhällliga värderingskoefficienterna i den dynamiska transportmodellen bytes ut mot privatekonomiska koefficienter. I så fall kommer inte negativa externa effekter för konsumenterna att behandlas i optimeringen och därigenom blir

---

<sup>1</sup> Om det inte finns några odelbarheter uppstår inga lösningsproblem i en modell av denna art. I så fall kan man t ex tillämpa den algoritm, som utarbetats av Haley, K.B. och Smith, A.J. (se Haley, Smith (1966)).

mönstret annorlunda. En jämförelse mellan duallösningen till det samhälleliga optimeringsproblemet och det privatekonomiska kan då ge information om vilka beskattnings- och subsidieåtgärder som krävs för att åstadkomma ett korrigerat beteende.<sup>1)</sup> Men denna analysteknik torde inte räcka till som underlag för bedömning av genomförbarheten av ett samhälleligt önskat lokaliseringsmönster. Det kan dock vara en möjlighet att omformulera den återgivna dynamiska transportmodellen i rekursiv form, så att värdekoefficienterna återger mikroenheternas värderingar som bestämda utifrån det mönster som uppnåtts i närmast föregående period. X-värdena skulle i så fall symbolisera den totala kapitalstocken i en viss sektor och ett visst läge i en viss tidsperiod.<sup>2)</sup> Problemet blir då:

Maximera

$$\sum_i \sum_j a_{ij,t} (X_{ij,t} - X_{ij,t-1})$$

$$\sum_j X_{ij,t} \leq (1 + \delta_i) X_{i,t-1}$$

$$\sum_j X_{ij,t} \geq (1 - \delta_i) X_{i,t-1}$$

$$\sum_i b_{ij,t} X_{ij,t} \leq (1 + \gamma_j) \sum_i b_{ij,t-1} X_{ij,t-1}$$

$$\sum_i b_{ij,t} X_{ij,t} \geq (1 - \gamma_j) \sum_i b_{ij,t-1} X_{ij,t-1}$$

$$\sum_i \sum_j c_{ij,t} X_{ij,t} \leq \bar{L}_t$$

$$X_{ij,t} \text{ är heltal } \geq 0$$

där  $\delta, \gamma$  = maximal förändringstakt och där X - total kapitalstock; övriga symboler återges i anslutning till tidigare modellen.

<sup>1</sup> Se Johan Lönnroths uppsats.

<sup>2</sup> En formell analys av denna programmeringsteori finns i Day, R.H. (1963).

En modell av denna typ återger företagens försök att period för period maximera sina vinster och kan då inte visa en utveckling som är den bästa på lång sikt ens ur företagens synvinkel.

Den dynamiska transportmodellen kan omformuleras så, att transport- och bostadspolitiken bestämmas inom modellen i stället för att dimensioneras i annan form. Transportsektorn kan då ses som en markproducerande sektor, där en ny transportled öppnar ett markområde för exploatering. Odelbarheterna måste här bli mycket uttalade eftersom det krävs kombinationer av komponenter i huvudvägnät och lokalvägnät. Effekten av en viss vägexploatering ( $X_{1jt}$ ) beror då på den lokala åtkomligheten hos exploateringsbar mark. Markutbudseffekten ( $\beta_{1jt}$ ) av en väginvestering varierar därför från område till område.

Bostadssektorn (inklusive bostadskomplement) ( $X_{2jt}$ ) kan behandlas på liknande sätt. Den kan ses som producent av nytt arbete, där arbetseffekten av en ny bostad varierar från läge till läge i enlighet med fasta koefficienter  $\gamma_{2jt}$ .

Modellen kan nu omformas på följande sätt:

Maximera

$$\sum_i \sum_j \sum_t a_{ijt} X_{ijt}$$

Restriktion:

$$\sum_j X_{ijt} \leq \bar{X}_{it}$$

$$\sum_{i=2}^n b_{ijt} X_{ijt} - \beta_{1jt} X_{1jt} \leq 0$$

$$\sum_i \sum_j c_{ijt} X_{ijt} - \sum_j \gamma_{2jt} X_{2jt} \leq 0$$

$$\sum_j \gamma_{2jt} X_{2jt} \leq \bar{L}_t$$

$$X_{ijt} \text{ är heltal } \geq 0$$

$\beta_{1jt}$  = markutbud per vägprojekt  
(i  $m^2$ ) i läge  $j$  och period  $t$

$\gamma_{2jt}$  = arbetsutbud per ny bostadsanläggning (i arbetstimmar)  
i läge  $j$  och period  $t$

### 7.9 Dynamiska problem rörande ekonomisk tillväxt, ränta och optimal planhorisont

De två senast presenterade modellerna visar, att lokaliseringsproblemet inte kan lösas utan information om de olika produktionssektorernas tillväxt i regionen som helhet över tiden.

För att binda upp tillväxten i var och en av produktionssektorerna krävs det att den långsiktiga ekonomiska tillväxten i Stockholmsregionen har analyserats i en lämplig tillväxtmodell. Det är fördelaktigt att ta Marx-Leontiefs tillväxtantaganden som utgångspunkt för en diskussion om detta struktur- och tillväxtproblem. En sådan tillväxtmodell är sluten, dvs endast utgångsläget är en yttre förutsättning för tillväxtförloppet. Slutenheten i modellen innebär att de olika sektorerna innefattar en sektor, som producerar exportvaror och som för sin produktion använder importvaror samt hushållssektorn, som för att växa kräver ett tillflöde av konsumtionsvaror, bostäder och infrastrukturkapital och som levererar arbete till andra sektorer. Modellen har följande formella struktur.

$$x = Ax + B \Delta x$$

där  $A$  är en input-output-matris ( $n \times n$ )

Antag att  $\Delta x = \alpha x$

och att  $B = TA$

där  $T = \begin{bmatrix} t_1 & & & & 0 \\ & t_2 & & & \\ & & t_3 & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & & t_n \end{bmatrix}$

och  $t_i$  = livslängden hos produkt  $i$

$$x = Ax + B\alpha x$$

$$\left[ (I - A) - \alpha B \right] x = 0 \quad (7.33)$$

$$\left[ B^{-1}(I - A) - \alpha I \right] x = 0 \quad (7.34)$$

$$p \left[ B^{-1}(I - A) - \beta I \right] = 0 \quad (7.35)$$

Ekv. (7.34) kan lösas genom att determinanten för matrisen sättes lika med noll. Härigenom erhålles ett polynom i  $\alpha$  av lika hög grad som antalet sektorer. På lång sikt kommer det maximala egenvärdet att bli den dominerande roten. Härigenom sammanfaller denna dominanta rot med den maximala tillväxttakten under balans. Till egenvärdesproblemet ekv. (7.34) hör ett dualt egenvärdesproblem (7.35) där  $p$  anger en prisvektor. Vid maximal tillväxt svarar tillväxttakten  $\alpha$  mot dess monetära motsvarighet  $\beta$ , som kan tolkas som den optimala räntan. Den optimala planhorisonten är  $1/\beta$  eftersom  $B = T \cdot A$ .

Det första problemet i en sektoranalys på lång sikt är givetvis att bestämma egenvektorn  $x$ , motsvarande det maximala egenvärdet, vilken anger sektorernas relationer under maximal balanserad tillväxt. Denna vektor får sedan jämföras med sektorernas relationer i ett verkligt utgångsläge, där ekonomin troligen ligger vid sidan av den maximala tillväxtens stig. Därefter återstår problemet att diskutera hur ekonomin skall föras från sina iakttagna utgångsrelationer över till de tillväxtoptimala, vilket är ett rimligt problem med en avlägsen tidshorisont i planeringen. Sedan återstår också problemet att avgöra när och hur ekonomin skall föras ut från denna ur tillväxtsynpunkt optimala sektorstruktur för att uppnå eventuella terminalvillkor för planeringen. Väsentliga problem i sammanhanget är givetvis uppskattning av den teknologiska utvecklingen under tillväxtprocessen, vilken tar sig uttryck i fallande värden i koefficientmatrisen  $A$ .

#### 7.10 Ett förslag till stegvis genomförande av en lokaliseringsanalys för Stockholmsregionen

1. Den bosättningspolitiska frågan utredes med utgångspunkt från "Bostadspolitik för Stockholmsregionen" med tonvikt på antisegregationsmålet under fri prisbildning på bostadsmarknaden.

2. Det ömsesidiga beroendet mellan transportsystemet och bostättningsmönstret utredes, varvid substitutionsproblemen i transportsektorn också analyseras.
3. Klustringsanalys genomföres, varvid de ömsesidiga beroendena av teknisk art studeras med hjälp av input-output-matriser, som utvecklas så att koefficienterna även återspeglar lagringstiden.
4. Den dynamiska "transportmodellen" utvecklas, varvid antalet kluster, antalet delperioder och antalet regionkärnor bestämmes. Värderingsmetod bestämmes och värderingsförsök genomföres.
5. Adaptivitetsproblemet i den dynamiska markpolitiken utredes.
6. Regionens sektoriella tillväxt analyseras och vissa restriktioner bestämmes för den dynamiska "transportmodellen".
7. De olika sektorernas ytanvändning per produktions- eller kapacitetsenhet beräknas och tillgången på exploateringsmark i olika kärnlägen beräknas för olika planperioder.
8. Den rekursiva modellens användningsmöjligheter för prognos och planering utredes.
9. Modeller utvecklas för kvalitativ analys av optimal lokalisering på längre sikt. Dessa modeller måste ha mycket enkel utformning, t ex som hos Mills.
10. Plangenomförandets problem studeras genom att utnyttja alternativa målfunktioner i den dynamiska "transportmodellen" respektive genom planmässiga störningar av förloppet. Beräkningen utföres i den rekursiva planeringsmodellen.
11. Den mera detaljerade transportpolitiken på kort, medellång och lång sikt utredes med olika partiella modeller.

7.11 Litteraturreferenser

Andersson, Åke E, 1970, *SOU 1970:15: Regionalekonomisk utveckling*. (Allmänna förlaget.) Bilaga 7: Storstadsproblematiken. Stockholm.

Andersson, Åke E, 1969, *Storstadsregionens enhetliga arbetsmarknad och transportsystemet*. (Föreningen för samhällsplanering.) Plan, nr 2. Stockholm.

Arvidsson, G & Nicklasson, H, 1968, *On Consumer's Surplus and Related Concepts*. (University of Lund.) Lund.

Haley, K B & Smith, A J, 1966, *Transportation Problems with Additional Restrictions*. Applied Statistics, Nr. 2.

Hicks, 1956, *A Revision of Demand Theory*. Oxford.

Hildebrand, G & Liu, T, 1965, *Manufacturing Production Functions in the United States 1957*. Cornell University.

Lefebvre, L, 1968, *Allocation in Space*. (North Holland.) Amsterdam.

Lundqvist, L m fl, 1970, *Matematiska modeller för fysisk planering*. (KSL, regionplanekontoret.) Stockholmsregionen ur kontaktsynpunkt - en case-study. Stockholm.

Serck-Hanssen, J, 1970, *Optimal Patterns of Location*. (North Holland.) Amsterdam.



### 8.1 Bakgrund och syfte

Grundproblemet för normativa studier av markanvändningsbeslut - och kanske gäller det allmänt normativ beslutsteori - är hur långt man bör abstrahera från existerande beslutsordningar. Har man ambitionen att få fram beslutskriterier som kan användas av dagens beslutsfattare, måste man acceptera den rollfördelning i beslutsprocessen som finns och därmed avstå från möjligheten att analysera och påverka institutionella förhållanden. Väljer man i stället att studera hypotetiska beslutssituationer - med fullständig konkurrens, med total centralplanering eller något annat fiktivt samhällsmönster - måste man avstå från att påverka dagsbesluten.

I Sverige fattas besluten om markens och andra naturresursers användning enligt ett komplicerat system av regler. Om det rör sig t ex om lokalisering av en större industrianläggning börjar det med markköpsförhandlingar mellan företag och kommun följt av stadsplanprövning, finansieringsförhandlingar, byggnadstillståndsprövning etc. De roller som olika intressegrupper tilldelats i denna beslutsprocess baseras naturligtvis på det ekonomiska system vi har - ett system med fri etableringsrätt, ett utbrett enskilt markägande, med planmonopol och viss expropriationsrätt för kommunerna och med vissa statliga kontroller, skatter och andra markpolitiska instrument.

Under senare år har man diskuterat en ökad statlig styrning av åtminstone de markanvändningsbeslut, som har betydelse för flera kommuner - lokalisering av miljöpåverkande stora industrianläggningar, större naturreservat etc. Man har startat ett förberedelsearbete för en "fysisk riksplanering" och man har i politiska program krävt markanvändningsbeslut baserade på "samhällsekonomiska" snarare än kommunalekonomiska eller företagsekonomiska kriterier. I denna diskussion har det ibland ansetts att det i första hand är de samhälleliga beslutskriterierna som saknas. Man har pekat på svårigheterna att väga "ekonomiska faktorer" mot "miljöfaktorer" när t ex industrin konkurrerar med friluftslivet om naturresurserna. Men frågan är om inte själva roll-

fördelningen i beslutsprocessen är en väsentligare förklaring till dagens markanvändningsmönster. En riksplan för naturresursernas användning, som utarbetats med aldrig så förfinade metoder blir betydelselös om man inte samtidigt ändrar på maktförhållanden och planeringssystem.

Den följande framställningen är ett försök att med hjälp av en starkt förenklad modell av markkonkurrens mellan två verksamheter, diskutera marknadens roll för markanvändningsbesluten och möjligheterna att centralt styra markanvändningsmönstret. Utifrån denna diskussion skall också ett försök göras att formulera ett program för forskning om markanvändningsbeslut.

## 8.2 En modell för markanvändningsbeslut

### Resurserna

Vi antar att man genom inventeringar av den typ som beskrivs i den fysiska riksplaneringens materialsammanställning har valt ut ett antal, för en framtida tidsperiod, tänkbara lokaliseringspunkter (eller lokaliseringsområden) för olika verksamheter (Fysisk riksplanering, Materialredovisning, 1969). Vi skall anta att två verksamheter på flera platser uppträder som konkurrenter om naturresurserna. Den ena verksamheten är en varuproducerande sektor med en internationell marknad och med produkter vars värde kan mätas i ett internationellt marknadspris. Den andra verksamheten skall representera en "icke-produktiv" sektor med svår-mätbar avkastning. För att vi skall få konkretion i framställningen antar vi att det gäller det nyligen aktualiserade förhållandet med oljeraffinering som konkurrerar med fritidsboende.

Vi antar att analysen avser en framtida tidsperiod - ett par decennier lång - indelad i diskreta tidsenheter, indicerade med ett tidsindex  $t$ , som löper mellan 0 och  $T$ . De platser som kan komma ifråga för nya raffinaderier eller fritidshus nummerar vi med index  $p = 1, \dots, P$ . Begreppet plats får här uppfattas som en symbol för en uppsättning någorlunda geografiskt bestämbara naturresurser - dvs mark, vatten, mineraler eller andra resurser som inte i någon nämnvärd omfattning kan tillföras platserna genom mänskliga insatser. I fortsättningen används ibland ordet mark synonymt med ordet naturresurser i denna mening. Innebörden av konkurrensen mellan de båda sektorerna är

att dessa knappa naturresurser på en plats bara kan användas av den ena parten. På plats  $p$  kan man under den studerade perioden mellan  $0$  och  $T$  antingen bygga ut raffinaderikapaciteten till högst  $K_p'$  ton oljeprodukter per tidsenhet eller bygga högst  $K_p'$  fritidshus (eller ge rum för  $K_p'$  nya fritidshushåll eller hur man nu vill mäta det). Den utmärkande egenskapen hos marken i den innebörd vi ger åt begreppet är just omöjligheten att substituera mot andra resurser i dessa två sektorer's produktion. (Markåtgången i vanlig mening per fritidshus varierar naturligtvis med bebyggelsestyp, sötvattenbehovet i raffinaderier varierar med produktionstekniken och vatten kan transporteras mellan platser genom tunnelbyggen, men utgångspunkten i vår analys skall vara att vissa resurser på varje plats en gång för alla är givna och begränsande för produktionens storlek).

Utöver dessa fasta naturresurser krävs för raffinaderier eller fritidsboende andra rörliga resurser - dvs resurser som kan tillskapas på produktionsplatsen eller som kan transporteras dit. Vi skall anta att produktionssättet är givet i analysen för både oljeprodukter och "fritidsboendetjänster", vilket gör det möjligt att utgå ifrån givna styckkostnader för att producera en produktionsenhet i respektive sektor på olika platser och i olika tidsperioder. Dessa kostnader omfattar både investerings- och driftskostnader i vanlig bemärkelse.

Antagandet om givna styckkostnader kan också uppfattas så, att de rörliga resurserna tillhandahålls "fria" av marknader, vars prisförhållanden inte påverkas av de studerade sektorernas efterfrågan. De knappa naturresurserna antas däremot vara samhällsägda (kommunalt eller statligt) och det pris som tas ut av samhället för användningen av dessa resurser skall vi kalla för lägesränta.

#### Oljeproducenternas markanspråk

Antag nu också att man genom efterfrågeprognoser (eller "anspråksinventeringar" om vi använder terminologin i den fysiska riksplaneringen) har kunnat få fram hur efterfrågan på oljeprodukter kommer att utvecklas under den studerade tidsperioden. Vi kan anta ett givet internationellt marknadspris på oljeprodukter, och att branschorganet för raffinaderier beräknat (eller bestämt sig för) att landet från ett importöverskott vid  $t = 0$  skall nå

självförsörjning med oljeprodukter vid  $t = T$ . Denna målsättning kan t ex baseras på att produktionskostnader plus transportkostnader inom landet för alla platser ligger lägre än det internationella priset på oljeprodukter.

Beroende på hur stor andel transportkostnaderna inom landet utgör av totalkostnaderna kan det finnas anledning att dela upp landet i delmarknader och anta olika transportkostnader per enhet mellan varje par av produktionsplats - marknad. Här skall för enkelhetens skull antas att transportkostnaderna inom landet endast varierar mellan produktionsplatser och inte mellan marknadsplatser. De totala produktionsanspråkens ökning från tidpunkten noll kan vi då beskriva med en enda exogen variabel som betecknas  $E(t)$ .

Om produktionsanspråken ständigt ökar kommer markanspråken att ständigt öka. Om oljebolagen samarbetar inbördes (genom sitt branschorgan) och om de inte tar hänsyn till (eller känner till) konkurrerande markanspråk kommer de att planera för att först ta den plats som har lägsta styckkostnader, för att successivt ta allt "dyrare" platser i anspråk - (på samma sätt som Ricardos<sup>1</sup>) jordbrukare tog allt sämre jordar i anspråk). Matematiskt kan vi framställa oljebolagens planeringskriterium så här: Om  $X_p(t)$  betecknar produktionsökningen (från tidpunkten noll) på plats  $p$  och om  $c_p$  anger kostnaden för att producera och avsätta en produktionsenhet i en godtycklig tidpunkt (vi kan göra  $c_p$  tidsberoende, men detta är inte väsentligt för resonemanget), så kommer man att planera för den produktionsstrategi, som i varje tidpunkt

$$\text{minimerar } \sum_p c_p X_p(t) \quad (8.1)$$

$$\text{under restriktionerna } \sum_p X_p(t) = E(t) \quad 0 \leq X_p(t) \leq K_p$$

Om vi bortser från eventuell lägesränta och antar att marknadspriset på oljeprodukter är  $s$ , kommer vinsten per produktions-

---

<sup>1</sup> D. Ricardo, engelsk nationalekonom.

enhet på plats  $p$  att vara  $s - c_p$ . Oljeföretagen kan antas vara beredda att betala upp till  $i$  närheten av denna summa per enhet i lägesränta vid en eventuell konkurrens om marken.

#### Fritidsboendets markanspråk

Fritidsboendet skall här få representera en sektor, där producenterna = konsumenterna = de som disponerar fritidshusen inte själva har möjlighet att långtidsplanera sina markbehov. Vi skall också låta denna sektor utmärkas av betydande transportkostnader inom landet - för att kunna se hur denna faktor påverkar markanspråken. Med transportkostnader skall vi då mena fritidshusägarens kostnader för resor mellan primärbostad och fritidshus. Av denna anledning skall efterfrågan separeras på delmarknader - geografiskt bestämda till fritidshusägarnas hemorter. Utöver produktionsplatserna med index  $p$  får vi då ett nytt system av rumsliga enheter - t ex landets större tätorter - numrerade med index  $q$ .

Till kostnaden för investering och drift av ett fritidshus under en tidsperiod kommer då för en fritidshusägare med fritidshus  $i$   $p$  och hemort  $i$   $q$  en transportkostnad (exklusive eventuell lägesränta). Totalkostnaden för att disponera ett fritidshus (= en produktionsenhet boendetjänster) betecknar vi med  $c'_{pq}$ . Kan vi på basis av dessa informationer räkna ut antalet tillkommande fritidshusägare under period  $t$ ? Självfallet måste vi då veta något om hur mycket potentiella fritidshusägare från olika orter är beredda att betala för fritidshus. Fritidsutredningen försökte en gång att få fram sådana siffror och det framgick att uppgiften är förenad med många problem, se SOU 1964:47, 1965:19 och 1966:33. Vi skall anta att samhället i stället upprättat bostadsbyggnadsprogram för fritidsbebyggelse - dvs målsättningar  $E'_q(t)$  för antalet, under perioden mellan 0 och  $t$ , tillkommande fritidshusägare med hemort  $i$   $q$ . Dessa program måste då vara konsistenta med kapaciteten för fritidsbebyggelse, dvs om  $X'_{pq}(t)$  anger antalet vid tidpunkten  $t$  tillkommande fritidshusägare  $i$   $p$  med hemort  $i$   $q$  måste det gälla att

$$\sum_q X'_{pq}(t) \leq K'_p \quad \sum_p X'_{pq} = E'_q(t) \quad X'_{pq} \geq 0 \quad (8.2)$$

Vidare antar vi att fritidsboendets markanspråk representeras av den fördelning av fritidshusägandet, som uppfyller dessa konsistenskrav och som vid varje tidpunkt minimerar de totala kostnaderna

$$\sum_p \sum_q c'_{pq} X'_{pq}(t) \quad (8.3)$$

Om vi inte känner de framtida fritidshusägarnas "nytta" av att få disponera olika platser kan vi inte heller - som vi gjorde för oljeföretagen - räkna fram de lägesräntor som fritidshusägarna kan tänkas vara villiga att betala. Däremot kan vi få fram de olika platsernas knapphetsvärden för fritidshusägarna som grupp. Dessa knapphetsvärden representeras av de dualpriser som hör till restriktionerna

$$\sum X'_{pq}(t) \leq K'_p$$

i linjärprogrammet (8.2). (För en precisering av begreppet dualpris (skuggpris) se t ex Gale (1960). Dualpriset vid tidpunkten  $t$  på kapaciteten  $K'_p$  på plats  $p$  kan tolkas som kostnadsminskningen totalt för fritidsboendet vid  $t$  och en ökning av  $K'_p$  med en enhet. Om samhället tillämpade dessa dualpriser som lägesräntor och om fritidshusägarna ensamma fick disponera marken skulle kostnadsskillnaderna för fritidshusägare med samma hemort utjämnas, men vi skulle inte ha någon garanti för att någon "rättvisa" skipades eftersom folk med olika hemort och med fritidshus på samma plats ändå skulle drabbas av olika kostnader och eftersom olika platser  $p$  säkert värderas olika högt, oavsett kostnaderna.

Fritidsboendets lägesräntor får vi snarare uppfatta som de ränteintäkter från konkurrerande markanvändare (i detta fall oljeraffinaderier), som samhället är berett att avstå ifrån för att bereda plats för den politiskt beslutade fritidsbebyggelsen. Om samhället beslutat om en viss fritidsbebyggelse får man också stå för dessa kostnader.

De dualpriser som hör till restriktionerna  $\sum X'_{pq}(t) = E'_q(t)$  kan vi i enlighet med detta resonemang tolka som den subvention



samhället är berett att betala fritidshusägaren med hemort i q. Vi kan s.a.s. uppfatta samhället i två roller - en som markägare och en som fritidsboendeproducent. Samhället som fritidsboendeproducent värderar fritidsboendet till ett visst belopp (olika beroende på fritidsboendekonsumenternas hemort), som delvis går åt till att betala samhället som markägare lägesräntor. Vi skall utnyttja detta synsätt i nästa avsnitt, där vi skall diskutera avvägningen mellan de olika sektorernas markanspråk.

#### Samhällets kriterier för markens användning

Den diskussion vi fört om de båda sektorernas markanspråk och om deras förmåga att betala lägesränta har varit avsedd att visa, att samhället inte kan överlåta marken åt den "fria" konkurrensen om man vill tillfredsställa förutbestämda anspråk. Innebörden av en sådan konkurrens skulle vara att marken gick till den sektor som spontant erbjuder den högsta lägesräntan.

Vi antar i stället att samhället är berett och har förmågan att direkt reglera markens användning. Det gäller då att få fram vilken sektor som drabbas av de största förlusterna om de förbjuds att använda en omstridd plats, således att väga samman oljeproducenternas och fritidshusboendets kostnader. Den totala kostnaden för de båda sektorerna blir

$$\sum_p c_p X_p(t) + \sum_p \sum_q c'_{pq} X'_{pq}(t) \quad (8.4)$$

Samhället bör nu välja den markfördelning som i varje tidpunkt minimerar denna kostnad under en rad restriktioner. Dels får en plats bara användas av en verksamhet och denna verksamhet får inte överskrida en given produktionskapacitet, dels måste de totala anspråken någonstans tillfredsställas. Dvs alla restriktioner i programmen (1) och (2) ovan måste vara tillfredsställda och dessutom kravet att  $X_p(t)$  och  $X'_{pq}(t)$  inte för någon kombination av p och q samtidigt får vara positiva.

Det är således (teoretiskt) möjligt att räkna ut vilket markanvändningsbeslut samhället bör fatta. (Om vi nu kan acceptera så verklighetsfrämmande resultat, som att t ex ett raffineri som i början etablerats på en plats, senare slås ut av fri-

tidsbebyggelse). Men hur skall detta beslut verkställas? Antag att den totalt bästa markanvändningen beskrivs av symbolerna  $\hat{X}_p(t)$  och  $\hat{X}_{pq}(t)$  att lägesräntan totalt vid tidpunkten  $t$  på plats  $p$  sättes till  $r_p(t)$  och att  $s'_q(t)$  är det belopp samhället - i enlighet med resonemanget i förra avsnittet - är berett att betala i subvention för en fritidsboendekonsument med hemort i  $q$ . Frågan är nu vilka krav på  $r_p$ , marknadspriset på oljeprodukter  $s$ , subventionerna  $s'_q$  och kostnaderna  $c_p$  och  $c'_{pq}$  som måste gälla för att den optimala markanvändningen skall vara förenlig med ett marknadsbeteende hos oljeproducenter och hos samhället (i egenskap av markägare och i egenskap av fritidsboendeproducent).

Ett klassiskt jämviktskrav är att lägesräntan på platser med ledig kapacitet är noll. Antag att plats  $p$  till sin fulla kapacitet utnyttjas av oljeproducenterna vid en viss tidpunkt. Lägesräntan per enhet bör då vara lika med vinsten, dvs

$$(r_p/K_p) = s - c_p \quad \text{när} \quad \hat{X}_p = K_p \quad (8.5)$$

Däremot bör de platser som inte enligt den optimala strategin bör användas av oljebolagen ha en lägesränta per enhet som är så stor att den avskräcker från oljeproduktion, dvs

$$(r_p/K_p) > s - c_p \quad \text{när} \quad \hat{X}_p = 0 \quad (8.6)$$

På motsvarande sätt bör det gälla på de platser som skall användas för fritidsbebyggelse av individer med hemort i  $q$  att samhällets värdering av boendet stämmer med kostnaderna, dvs att

$$s'_q = c'_{pq} + (r_p/K_p) \quad (8.7)$$

och att för platser där det gäller att hålla fritidsbebyggelsen borta, att

$$s'_q < c'_{pq} + (r_p/K_p) \quad (8.8)$$

Nu kan man visa att det endast i speciella fall går att hitta ett sådant system av lägesräntor och subventioner - om man inte är beredd att släppa på kravet att en plats endast får användas



av en sektor. Att marknaden inte alltid fungerar i denna typ av modeller är ett fenomen av samma art som det som diskuteras av Koopmans och Beckmann i en berömd uppsats, se Koopmans, Beckmann (1957).

Försöket att sätta ett pris på naturresurserna - eller på miljön för att knyta an till ett välkänt slagord - har (i denna starkt förenklade modellvärld) således inte fungerat. Vi kommer inte ifrån de direkta markanvändningsregleringarna - den fysiska planeringen - om vi vill möta från början målsatta markanspråk.

### 8.3 Ett forskningsprogram

Den modell som presenterats är avsedd att illustrera ett mer generellt synsätt på markanvändningsplanering - eller mer allmänt - över rummet och tiden fördelade investeringsbeslut. Stommen i en analys av markanvändningsplaner bör vara en enkel, teoretisk allokeringsmodell, som gör det möjligt att i grova drag få fram ett "samhällsoptimalt" markanvändningsmönster. Dagens beslutsordning och beslutskriterier får sedan granskas utifrån detta mönster. Ett mera fullständigt program för studier av markanvändningsplanering bör omfatta följande fyra huvudpunkter:

1. En analys av de argument som företag, kommuner, statliga verk, opinionsgrupper etc har använt inför lokaliseringsbeslut och en analys av hur de olika intressena i praktiken har vägts mot varandra när besluten fattas. Två uppsatser vid Nationalekonomiska institutionen, Göteborgs universitet, som behandlar dels lokalisering av kärnkraftverk, dels lokaliseringen av OK-raffinaderiet i Brofjorden kan sägas utgöra exempel på denna typ av studier, se Gustavsson, Hansson & Nyman (1970) och Jansson, Nilsson & Olsson (1970).
2. Ett försök att formulera kriterier för hur de olika intressena bör vägas mot varandra i en hypotetisk beslutssituation - där staten har möjlighet att fatta de beslut (eller decentralisera de beslut) man önskar.
3. Ett försök till jämförelse mellan (å ena sidan) ett prognostiserat framtida lokaliseringsmönster vid oförändrad besluts-

ordning och å andra sidan ett konstruerat "optimalt" lokaliseringsmönster. Dessa studier får till en början bli partiella - avse vissa sektorer eller områden.

4. En diskussion av de förändringar i ägandeförhållanden, byggnadslagstiftning, skattesystem etc, som skulle krävas för statlig styrning mot ett önskat lokaliseringsmönster.

8.4 Litteraturreferenser:

Fysisk riksplanering, *Materialredovisning*, 1969. (Kommunikationsdepartementet). Stencil K 1969:13, 14, 15. Stockholm.

Gale, 1960, *The Theory of Linear Economic Models* (McGraw Hill). New York.

Gustavsson, Hansson, Nyman, 1970, *OK-lokaliseringen* (National-ekonomiska institutionen, Göteborgs universitet). Stencil. Göteborg.

Jansson, Nilsson, Olsson, 1970, *Kriterier för lokalisering av kärnkraftverk*. (Nationalekonomiska institutionen, Göteborgs universitet). Stencil. Göteborg.

Koopmans, T C, Beckmann, M J, 1957, *Assignment Problems and the Location of Economic Activities*. *Econometrica*, Vol 25, p 53-76.

SOU 1964:47, 1965:19, 1966:33, *Friluftslivet i Sverige*. (Statens offentliga utredningar). Stockholm.

## NÅGRA ORDFÖRKLARINGAR

adaptiv	ett adaptivt kontrollproblem består i att bestämma optimal kontroll för ett system utan att från början ha fullständig information om systemets struktur. Med ledning av resultaten av vidtagna åtgärder ökar successivt kunskapen om systemstrukturen
allokering	fördelning av givna resurser på olika användningsområden
autokorrelation	statistisk storhet, som anger graden av samvariation i t ex konsekutiva observationer i en tidsserie
avbildning	regel för tillordning av element i en mängd till element i en annan, t ex $y = f(x)$ , avbildningen är $f$ i detta fall
dekomposition	nedbrytning av ett optimeringsproblem i mindre delar, den mest kända metoden för linjära problem är Dantzig-Wolfes metod
determinant	ett tal som har stor betydelse för egenskaperna hos en kvadratisk matris och därmed för t ex linjära ekvationssystem
deterministisk	ingen slumpmässighet förekommer i vare sig variabler eller samband i en sådan matematisk modell
diagonaldominant	om i en kvadratisk matris diagonalelementen är stora relativt övriga element så har matrisen denna benämning
diskret	motsatsen till kontinuerlig, en sådan stor-

het kan bara anta vissa värden, t ex bara heltal

dual	se även dualproblem, primalproblem. Mest välkänd är teorin för dualitet vid linjär programmering, men motsvarande resultat finns även för icke-linjär programmering och heltalsprogrammering. Intäktsmaximering och kostnadsminimering är exempel på duala problem i linjär produktionsteori
dualproblem	ett linjärprogram kan alltid ges två ekvivalenta utformningar, primalproblemet och dualproblemet
egenvärde	varje kvadratisk matris har en uppsättning sådana, erhållna ur ekvationen $Ax = \lambda x$ , för vissa $x$ kan $A$ ersättas med egenvärdena $\lambda$
ekonometrisk	ekonometri behandlar problemet att statistiskt fastställa samband i (ekonomiska) modeller, dvs att mäta och testa samband
endogen	bestämd inom ramen för en matematisk modell
ensemble	en samling tänkta kopior av ett studerat system, termen användes i samband med statistiska modeller
entropi	ett fördelningsmått som används inom statistik och fysik, ju jämnare fördelning, desto högre entropi
ergodicitet	hypotes som rättfärdiggör ersättning av tidsmedelvärden med medelvärden över tvärsnitt vid en given tidpunkt

estimation	bestämning av parametervärden i en matematisk modell ur observationer med statistisk metodik
Euler-Lagranges ekvationer	ett ekvationssystem som ger den stationära lösningen till optimeringsproblem inom variationskalkylen
Eulers metod	enkel metod för approximation av derivata med differenskvot
exogen	a priori given utanför ramen av en matematisk modell
external economies	sammanfattande term för vissa ömsesidiga beroenden inom det ekonomiska systemet
fasrum	detta erhålles om alla de storheter som krävs för att karakterisera ett systems tillstånd avsätts längs varsin axel i ett rätvinkligt koordinatsystem
funktional	ex.: om $y = F(z,x)$ där argumentet $z$ är en funktion $z = g(x)$ , så kallas $F$ en funktional
graf	den kurva i rummet som en funktion beskriver eller, likvärdigt, mängden av punkter på denna kurva
heuristisk metod	sammanfattande namn på metoder att via förenklingar och approximationer söka bra lösningar (ger ej med säkerhet optimum)
hyperkub	beteckning för en kub i mer än tre dimensioner
input	i kontrollteoretiska och reglertekniska sammanhang en beteckning för inmatning av variabelvärden i ett studerat system

invers	om $y = f(x)$ och funktionen $f$ uppfyller vissa villkor finns det en funktion $g$ från $y$ till $x$ sådan att $x = g(y)$ , vilken kallas inversen till $f$
kausaltitet	orsakssamband gäller
kedjeregeln	deriveringsregel, om $y = f(u(x))$ så ger denna att $y' = \frac{df}{du} \frac{du}{dx}$
komposition	sammansättning av två funktioner
konsistens	en teori är konsistent om den är logiskt motsägelsefri
kontinuerlig	motsatsen till diskret. En kontinuerlig variabel antar godtyckligt tätt liggande värden i ett intervall
konvexitet	en konvex funktion är sådan att varje punkt på en rät linje mellan två funktionsvärden ligger ovanför funktionens värde i samma punkt. Motsatsen till konvex är konkav
Kuhn-Tucker-villkor	nödvändiga villkor för att ett minimum skall ha erhållits i ett matematiskt minimeringsproblem
linjärprogram	ett optimeringsproblem med bivillkor där både kriterium och bivillkor är linjära funktioner
maximum-likelihood-metod	statistisk estimationsmetod, där parametrar bestäms genom maximeringsförfaranden
metriskt rum	en viss typ av matematiskt rum, där begreppet avstånd finns definierat

minimum	för en konvex funktion finns endast ett minsta värde, det globala. I annat fall kan även lokala minima förekomma
optimum	det högsta (lägsta) värde som en kriteriefunktion, avpassad för maximering (minimering) kan anta, eller motsvarande variabelvärden
output	de variabelvärden som kan avläsas som resultat av kombinationen av ett systems inre egenskaper och de inmatade variabelvärdena
paretoeffektiv	ett sådant alternativ i en alternativmängd har egenskapen att effektiviteten i något avseende ej kan höjas utan att samtidigt effektiviteten i något annat avseende sänks
permutation	omordning av en ordnad mängd element, $(2,1,3)$ , är t ex en permutation av $(1,2,3)$
positivt definit	en egenskap hos vissa kvadratiska matriser som gör ett kvadratisk minimeringsproblem i termer av denna matris konvext
positiv linjär avbildning	en linjär avbildning $y = ax + b$ är positiv om $a > 0$ .
prediktion	förutsägelse; det är ett mål med de flesta beskrivande modeller att använda dem för förutsägelser om framtiden
primalproblem	varje linjärprogrammeringsproblem kan ges två ekvivalenta formuleringar, primal och dual
probabilistisk	styrd av sannolikheter; i en sådan modell är variabler och samband behäftade med slumpmässighet



regressionsanalys	statistisk metod för att anpassa matematiska funktioner till ett observationsmaterial
rum	en mängd element och ett antal tillåtna matematiska operationer, se t ex vektorrum
skalekonomier	stordriftsfördelar, dvs iakttagelsen att produktionskostnaden/enhet sjunker med växande produktionsskala
stationär lösning	två betydelser: 1) en svagare egenskap än maximum eller minimum hos en funktion 2) ett över tiden konstant tillstånd hos ett system
stokastisk	slumpmässig; motsatsen till en deterministisk variabel kallas en stokastisk variabel
tillståndsrum	ekvivalent med beteckningen fasrum, ett systems tillstånd kan ses som en punkt i detta rum
topologiskt rum	här finns inte begreppet avstånd definierat, öppna mängder spelar en central roll
trajektorier	bana i tillståndsrummet utgående t ex från någon känd startpunkt
transformation	i matematisk mening ekvivalent med begreppet avbildning
transient	övergående; om ett system utsätts för en störning återgår det först så småningom till jämvikt, övergångsperioden kallas transientfas

triangelolikhet	det faktum att summan av längderna av två sidor i en triangel alltid är större än eller lika med längden av den tredje
trim-problem	ett problem som innebär att med given ordersammansättning skära ett antal pappersrullar så att den totala spillmängden minimeras
unimodal	en sådan funktion har bara ett maximum (eller minimum)
vektorrum	elementen består här av vektorer, som karakteriseras av storlek och riktning och de tillåtna operationerna är addition och multiplikation med skalära tal
ändligt dimensionellt rum	om rummet är t ex ett vektorrum har vektorerna ändligt många komponenter; det finns bara ett ändligt antal koordinat-axelriktningar
öppen mängd	en sfär med skalet undantaget är exempel på en öppen mängd

## MATEMATISKA SYMBOLER

$\frac{\partial F}{\partial x}$	partialderivatan av $F$ med avseende på $x$ .
$f:A \rightarrow B$	funktionen $f$ avbildar mängden $A$ på mängden $B$ . Funktionen $f$ tillordnar till varje element i $A$ ett eller flera element i $B$ .
$x \in X$	elementet $x$ tillhör mängden $X$ .
$L_1 \text{ omm } L_2$	$L_1$ gäller om och endast om $L_2$ gäller, dvs. $L_2$ är ett nödvändigt och tillräckligt villkor för $L_1$ .
$\exists$	det existerar
$\ni$	så att, med den egenskapen att
$\{ \}$	mängden av det element som innesluts av klamrarna
$E\{u\}$	medelvärdet av den stokastiska variabeln $u$
$ $	betingat av
$\Pi$	produkttecken (jämför $\Sigma$ för summa)
$a < b$	$a$ domineras av (är mindre än) $b$ i en partialordning
$uSv$	$S$ betecknar en relation, t ex $<$ . $uSv$ är sann om $u$ står i relationen $S$ till $v$
$f \circ u$	sammansättning, om $f$ är en transformation från mängden av nyttofunktionaler till mängden av nyttofunktionaler och $f$ avbildar $u$ på $v$ kan nyttofunktionalen $v$ skrivas som sammansättningen $f \circ u$ .
$A \subset B$	mängden $A$ innehålls i mängden $B$
$\equiv$	identiskt lika med
$\tau \propto 10^{-6}$	$\tau$ är av storleksordningen $10^{-6}$



**R15: 1972**

**Denna rapport avser anslag Bs 471 från Statens råd för byggnadsforskning till tekn. lic. Anders Karlqvist, matematiska institutionen, KTH, Stockholm. Försäljningsintäkterna tillfaller fonden för byggnadsforskning.**

**Distribution: Svensk Byggtjänst, Box 1403, 111 84 Stockholm  
Grupp: samhällsplanering**

**Pris: 28 kronor**